Gemini Science Archive

Functional and Performance Requirements Document (Revised Initial)

Version 04

30 November 2000

Norman Hill, Séverin Gaudet, Daniel Durand, David Schade, David Bohlender

National Research Council of Canada
Herzberg Institute of Astrophysics
Canadian Astronomy Data Centre

Felipe Barrientos, Felipe Richardson

Proyecto Gemini-Chile
Comisión Nacional de Investigatión Científica y Tecnológica

Table Of Contents

1. Introduction 1
1. Purpose 1
2. Overview 1
3. Abbreviations and Acronyms 1
4. Glossary 3
5. References 10
2. General Description 11
1. History 11
2. Relation to Current Projects 11
2.1 CADC Archive Development 11
2.2 The Gemini On-line Data Processing System 12
3. Relation to Predecessor and Successor Projects 12
3.1 The CADC Archive Infrastructure 12
3.2 The Gemini Engineering Archive 12
3.3 The Gemini Data Handling System 12
3.4 Ongoing development of the GSA 12
4. Function and Purpose 13
5. Mission Statement 13
6. User Characteristics 13
7. Relation to Other Systems 14
3. Functional and Performance Requirements 17
1. Introduction 17
2. GSA Software Requirements 39
2.1 Functional Requirements 40
SR1.1 The GSA shall archive all data. 40
SR 1.2 The GSA shall respect proprietary periods assigned by Gemini. 41
SR 1.3 The GSA shall be able to automatically process data. 41
SR 1.3.1 Data processing shall respect Gemini policy on proprietary periods. 41
SR 1.3.2 Data processing shall proceed promptly. 42
SR 1.4 Preview data should be generated for each dataset. 42
SR 1.4.1 Generate previews for associations. 42
SR 1.4.2 Remove previews when release date is changed. 43
SR 1.4.3 Preview should be mirrorable. 43
SR 1.5 Collect information needed to track of publications based on archive data. 43
SR2.1 The GSA catalogue shall contain all data descriptors. 44
SR 2.2 The catalogue shall include environmental information. 44
SR 2.3 The catalogue shall include accurate WCS information. 44
SR 2.4 Proposal information should be associated with data supersets. 44
SR 2.5 The catalogue should include image descriptors derived from the data. 45
SR 2.5.1 Remove derived image descriptors for proprietary data. 45
SR 2.6 Release date is a data descriptor. 45
SR 2.7 The GSA catalogue should contain associated datasets. 45
SR 2.8 Data quality assessment shall be in the catalogue. 46
SR 2.9 The science program shall be associated with the catalogue data. 46

SR 2.10 The electronic observing log shall be associated with the catalogue. . . . 46
SR 2.11 Proposal id shall be a descriptor. . . . 46
SR 2.12 The GSA catalogue shall be mirrorable. . . . 46
SR 2.13 Calibration data shall be associated with science data. . . . 47
SR 2.14 The GSA should have an object catalogue. . . . 47
SR 2.14.1 Object parameters should be extracted from object catalogues. . . . 47
SR 2.14.2 The GSA object catalogue should be up to date. . . . 48
SR 2.14.3 The object catalogue should include object descriptors derived from data. . . . 48
SR 2.14.4 Remove derived objects for proprietary data. . . . 48
SR 2.15 Publications should be linked to data supersets. . . . 48
SR 2.16 The GSA shall store Gemini hardware and software versions. . . . 49
SR3.1 The GSA shall have an interactive user interface. . . . 49
SR 3.2 Select subsets of the GSA catalogue. . . . 49
SR 3.2.1 Select subsets of the GSA catalogue based on coordinate region search. . . . 49
SR 3.2.2 GSA users should be able to select data supersets based on object descriptor. . . . 50
SR 3.2.3 Support cross catalogue searches. . . . 50
SR 3.2.4 The GSA should support a batch mode interface. . . . 50
SR 3.3 View descriptors for a subset of the catalogue. . . . 51
SR 3.3.1 There shall be on-line help describing all catalogue data descriptors. . . . 51
SR 3.3.2 GSA users can view a list of calibration data supersets. . . . 51
SR 3.3.3 View the list of publications associated with a data superset. . . . 51
SR 3.3.4 View descriptors of objects in a data superset field of view. . . . 51
SR 3.3.5 View proposal information associated with a data superset. . . . 52
SR 3.3.6 Sort the displayed subset of the catalogue. . . . 52
SR 3.3.7 Sort the displayed objects. . . . 52
SR 3.3.8 Plot a subset of data supersets in spatial coordinates. . . . 52
SR 3.3.9 GSA users shall be able to view the electronic observing log. . . . 52
SR 3.3.10 View environmental data. . . . 52
SR 3.3.11 Display science programs. . . . 53
SR 3.3.12 Display the members of association data supersets. . . . 53
SR 3.3.13 Save a set of data supersets to a file. . . . 53
SR 3.3.14 Load a set of data supersets from a file. . . . 53
SR 3.4 GSA users shall be able to select data for retrieval. . . . 53
SR 3.4.1 Users shall retrieve data via Internet, EXABYTE, or CD-ROM. . . . 54
SR 3.4.2 Only authorized users shall be able to retrieve proprietary data. . . . 54
SR 3.4.3 GSA users should be able to choose to have data processed. . . . 54
SR 3.4.4 Data processing shall not use proprietary data as input. . . . 54
SR 3.4.5 Processing recipes shall be documented. . . . 55
SR 3.4.6 Authorized Gemini staff can retrieve proprietary data. . . . 55
SR 3.4.7 GSA users can retrieve calibration data. . . . 55
SR 3.4.8 Retrieve environmental data. . . . 55
SR 3.4.9 It shall be possible to retrieve science programs. . . . 55
SR 3.4.10 It shall be possible to retrieve proposal information. . . . 56
SR 3.4.11 It shall be possible to retrieve the electronic observing log. . . . 56
SR 3.4.12 It should be possible for users to monitor the progress of data requests. . . . 56
SR 3.4.13 Batch mode requests. . . . 56
SR 3.4.14 Document user data requests. . . . 56

SR 3.4.15 Retrieve associated data supersets. 57
SR 3.4.16 Provide data processing recipes for associations. 57
SR 3.5 A general description of the content of the GSA shall be available. 57
SR 3.5.1 A description of the Gemini instruments shall be available. 57
SR 3.5.2 A description of the Gemini facilities shall be available. 58
SR 3.6 Proposal information should be searchable. 58
SR 3.6.1 Find all data supersets associated with proposals. 58
SR 3.7 Preview display for each science data superset. 58
2.2 Data flow Requirements 58
SR4.1 Maximum and average requirements for catalogue ingest. 58
SR 4.2 Maximum and average requirements for raw data ingest. 59
SR 4.3 Maximum and average requirements for data retrieval by users. 59
SR 4.4 Preview should be display promptly. 59
SR 4.5 Access to the GSA catalogue must be interactive. 59
SR 4.6 Data for Internet retrieval should be available promptly. 60
2.3 Operational Requirements 60
SR5.1 Participate in catalogue unification efforts. 60
SR 5.2 Data shall be physically secure. 60
SR 5.3 Data shall be electronically secure. 60
SR 5.4 GSA staff shall respect proprietary periods. 60
SR 5.5 GSA operations staff shall support GSA users. 61
SR 5.6 GSA operations staff shall support ongoing development of Gemini. 61
SR 5.7 GSA operations staff shall incorporate evolving technologies into the GSA. 61
SR 5.8 The GSA operations staff should add new features to the GSA. 61
2.4 Resource Requirements 61
SR6.1 GSA development cost. 61
SR 6.2 GSA operating cost. 62
2.5 Requirements on Gemini 62
SR7.1 Gemini shall collect instrument configuration descriptors. 62
SR 7.2 Gemini shall collect telescope configuration descriptors. 62
SR 7.3 Gemini shall record weather information. 62
SR 7.4 Gemini shall collect accurate WCS information. 62
SR 7.5 Gemini shall make proposal information available to the GSA. 62
SR 7.5.1 Gemini shall provide passwords for executed proposals. 62
SR 7.5.2 Gemini shall provide information to link proposals with datasets. 62
SR 7.6 Gemini shall provide a data quality assessment of each dataset. 63
SR 7.7 Gemini should track publications associated with data. 63
SR 7.8 Gemini shall provide an electronic observing log. 63
SR 7.9 Supply executed science programs to the GSA. 63
SR 7.10 Gemini shall provide a general description of the telescope. 63
SR 7.10.1 Gemini shall provide a description of the instruments. 63
SR 7.10.2 Gemini shall provide a description of Gemini facilities. 64
SR 7.11 Gemini shall provide descriptions of data descriptors. 64
SR 7.12 Gemini shall provide raw science data to the GSA. 64
SR 7.13 Gemini shall ensure that calibration data is available. 64

SR 7.13.1 Gemini shall provide information to identify calibration data. . . . 65
SR 7.14 Gemini should provide data processing recipes. . . . 65
SR 7.14.1 Gemini should provide descriptions of data processing recipes. . . . 65
SR 7.15 Gemini shall provide a release date for each dataset. . . . 65
SR 7.15.1 Gemini shall inform the GSA when release dates change. . . . 65
SR 7.15.2 Inform the GSA which Gemini staff are permitted to access proprietary data. . . . 65
SR 7.16 Gemini shall provide support for GSA users. . . . 65
SR 7.17 Gemini shall provide version information. . . . 66
SR 7.18 Gemini shall provide test data. . . . 66
2.6 External Requirements . . . 66
SR8.1 GSA needs access to SIMBAD and NED catalogues. . . . 66
SR 8.2 A unified astronomical archive catalogue should be created. . . . 66
4. Effect of the GSA on Gemini Systems . . . 67
1. Introduction . . . 67
2. Gemini Components . . . 70
2.1 Gemini Observing Tool . . . 70
2.2 Gemini Engineering Archive . . . 71
2.3 Gemini Observing Database . . . 71
2.4 Electronic Observing Log . . . 72
2.5 Gemini Operations Staff . . . 72
2.6 Observatory Control System . . . 74
2.7 DHS Storage Server Archive media . . . 74
2.8 Telescope Control System . . . 75
2.9 Gemini Instruments . . . 75
2.10 Gemini facilities . . . 75
3. WCS coordinates . . . 75
4. Gemini Policy . . . 76
4.1 Distribution of proprietary data to PI . . . 76
4.2 Distribution of proprietary data to non-PI. . . . 76
4.3 Proprietary periods for various classes of data . . . 76
4.4 Gemini support for GSA users . . . 76
5. Review Comments and Responses . . . 77
1. Final Review of the FPRD . . . 77
1.1 Inger Jørgensen . . . 77
2. Draft Review Comments . . . 77
2.1 Fred Gillett . . . 77
2.2 Ted von Hippel . . . 79
2.3 Inger Jørgensen . . . 81
2.4 Felipe Barrientos . . . 84

Created:February 16, 2000
Modified:November 30, 2000

Gemini Science Archive

Chapter 1

Introduction

1. Purpose

This document describes the functional and performance requirements for the GSA, including the origin of each requirement, the assumptions about the characteristics and performance of the Gemini systems, and indicates if the current performance of these Gemini systems support the GSA requirements.

2. Overview

This document is divided into the following chapters:

- Chapter 1 is an introduction to this document.
- Chapter 2 gives a general description of the GSA, its reason for existence, and describes how it fits within Gemini.
- Chapter 3 gives a detailed list of the functional and performance requirements, as derived from the user requirements listed in [8], and the requirements listed in [6]. Chapter 3 includes separate sections listing the requirements on other Gemini systems, and requirements that are external to both the GSA and to Gemini.
- Chapter 4 lists each of the Gemini subsystems (in this chapter, we also include Gemini operations staff as a subsystem) that we believe will be effected by the requirements on Gemini listed in Chapter 3, lists the requirements which we believe effect that subsystem, and describes our assumptions on whether or not the subsystem currently has the functionality required. This chapter also includes a section describing the policy decisions that we believe must be made be Gemini in order to clarify aspects of the development and operation of the GSA.

3. Abbreviations and Acronyms

API	Application Programming Interface.
ASCII	American Standard Code for Information Interchange.
BDS	The abbreviation used for the Gemini Science Archive bulk data storage subsystem.
CADC	Canadian Astronomy Data Centre.
CAT	The abbreviation used for the Gemini Science Archive catalogue subsystem.
CD-R	Compact Disc Recordable.
CD-ROM	Compact Disc Read Only Memory.
CoI	Co-Investigator (see glossary).
DAT	Digital Audio Tape.
DBMS	Data Base Management System.

DHS	Data Handling System.
DI	The abbreviation used for the Gemini Science Archive data ingest subsystem.
DP	The abbreviation used for the Gemini Science Archive data processing subsystem.
DPR	Data Processing Recipe (see glossary).
DR	The abbreviation used for the Gemini Science Archive data Retrieval subsystem.
DVD	Digital Video Disk (sometimes referred to as Digital Versatile Disk).
DVD-R	Digital Video Disk (sometimes referred to as Digital Versatile Disk), recordable.
EMSS	Einstein Medium Sensitivity Survey.
ESA	European Space Agency.
FITS	Flexible Image Transport System.
FPRD	Gemini Science Archive Functional and Performance Requirements Document (see glossary).
FTP	File Transfer Protocol.
GB	Gigabytes (1073741824 bytes).
GEA	Gemini Engineering Archive (see glossary).
GSA	Gemini Science Archive (see glossary).
GSC	Gemini Science Committee.
GSCG	Gemini Software and Controls Group.
GSRS	Gemini Software Requirements Specification.
GUI	Graphical User Interface.
HIA	Herzberg Institute of Astrophysics.
HROS	High Resolution Optical Spectrograph.
HST	Hubble Space Telescope.
HTTP	HyperText Transfer Protocol.
ICD	Interface Control Document.
IRAS	Infrared Astronomical Satellite (see glossary).
ISO	International Standards Organization.
IVO	International Virtual Observatory (see glossary).
MB	Megabytes (1048576 bytes).
MEF	Multi-Extension FITS file.
NED	NASA/IPAC Extragalactic Database (see glossary).
NGST	Next Generation Space Telescope.
NIRI	Near InfraRed Imager (see glossary).
NVO	National Virtual Observatory (see also IVO).
OCDD	Gemini Science Archive Operational Concept Definition Document (see glossary).
OT	Gemini Observing Tool.
PI	Principal Investigator (see glossary).
PSF	Point Spread Function (see glossary).

RDBMS	Relational Database Management System.
S/N	Signal to Noise ratio.
SDD	Software Design Description.
SIMBAD	Set of Identifications, Measurements and Bibliography for Astronomical Data (see glossary).
SP	Science Program (see glossary).
SQL	Structured Query Language (see glossary).
TB	Terabytes (~1.1×10^{12} Bytes).
TBD	To Be Determined.
UI	User Interface. In this document, this refers to the archive user interfaces to the GSA.
URL	Universal Resource Locator.
WCS	World Coordinate System.
WPD	Work Package Description.
WWW	World Wide Web.

4. Glossary

Altair — A natural guide star adaptive optics system being built for the Gemini north telescope.

Archive — In the GSA documents, the term archive is the set of scientific astronomical data, intended to be accessible to the general astronomical community. An archive provides the tools necessary to allow an astronomer to easily identify and retrieve those data which will contribute to a specific area of study.

Associated Data Superset — This is used to refer to the concept of using data processing to merge several inputs into a single output. For example, several datasets containing data for the same field could be averaged into a single associated data superset. Within the CADC, we tend to refer to associations as image data of the same field, however we are trying to generalize the concept to include other associations, such as mosaics of adjoining images, averaging of spectral data of a target (where the spectra cover the same band), and concatenation of spectra of a target (where the spectra do not cover the same band). An associated data superset may be a virtual concept, with only a database entry describing how members could be combined, or it may be a FITS file which contains the results of data processing. See also Data Superset.

Association — See “Associated Data Superset”.

Astronomical Server URL — A standard proposed which will allow standardized access to astronomical catalogues over the world wide web. (See http://vizier.u-strasbg.fr/doc/asu.html.)

Calibration — This is the processing of raw data to remove instrument or environmental signatures. This is usually the first processing step towards making raw data scientifically useful.

Catalogue — There are two types of catalogues which are significant to the GSA.

- An archive catalogue contains descriptor values for each scientific observation in an archive, and it is the primary tool used by astronomers when identifying data for

retrieval. The descriptor values are properties of the data as a whole, e.g. WCS information, S/N for the data, primary target, etc.

- An object catalogue is a list of astronomical objects, with associated descriptors. In this type of catalogue the descriptors describe the objects. An object catalogue may be derived from the data in a single archive (e.g. an HST object catalogue is under development at the CADC), from a single project (e.g. EMSS), or they may be compiled from a variety of sources (e.g. SIMBAD, or NED). Entries in an object catalogue could have links to an archive catalogue, identifying the data from which the object descriptors were derived.

Classical Observing — This is a mode of observing where the execution of a science program is overseen by a PI or a CoI who is present at the telescope while the science program is being executed. See also Queue Observing, Remote Observing, and Service Observing.

Client — A computer program which uses a **"Server"** to carry out some action which it cannot otherwise do itself. The client commands the server and uses it as a system resource.

Co-Investigator — A person collaborating with a Principal Investigator (see glossary) on a project.

Configuration — A description of the state of one or more components of a system. For example, a Gemini instrument controller configuration might contain the location of the filter wheel, position of the grating etc. A state described by a configuration can be dynamic. For example an instrument could be continuously changing the state of the chopper in one "configuration".

Context diagram — A special case of a "Data Flow Diagram (DFD)" in which a single bubble represents the entire system. All boxes in a context diagram represent objects in the outside world.

Cross catalogue query — A query that correlates the content of two or more catalogues. An example of a cross archive query would be "find all NIRI data from the Gemini Science Archive where there also exists at least one observation in the JCMT SCUBA catalogue with an overlapping field of view". (See also Multi-archive query.)

Cross archive search — See "Cross catalogue query".

Data Handling System (DHS) — That part of the Gemini Control System responsible for displaying, processing, saving and archiving data in a standard way.

Data Flow Diagram (DFD) — A diagram showing how data is handled by a software system.

Data label — A unique label which is used to identify a dataset created by the Gemini Telescopes. A data label can be used to uniquely identify the dataset for as long as the dataset exists.

Data Mining — The extraction of information or knowledge from large databases.

Data Package — The assembly of data packaged and distributed together on removable media (CDs or DVDs) as either science data, facility data, or baseline calibration data. A data package includes all relevant observational data (pixel data and meta-data) and may also include relevant PR images and information. A data package may contain data from one or more observations. A data package may require one or more volume of the removable media.

Data Processing Recipe (DPR) — A recipe describing how to process a particular type of data.

Data Quality — A single descriptor which gives an indication of the usability of a dataset. Tentatively, the data quality for data from the Gemini telescopes should categorize the data quality as follows:

- The data does not meet Gemini standards and has no scientific value. This data will not be visible to GSA users.
- The data meets Gemini standards, but does not meet the standards required for the science program.
- The data meets Gemini standards and meets the standards required for the science program.

Data superset — A data superset is either a single dataset, or an associated data superset (see "Associated Data Supserset").

Dataset — Data created from a single OBSERVE command. A dataset consists of header data describing the dataset and one or more frames. Each of the frames has a separate header containing any information specific to the individual frames. All data from a dataset is eventually stored in a single FITS file, with each frame stored in a FITS extension.

Descriptors — The data which describes a dataset. This would include data from a dataset's FITS file headers, and data from other sources, such as weather data, and proposal information.

DHS database — The DHS database contains records of datasets collected by the telescope, along with a selected set of header parameters, the history logs from the various principal systems, the data dictionary, information about archive media created by the DHS, etc.

External Requirements — These are requirements which are beyond the scope of the GSA implementation and operations, and the Gemini project as a whole. See also Functional Requirements and Usability Requirements.

File Transfer Protocol (FPT) — A widely available protocol for transferring data files over the internet.

FITS extension — A FITS extension is part of a FITS format file. FITS extensions are used to allow more than one image or table to be stored in a single FITS file. Each FITS extension has a separate header area, which contains meta data describing the data content of the extension.

Note: The content of a FITS extension could be meta-data further describing the data in the primary FITS data, or describing the data in another FITS extension.

FITS format file — A file obeying the Flexible Image Transport System (FITS) standard (see http://fits.gsfc.nasa.gov). For the GSA project, this document the term "FITS format file" will refer to a file written to disk.

FITS header — A FITS header as written to a FITS format file. This consists of a series of 80-byte ASCII card images. Each record contains a unique 8 character keyword identifying the data item, the value of the item encoded as an ASCII string, and an optional comment describing the item. The value of a data item must be a single integer, floating point or string value.

Functional Requirements — These are the requirements the GSA must meet in order to make possible and practical, all of the science described in the GSA OCDD. See also Usability Requirements and External Requirements.

Gemini Engineering Archive (GEA) — The purpose of the GEA is to collect data from all of the Gemini systems and to allow efficient retrieval of the data for the purpose of analysing the system behaviour. To avoid duplication, the GEA may contain pointers to data in existing databases such as the DHS permanent store and Observing database. The DHS permanent store contains all image data, image processing histories and recipes, and completed science programs. The Observing database contains planned, active and completed science program information.

Gemini Science Archive (GSA) — An archive system intended to preserve scientific data collected by the Gemini Telescopes. After the proprietary period for the data has expired, the data is made available to the general astronomical community.

Gemini Science Archive Functional and Performance Requirements Document (FPRD) —

This document describes the functional and performance requirements to be included in the Science Archive, and describes functional and performance requirements required from other Gemini Systems.

Gemini Science Archive Operational Concept Definition Document (OCDD) — This document describes the science cases for which the Science Archive is designed, contains the scenarios which show how a GSA user would interact with the GSA in order to collect the data required for each of the science cases, and lists the user requirements which arise from the scenarios.

Gemini Science Archive User — For the purposes of these documents, GSA users are considered to be professional astronomers, or students in a university level astronomy program. The GSA is not intended to provide data to the general public.

Hipparcos — Hipparcos is a pioneering space experiment dedicated to the precise measurement of the positions, parallaxes and proper motions of the stars. The intended goal was to measure the five astrometric parameters of some 120 000 primary program stars to a precision of some 2 to 4 milliseconds of arc.

Infrared Astronomical Satellite (IRAS) — An infrared space telescope. It conducted an all-sky survey at wavelengths ranging from 8 to 120 microns in four broad-band photometric channels centred at 12, 25, 60, and 100 microns. Some 250,000 point sources were detected. A catalogue of small (< 8') extended sources gives the characteristics of some 20,000 objects. An atlas of images covering the entire sky gives the absolute surface brightness at each of the four survey wavelengths.

Instrument — A device, consisting of a detector and a means of dispersing and/or focusing light onto that detector, designed to extract a particular type of information from the radiation gathered by the Gemini telescope. An instrument is specially designed for astronomy research. The wavefront sensors on the Gemini telescope can be regarded as an instrument.

Instrument Control System (ICS) — That part of the Gemini Control System which is responsible for controlling and operating a scientific instrument. There is a separate Instrument Control System for each available scientific instrument.

Interactive Observing — A mode of observing where details of the sequence of events are decided in real-time while performing the observation. See also Planned observing.

Interface Control Document (ICD) — A document which either describes a general interface between Gemini systems, or which describes in detail the interface between two specific Gemini systems.

International Virtual Observatory (IVO) — Also known as the National Virtual Observatory (NVO). This is a proposed virtual observatory which will incorporate catalogues and archive data from a variety of surveys and observatories.

ISO 9660 — The international standard defining the CD-ROM data format.

Logging — The act of keeping a record of happenings during an observing session. “Logging” encompasses the observing log, system history records and any engineering logs made for the day crew.

Meta-data — Meta-data is data which describes data. This can be a recursive concept in that the header of a FITS file contains keywords and values which are Meta-data for an image stored in the file, but there could also be Meta-data describing the FITS keywords and values (permitted range or values, units, description, etc.).

Multi-catalogue query — A query which spans more that one catalogue, but does not directly correlate the content of the two catalogues. An example of a multi-catalogue query would be “find all images of a specified object in the Gemini catalogue, in the HST catalogue, or in the CFHT catalogue. (See also cross archive query.)

Multi-archive search — See “Multi catalogue query”.

NED — NED is a catalogue built around a master list of extragalactic objects for which cross-identifications of names have been established, accurate positions and redshifts entered to the extent possible, and some basic data collected. Bibliographic references relevant to individual objects have been compiled, and abstracts of extragalactic interest are kept on-line. Detailed and referenced photometry, position, and redshift data, have been taken from large compilations and from the literature. NED’s data and references are being continually updated, with revised versions being put on-line every 2-3 months.

NICMOS — An HST instrument for both imaging and multi-object spectroscopy.

NIRI — A Gemini north instrument designed for near infrared imaging and spectroscopy.

Non-conforming instrument — An “Instrument” which does not conform to the Gemini hardware or software requirements. Such an instrument is only allowed as a **“Visiting instrument”.**

Observation — An observation is the smallest scheduable event for the Gemini telescopes. A Science Program may consist of multiple observations, and an observation may consist of several datasets, each with a different instrument and/or telescope configuration.

Observing log — The observing log is a list of all observations taken and key attributes of the observations, including time of observation, observer comments, object of observation, etc.

Observing Modes — As defined in the Gemini Software Design Description, p5-46, the observing modes are: interactive, queue-based, service and remote.

Observing Tool — The primary astronomer interface of the Gemini control system.

Planned Observing — Planned observing is a science program where the details of the observing process are planned in advance. Planned observing is required for both Queue Observing and Service Observing, and is strongly encouraged for classical users. See also Interactive Observing.

Point Spread Function — A measure of image quality.

Preview data — A representation of a dataset intended for display to an archive user. The intention of preview data is to allow the archive user to quickly view an image or spectral plot, giving an indication of the data quality and usefulness. Preview data is a calibrated, compressed (possibly with lossy compression) version of the data.

Principal Investigator (PI) — The Gemini user identified as being the person who has sole access to the science archive data associated with his/her science program for the proprietary period.

Proprietary Period — The period in which data in the GSA is not released to the general public. From the point of view of the GSA operations, the initial public release date for each dataset is set by Gemini, and may be changed by Gemini at any time.

Prototype — An implementation of the functionality of some aspect of the GSA prior to the delivery of the fully finished system. The purpose of the prototype is to clarify an aspect of the design.

Queue Observing — In this mode of observing, an astronomer submits a science program to an observatory, and the science program is added to a list of science programs waiting to be executed. The execution of the science program is overseen by observatory staff. The scheduling of the program execution can by based on many factors, including time in the list, current observing conditions, and the conditions required by the science program. Queue mode observing also requires that the science program consists of planned observations. See also Classical Observing, Remote Observing, and Service Observing.

Recipe — See Data Processing Recipe.

Remote Observing — This is similar to Classical observing, except that the Observer oversees the execution of the science program from a remote site (usually their home institution). See also Classical Observing, Queue Observing, and Service Observing.

Rock Ridge — An extension to ISO 9660 to allow Unix-specific information such as links, file attributes and native Unix naming to be stored along with the ISO 9660 information.

ROSAT — ROSAT is an X-ray space telescope which went into operation in June 1990. As the primary objective ROSAT has performed in a scan mode the first all-sky survey with imaging telescopes in the soft X-ray band of 0.1 keV - 2 keV (corresponding to wavelengths of 100 Å - 6 Å) as well as in the adjacent extreme ultraviolet region of 0.04 keV - 0.2 keV (corresponding to wavelengths of 300 Å- 60 Å). More than 60,000 X-ray sources have been detected with the ROSAT all-sky survey, larger by almost two orders of magnitude than the 840 sources of the catalogue of the previously largest all-sky survey of the HEAO-I satellite.

Science program — A formal description of an observer's plan for using a Gemini Telescope. Science programs are produced by the Gemini Observing Tool. A science program is suitable for near-automatic execution. It usually consists of an unordered group of Observations.

Server — Any computer program which carries out some form of service for one or more client computer programs. Typically, a server will be on a remote machine, and therefore it has direct access to resources not directly available to client programs. A server is completely subservient to its clients. It responds to commands from the clients but does not initiate dialogue with clients.

Service Observing — This is similar to queue observing, in the sense that the execution of a science program is overseen by a member of the observatory staff, instead of one of the a members on the science team who submitted the science program. Service observing

also requires that the science program consists of planned observations. See also Classical Observing, Queue Observing, and Remote Observing.

SIMBAD — The SIMBAD astronomical database provides basic data, cross-identifications and bibliography for astronomical objects outside the solar system. SIMBAD can be queried by object name, coordinates and other criteria (filters).

Status — Attributes which define the state of a hardware or software system.

STIS — An HST instrument for both imaging and spectroscopy.

Structured Query Language (SQL) — A standard language, common in commercial database management systems, used for specifying how to search a relational database for items matching certain criteria. The criteria can involve arithmetic and logical combinations of record properties (tuples).

Telescope — In the Gemini software documentation the word "Telescope" is used to refer to the Gemini 8m telescope together with all of its associated peripheral devices. These include the telescope mount, the primary and secondary mirrors, together with their support and control structures, the Cassegrain rotator, and all the devices associated with acquisition and guiding and with the adaptive optics. Some of the latter devices can also be regarded as instruments.

Transportable media — Media that will contain the results of an observation program for transport to the observers home facility, and to the archive centre. This is base-lined to be either DAT or EXABYTE tapes, or CD-ROM.

Universal Resource Locator (URL) — The unique address of an internet resource, such as a hypertext document, gopher page, mail address or a computer. The address adheres to a communications protocol described in RFC1738 (http://www.cis.ohio-state.edu/htbin/rfc/rfc1738.html).

Unix — A standard host-level operating system. The Gemini baseline for "Unix" is "Solaris" marketed by Sun Microsystems.

Usability Requirement — This is a requirement which does not enable any new functionality, but which make the GSA easier to use.

Visiting instrument — A self-contained **"Instrument"** which is brought to the Gemini telescope for temporary use. This type of instrument is intended only for the use of the instrument owner.

WFPC2 — An HST imaging instrument.

Work package — A well-defined unit of work, described by a workscope document. A work package usually consists of the development of a subsystem or a layer of utilities, which can be devolved to a collaborating group.

World Coordinate System (WCS) — A mapping of the pixel column/row (x/y) coordinate system of an image into a coordinate system that has meaning in the real world. In astronomy, the would coordinate system is most often right ascension and declination, but could also include other coordinate systems, such as galactic coordinates. A WCS could also be associated with a spectral dataset, where one of the world coordinates would be wavelength.

World Wide Web — A part of the internet designed to allow easy navigation of the network of computers through the use of GUI's and hypertext links between different URL's or addresses.

5. References

[1] *Engineering Archive Requirements*, Phil Puxley, Dayle Kotturi, Jim Oschmann, Bret Goodrich, Gemini, Daniel Durand, David Schade, David Bohlender, Séverin Gaudet, HIA, http://www.gemini.edu/documentation/webdocs/spe/spe-c-g0073v02.pdf

[2] *ESA Guide to the Software Requirements Definition Phase*, ESA PSS-05-03, ESA Board for Software Standardisation and Control

[3] Galfit

[4] *Gemini Archive Position Paper*, Andy Woodsworth, January 3, 1992, http://www.hia.nrc.ca/pub/Gemini_HIA/GSA/References/cadcPosition.pdf

[5] *The Gemini Science Archive: A Proposal by the Canadian Astronomy Data Centre*, Dennis R. Crabtree, September, 1995, http://www.hia.nrc.ca/pub/Gemini_HIA/GSA/References/proposal.pdf

[6] *Gemini Science Archive Conceptual Design Study Work Scope* No. 9414257-GEM02012, http://www.hia.nrc.ca/pub/Gemini_HIA/GSA/Workscope/workscope.pdf

[7] *Gemini Science Requirements*, SPE-PS-G001, Version 2., http://www.gemini.edu/documentation/webdocs/spe/spe-ps-g0001.pdf

[8] *GSA Operational Concept Definition Document*, gsa_OCDD/04, Norman Hill, Séverin Gaudet, Daniel Durand, David Schade, David Bohlender, NRC, Felipe Barrientos, Felipe Richardson CONICYT, http://www.hia.nrc.ca/pub/Gemini_HIA/GSA/OCDD/OCDD_initial.pdf

[9] *Recommendations of Science Archive Workshop*, http://www.gemini.edu/sciops/data/scia-arch-rec.html

[10] *The Scientific Case For a Gemini Data Archive*, David Schade, Daniel Durand, Jean-René Roy, Maria-Teresa Ruiz, http://www.hia.nrc.ca/pub/Gemini_HIA/GSA/References/science_case.pdf

[11] SExtractor Online documentation, Emmanuel Bertin, http://terapix.iap.fr/sextractor/

[12] *Summary of Science Archive Workshop Presentations*, http://www.gemini.edu/sciops/data/scia-arch-wkshop.html

Created:February 7, 2000
Modified:November 21, 2000

Gemini Science Archive

Chapter 2

General Description

This chapter is a copy of chapter to of the GSA OCDD document.

1. History

The Gemini Science Archive was originally proposed in a position paper [4] written by the CADC in early 1992. This paper described the value of a Gemini Archive, and described how it could be implemented. This was followed in September 1995 by a proposal [5] by the CADC to design, implement, and operate a Gemini Science Archive based on the existing archive infrastructure maintained at the CADC.

At its April 1998 meeting the Gemini Science Committee (GSC) reviewed the science cases for a Gemini archive presented in [10], and endorsed the following resolution concerning a Science Archive for Gemini:

> *"The GSC thinks a Gemini science archive is scientifically compelling. Such an archive would provide the scientific community with tools for effective on-line access to all Gemini science data and supporting information in order to promote further scientific exploitation of those data. The Gemini Science Archive should guarantee that the valuable datasets obtained with the Gemini Telescopes are usable by future generations for research and education. The Project should further study the implementation and the resources required for a Gemini science archive."*

The Gemini Board has endorsed the GSC views on a Gemini Science Archive, and agreed that the Project should study the specifications and requirements for a science archive

In order to provide a basis for the science archive conceptual design activities, Gemini sponsored a workshop to assess the feasibility of a viable Gemini Science Archive and recommend prioritized specifications and requirements for the science archive. The Science Archive Workshop was held on Monday 14 September 1998, in the Gemini Hilo Base Facility, Hilo, Hawaii. The workshop recommendations can be found in [9]. A summary of the workshop presentations can be found in [12].

The workshop recommendations described in [9], formed the basis of the Gemini Science Archive Conceptual Design Study work scope [6], and this document has been produced as part of that work scope.

2. Relation to Current Projects

2.1 CADC Archive Development

The CADC currently supports archives for several telescopes, including the HST, the CFHT, JCMT, and several others. It is expected that the development of the GSA by the CADC will have

benefits to these other archives, and that development efforts for these other archives will in turn benefit the GSA. In particular, the GSA may benefit from the current development efforts at the CADC, aimed at creating an astronomical data mining system.

The other archives at the CADC will also be an advantage in attempting to initiate the unified catalogue required for multi-archive and cross archive searches. The CADC can represent the interests of all of the archives supported by the CADC, including the GSA, which we believe will increase the chances of creating a successful unified archive. It is also possible to create a very useful unified archive consisting only of the archives supported by the CADC.

2.2 The Gemini On-line Data Processing System

The Gemini Telescopes currently plan to implement a On-Line Data Processing system (OLDP) to be used for near real-time evaluation of data quality during telescope operations. We hope that the GSA will benefit from the existence of this data processing system by re-using the data processing recipes, and possibly the data processing infrastructure that must be developed to support the OLDP. It is also feasible that the implementation of the OLDP could be done as a shared task with the GSA development.

3. Relation to Predecessor and Successor Projects

3.1 The CADC Archive Infrastructure

The GSA will build on the software, hardware, and staffing infrastructure created by the CADC for existing archives. The experience gained, lessons learned and software developed during the development of these archives will be applied to the development of the GSA.

3.2 The Gemini Engineering Archive

The Gemini Telescopes have currently implemented a Gemini Engineering Archive (GEA), which records much of the data required by the GSA. The existence of the Engineering archive will simplify the task of gathering data required for the GSA.

3.3 The Gemini Data Handling System

As part of a separate work package, the CADC developed the Data Handling System (DHS) for the Gemini telescopes. As part of the development of the DHS, the CADC developed an intimate knowledge of the data handling infrastructure of the Gemini telescopes, and considerable knowledge of other elements of the Gemini telescopes. This knowledge will be extremely valuable in the design and implementation of the GSA.

3.4 Ongoing development of the GSA

After the initial design and implementation of the GSA is complete, an ongoing development process should be instituted as a part of the GSA operations. This ongoing development will allow advanced capabilities to be added to the GSA. These advanced capabilities may be capabilities which cannot be implemented within the initial GSA development budget, or they may be capabilities that are not yet identified.

4. Function and Purpose

The Gemini Science Archive should provide the scientific community of the partner countries with on-line access to all Gemini scientific observations to ensure full scientific exploitation of those data. The Gemini Science Archive should be considered an integrated part of the planning, observation, calibration, data reduction, and data distribution processes that occur at Gemini. The Gemini Science Archive should guarantee that the valuable datasets obtained with the Gemini Telescopes are preserved for use by future generations for research and education.

A well-designed and properly implemented Science Data Archive, as part of the capabilities of the Gemini observatories, would be a major contribution toward the full exploitation of the unique characteristics of the Gemini telescopes. An effective archive would boost scientific productivity and would ensure that maximum value was extracted from the expensive-to-obtain observational data. Archiving requirements are consistent with, and would help to optimise, queue mode observing requirements. In the long term, a science archive would enable many of the high level science goals outlined in the Gemini Science Requirements document [7]. Some specific goals— for example, understanding the relationship between quasars and their host galaxies, probing variable phenomena in stars and galaxies, dynamics of galactic nuclei, would be unattainable without archival access to the complete set of Gemini observations. The collective impact of the full Gemini Science Archive would far exceed what could be produced by the programs of single observers or teams of observers. It would span a wider range in physically important parameters such as redshift and luminosity of active galactic nuclei or galaxies. Furthermore, many high-priority Gemini projects would use the Gemini data in conjunction with Hubble Space Telescope, HIPPARCOS, ROSAT, IRAS, NGST and other archival data.

Access to Gemini data by the wider community, after the proprietary period, would ensure that the data are fully exploited and that maximum scientific value is returned to the astronomical communities and, ultimately, the citizens that support Gemini. The true legacy of the Gemini observatories, in addition to the initial discoveries would be the collection of excellent observations produced by its innovative instruments and this legacy, in the form of the **first scientifically effective archive of ground-based observations,** would play an important role in astrophysical research long after Gemini itself ceases operations.

5. Mission Statement

Following is the mission statement for the GSA:

The Gemini Science Archive should provide the scientific community with tools for effective on-line access to all Gemini science data and supporting information in order to promote further scientific exploitation of those data. The Gemini Science Archive should guarantee that the valuable datasets obtained with the Gemini Telescopes are usable by future generations for research and education.

6. User Characteristics

The users of the GSA are primarily, but not limited to, astronomers from the Gemini member countries. It is expected that the GSA users will use the archive for the following general purposes:

- Applying new techniques to re-analyse data obtained by the original observer.

- Creating sets of data larger than could be practically created with PI observations.
- Verification of the findings of the original observer.
- Augmenting original observations collected with either the Gemini Telescopes, or with other telescopes.
- Evaluation of the performance and capabilities of Gemini Telescopes and instruments by potential users.
- Planning for future observations of the same or similar fields of view.
- Creating new datasets by spanning observing runs.
- Examining historical behaviour of time varying phenomena.
- Proposal preparation for PI observations at Gemini and other telescopes.
- As a teaching or learning tool.

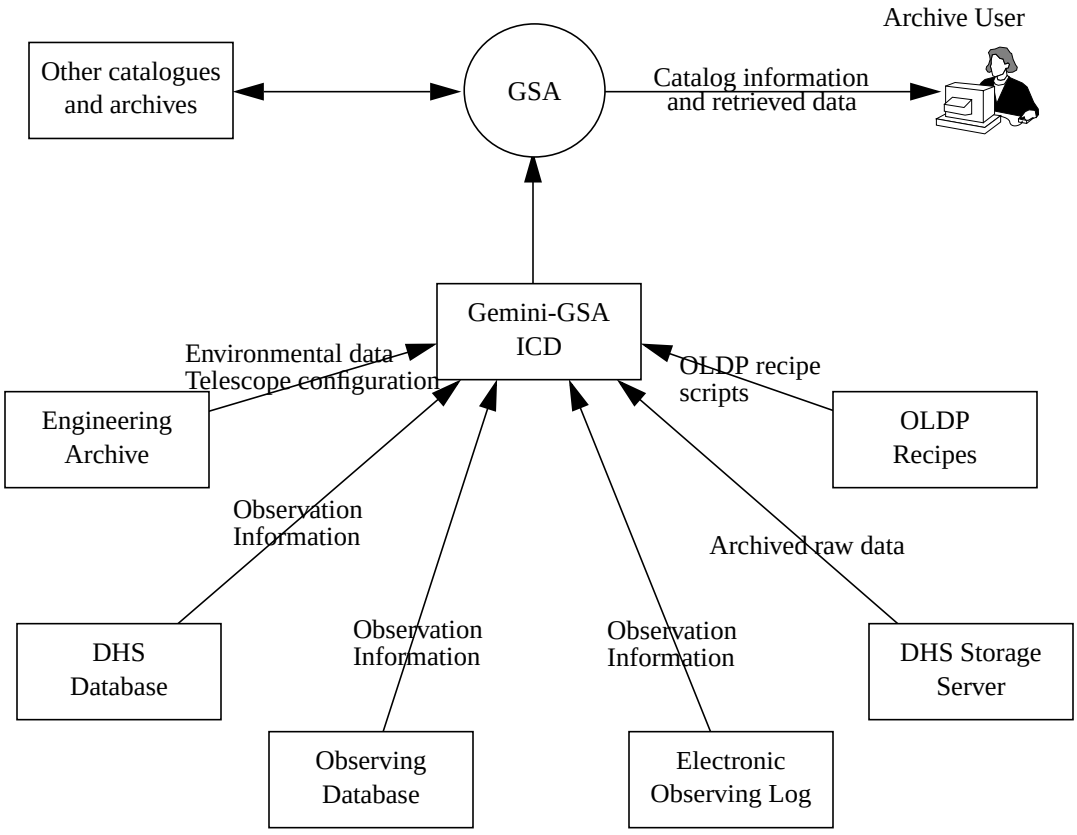
7. Relation to Other Systems

The GSA relies on data from a variety of other Gemini systems as input, and will therefore need interfaces to these systems. These systems include:

- The Gemini DHS Storage Server will produce removable media containing the raw datasets which will be used by the GSA.
- The Gemini Engineering Archive will record telescope activity and environmental data which will become meta-data for the raw datasets, and which will help identify calibration data associated with the raw data.
- The Gemini DHS database will provide an independent list of raw datasets which will be used to verify that all expected data is received on removable media.
- The Gemini Observing Database will provide information about the structure of the observations, the data processing done, and the public information from the original proposal.
- The Gemini Electronic observing log will describe the actions of the observer during the observing run, which will become meta-data for the raw data.
- Data processing recipes and rules for identifying calibration data, used by Gemini for data quality analysis will be the basis for the GSA automatic data processing.

The interface to each of these data sources will be described in an ICD. The context diagram for the GSA is shown in Figure 1 on page 15.

FIGURE 1. GSA Context Diagram



Gemini Science Archive

Chapter 3 Functional and Performance Requirements

1. Introduction

This chapter describes the software requirements for the GSA. These requirements are derived from the user requirements listed in the GSA OCDD [8], and from the requirements listed in the GSA work scope[6].

These software requirements will be used as the basis of the next phase of the development of the GSA, which will be the GSA to Gemini ICD, and the GSA Conceptional Design Document.

The software requirements have been placed into several different subsections based on the category of the requirement:

- Section 2.1.1 on page 40 lists the general requirements for the GSA.
- Section 2.1.3 on page 49 lists the requirements on the GSA user interface.
- Section 2.2 on page 58 shows the data input rates that the GSA must accept from the telescopes, and the data output rates that the GSA must be able to supply to users.
- Section 2.3 on page 60 list the requirements that affect GSA operations, but which will not affect the design or implementation of the GSA.
- Section 2.4 on page 61 lists the resource requirements on the GSA specified by the Gemini project.
- Section 2.5 on page 62 lists the requirements that Gemini must meet in order for the GSA to function properly.
- Section 2.6 on page 66 list external requirements, outside the scope of both the GSA and the Gemini project.

Note that some of the requirements listed in this documents may have to be dropped from the GSA implementation, either a) to keep the GSA within the available budget, b) because implementation of the requirement within the GSA is impractical, or c) because the impact of the requirement on Gemini operations is unacceptable. If requirements are to be dropped, this document, in combination with [8] will provide the basis for predicting the impact of the requirement on the functionality of the GSA. The user requirements listed in the work scope [6] and the OCDD [8], which have not been incorporated into the software requirements are listed in Table 1 on page 18.

Table 2 on page 18 and Table 3 on page 29 provide a cross reference between the software requirements and their sources in [6] and [8]. The source requirement descriptions used in the tables are the short descriptions of the requirements in [8], or a condensed version of the requirements given in [6].

TABLE 1. User requirements not included in software requirements

UR1.18 Reject data supersets based on field selection criteria. This requirement is considered to be unattainable.
[6] Section 7.2.1 (h) Provide software and hardware for browsing and retrieval. This is a very vague requirement which can be applied to all of the software requirements.
[6] Section 7.2.2 (d) through (f). These requirement numbers are missing from the work scope, apparently due to a typo in the list numbering

TABLE 2. Software requirement to source cross reference

SR1.1 The GSA shall archive all data. [6] Section 7.2.1 (a) Provide secure data storage. [6] Section 7.2.1 (g) Archive all science data. [6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
SR1.2 The GSA shall respect proprietary periods assigned by Gemini. [6] Section 7.2.1 (b) Respect proprietary periods. UR2.2 Automatic re-processing of data should be an option
SR1.3 The GSA shall be able to automatically process data. [6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode. [6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing. [6] Section 7.2.2 (b) Provide on-the-fly re-calibration. [6] Section 7.2.2 (h) Provide data mining capabilities. UR1.13 Display object parameters UR2.2 Automatic re-processing of data should be an option UR2.3 Automatic re-processing inputs and options should be documented UR2.4 Users should be able to choose one or more processing recipes UR2.6 GSA users may preview data for selected data supersets UR2.9 GSA catalogue needs descriptors derived from the data UR2.10 The catalogue should contain information on associated data
SR1.3.1 Data processing shall respect Gemini policy on proprietary periods. [6] Section 7.2.1 (b) Respect proprietary periods.
SR1.3.2 Data processing shall proceed promptly. UR1.19 Data requested by users should be available quickly UR1.20 The GSA should be available 24 hours a day
SR1.4 Preview data should be generated for each dataset. [6] Section 7.2.1 (b) Respect proprietary periods. [6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode. [6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing. [6] Section 7.2.2 (g) Provide visualization and plotting tools. UR2.6 GSA users may preview data for selected data supersets UR2.7 Preview must be display promptly
SR1.4.1 Generate previews for associations. [6] Section 7.2.1 (b) Respect proprietary periods. [6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode. [6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.

TABLE 2. Software requirement to source cross reference

[6] Section 7.2.2 (g) Provide visualization and plotting tools.
UR2.6 GSA users may preview data for selected data supersets
UR2.7 Preview must be display promptly
SR1.4.2 Remove previews when release date is changed.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue..
UR2.10 The catalogue should contain information on associated data
SR1.4.3 Preview should be mirrorable.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue..
SR1.5 Collect information needed to track of publications based on archive data.
UR1.1 The archive data must be fully described
UR2.10 The catalogue should contain information on associated data
SR2.1 The GSA catalogue shall contain all data descriptors.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
UR1.1 The archive data must be fully described
SR2.2 The catalogue shall include environmental information.
[6] Section 7.2.1 (e) Provide atmospheric monitoring data.
UR1.1 The archive data must be fully described
SR2.3 The catalogue shall include accurate WCS information.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
[6] Section 7.2.1 (m) Provide catalogue search capabilities.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.1 The archive data must be fully described
UR1.4 Select sets of data supersets based on coordinate
UR1.12 Select data supersets based on object parameter
UR1.17 Support unification of astronomical catalogues
UR1.21 The GSA should have an object catalogue
UR2.14 It should be possible to plot the fields of view of a set of data supersets
UR2.15 Select a set of data supersets based on spatial density
UR2.17 Search based on a list of coordinates
SR2.4 Proposal information should be associated with data supersets.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR1.14 Proposal information for public data must be available
UR1.15 Proposal information must be searchable
SR2.5 The catalogue should include image descriptors derived from the data.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
UR2.9 GSA catalogue needs descriptors derived from the data

TABLE 2. Software requirement to source cross reference

SR2.5.1 Remove derived image descriptors for proprietary data.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR2.9 GSA catalogue needs descriptors derived from the data
SR2.6 Release date is a data descriptor.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
UR1.1 The archive data must be fully described
UR2.10 The catalogue should contain information on associated data
SR2.7 The GSA catalogue should contain associated datasets.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR1.3 Select sets of the data supersets in the GSA
UR1.4 Select sets of data supersets based on coordinate
UR1.5 The GSA must display data superset descriptors to the user
UR1.9 GSA users may select sets of data supersets for retrieval
UR1.12 Select data supersets based on object parameter
UR2.2 Automatic re-processing of data should be an option
UR2.6 GSA users may preview data for selected data supersets
UR2.10 The catalogue should contain information on associated data
UR2.17 Search based on a list of coordinates
SR2.8 Data quality assessment shall be in the catalogue.
[6] Section 7.2.1 (j) Provide data quality assessment for each dataset..
UR2.16 Data quality indicator must be one of the data descriptors
SR2.9 The science program shall be associated with the catalogue data.
UR2.19 View and retrieve the science program
SR2.10 The electronic observing log shall be associated with the catalogue.
[6] Section 7.2.1 (d) Provide electronic observing log.
UR1.1 The archive data must be fully described
SR2.11 Proposal id shall be a descriptor.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
UR1.14 Proposal information for public data must be available
UR1.16 Search for data based on proposal
SR2.12 The GSA catalogue shall be mirrorable.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
UR2.6 GSA users may preview data for selected data supersets
UR2.10 The catalogue should contain information on associated data
UR2.13 Publications should be linked to the data
SR2.13 Calibration data shall be associated with science data.
[6] Section 7.2.1 (f) Ensure calibration material is available.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.

TABLE 2. Software requirement to source cross reference

[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.1 The archive data must be fully described
UR1.10 Calibration data must be easily available
UR2.2 Automatic re-processing of data should be an option
UR2.6 GSA users may preview data for selected data supersets
UR2.9 GSA catalogue needs descriptors derived from the data
UR2.10 The catalogue should contain information on associated data
SR2.14 The GSA should have an object catalogue.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.12 Select data supersets based on object parameter
UR1.13 Display object parameters
UR1.21 The GSA should have an object catalogue
UR2.10 The catalogue should contain information on associated data
SR2.14.1 Object parameters should be extracted from object catalogues.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.12 Select data supersets based on object parameter
UR1.13 Display object parameters
SR2.14.2 The GSA object catalogue should be up to date.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.13 Display object parameters
SR2.14.3 The object catalogue should include object descriptors derived from data.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.13 Display object parameters
SR2.14.4 Remove derived objects for proprietary data.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
[6] Section 7.2.2 (h) Provide data mining capabilities.
SR2.15 Publications should be linked to data supersets.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
UR1.1 The archive data must be fully described
UR2.10 The catalogue should contain information on associated data
UR2.13 Publications should be linked to the data
SR2.16 The GSA shall store Gemini hardware and software versions.
UR1.1 The archive data must be fully described

TABLE 2. Software requirement to source cross reference

SR3.1 The GSA shall have an interactive user interface.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR2.12 Access to the GSA catalogue must be interactive
SR3.2 Select subsets of the GSA catalogue.
[6] Section 7.2.1 (m) Provide catalogue search capabilities.
UR1.3 Select sets of the data supersets in the GSA
UR1.9 GSA users may select sets of data supersets for retrieval
UR2.9 GSA catalogue needs descriptors derived from the data
UR2.10 The catalogue should contain information on associated data
UR2.11 A single query needs to address any or all instruments
SR3.2.1 Select subsets of the GSA catalogue based on coordinate region search.
UR1.4 Select sets of data supersets based on coordinate
UR2.17 Search based on a list of coordinates
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.12 Select data supersets based on object parameter
UR1.21 The GSA should have an object catalogue
SR3.2.3 Support cross catalogue searches.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
UR1.17 Support unification of astronomical catalogues
SR3.2.4 The GSA should support a batch mode interface.
UR2.21 The GSA should support a batch mode interface
UR2.22 The GSA should be accessible to users with poor internet connections
SR3.3 View descriptors for a subset of the catalogue.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR1.1 The archive data must be fully described
UR1.5 The GSA must display data superset descriptors to the user
UR2.9 GSA catalogue needs descriptors derived from the data
UR2.10 The catalogue should contain information on associated data
UR2.12 Access to the GSA catalogue must be interactive
SR3.3.1 There shall be on-line help describing all catalogue data descriptors.
UR1.2 There should be meta data describing the data descriptors
UR1.13 Display object parameters
UR1.21 The GSA should have an object catalogue
UR2.9 GSA catalogue needs descriptors derived from the data
SR3.3.2 GSA users can view a list of calibration data supersets.
UR1.10 Calibration data must be easily available
SR3.3.3 View the list of publications associated with a data superset.
UR2.10 The catalogue should contain information on associated data
UR2.13 Publications should be linked to the data

TABLE 2. Software requirement to source cross reference

SR3.3.4 View descriptors of objects in a data superset field of view.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
UR1.13 Display object parameters
UR1.21 The GSA should have an object catalogue
UR2.10 The catalogue should contain information on associated data
SR3.3.5 View proposal information associated with a data superset.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR1.14 Proposal information for public data must be available
UR2.10 The catalogue should contain information on associated data
SR3.3.6 Sort the displayed subset of the catalogue.
UR2.8 Sort the displayed set of data supersets
SR3.3.7 Sort the displayed objects.
UR1.21 The GSA should have an object catalogue
UR2.8 Sort the displayed set of data supersets
SR3.3.8 Plot a subset of data supersets in spatial coordinates.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
UR2.14 It should be possible to plot the fields of view of a set of data supersets
SR3.3.9 GSA users shall be able to view the electronic observing log.
[6] Section 7.2.1 (d) Provide electronic observing log.
[6] Section 7.2.1 (f) Ensure calibration material is available..
UR2.10 The catalogue should contain information on associated data
SR3.3.10 View environmental data.
[6] Section 7.2.1 (e) Provide atmospheric monitoring data.
UR2.10 The catalogue should contain information on associated data
SR3.3.11 Display science programs.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR2.10 The catalogue should contain information on associated data
UR2.19 View and retrieve the science program
SR3.3.12 Display the members of association data supersets.
UR1.1 The archive data must be fully described
UR2.10 The catalogue should contain information on associated data
SR3.3.13 Save a set of data supersets to a file.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
UR2.15 Select a set of data supersets based on spatial density
SR3.3.14 Load a set of data supersets from a file.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
UR2.15 Select a set of data supersets based on spatial density
SR3.4 GSA users shall be able to select data for retrieval.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
UR1.9 GSA users may select sets of data supersets for retrieval
UR2.12 Access to the GSA catalogue must be interactive
SR3.4.1 Users shall retrieve data via Internet, EXABYTE, or CD-ROM.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
UR1.9 GSA users may select sets of data supersets for retrieval

TABLE 2. Software requirement to source cross reference

SR3.4.2 Only authorized users shall be able to retrieve proprietary data.
[6] Section 7.2.1 (a) Provide secure data storage.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR1.11 Retrieve proprietary data
UR2.2 Automatic re-processing of data should be an option
SR3.4.3 GSA users should be able to choose to have data processed.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR2.2 Automatic re-processing of data should be an option
UR2.4 Users should be able to choose one or more processing recipes
SR3.4.4 Data processing shall not use proprietary data as input.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR2.2 Automatic re-processing of data should be an option
UR2.4 Users should be able to choose one or more processing recipes
SR3.4.5 Processing recipes shall be documented.
UR2.3 Automatic re-processing inputs and options should be documented
UR2.4 Users should be able to choose one or more processing recipes
UR2.5 Information about re-processing recipes should be available
SR3.4.6 Authorized Gemini staff can retrieve proprietary data.
[6] Section 7.2.1 (a) Provide secure data storage.
[6] Section 7.2.1 (b) Respect proprietary periods.
Email message from F.G.?
SR3.4.7 GSA users can retrieve calibration data.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (f) Ensure calibration material is available.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
UR1.10 Calibration data must be easily available
SR3.4.8 Retrieve environmental data.
[6] Section 7.2.1 (e) Provide atmospheric monitoring data.
UR1.10 Calibration data must be easily available
SR3.4.9 It shall be possible to retrieve science programs.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR2.19 View and retrieve the science program
SR3.4.10 It shall be possible to retrieve proposal information.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR1.14 Proposal information for public data must be available
SR3.4.11 It shall be possible to retrieve the electronic observing log.
[6] Section 7.2.1 (d) Provide electronic observing log.
[6] Section 7.2.1 (f) Ensure calibration material is available.
UR1.10 Calibration data must be easily available
SR3.4.12 It should be possible for users to monitor the progress of data requests.
UR2.18 Monitor progress of data requests

TABLE 2. Software requirement to source cross reference

SR3.4.13 Batch mode requests.
UR2.21 The GSA should support a batch mode interface
UR2.22 The GSA should be accessible to users with poor internet connections
SR3.4.14 Document user data requests.
UR2.20 The content of user requests should be documented
SR3.4.15 Retrieve associated data supersets.
UR2.10 The catalogue should contain information on associated data
SR3.4.16 Provide data processing recipes for associations.
UR2.10 The catalogue should contain information on associated data
SR3.5 A general description of the content of the GSA shall be available.
UR1.6 A general description of the GSA must be available to users
SR3.5.1 A description of the Gemini instruments shall be available.
UR1.7 A description of the instruments must be available
SR3.5.2 A description of the Gemini facilities shall be available.
UR1.8 A description of Gemini facilities must be available
SR3.6 Proposal information should be searchable.
[6] Section 7.2.1 (m) Provide catalogue search capabilities.
UR1.15 Proposal information must be searchable
UR1.16 Search for data based on proposal
SR3.6.1 Find all data supersets associated with proposals.
UR1.16 Search for data based on proposal
SR3.7 Preview display for each science data superset.
[6] Section 7.2.1 (m) Provide catalogue search capabilities.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
UR2.6 GSA users may preview data for selected data supersets
UR2.10 The catalogue should contain information on associated data
SR4.1 Maximum and average requirements for catalogue ingest.
[6] Section 7.2.1 (g) Archive all science data.
SR4.2 Maximum and average requirements for raw data ingest.
[6] Section 7.2.1 (g) Archive all science data.
SR4.3 Maximum and average requirements for data retrieval by users.
UR1.9 GSA users may select sets of data supersets for retrieval
UR1.19 Data requested by users should be available quickly
UR2.2 Automatic re-processing of data should be an option
SR4.4 Preview should be display promptly.
UR2.6 GSA users may preview data for selected data supersets
UR2.7 Preview must be display promptly
SR4.5 Access to the GSA catalogue must be interactive.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
UR1.3 Select sets of the data supersets in the GSA
UR1.4 Select sets of data supersets based on coordinate
UR1.5 The GSA must display data superset descriptors to the user
UR1.12 Select data supersets based on object parameter
UR2.12 Access to the GSA catalogue must be interactive
UR2.17 Search based on a list of coordinates

TABLE 2. Software requirement to source cross reference

SR4.6 Data for Internet retrieval should be available promptly.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR1.9 GSA users may select sets of data supersets for retrieval
UR1.19 Data requested by users should be available quickly
UR1.20 The GSA should be available 24 hours a day
UR2.2 Automatic re-processing of data should be an option
UR2.4 Users should be able to choose one or more processing recipes
SR5.1 Participate in catalogue unification efforts.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.17 Support unification of astronomical catalogues
UR2.1 Support multi archive searches
UR3.1 Other astronomy archives must participate in unification efforts
SR5.2 Data shall be physically secure.
[6] Section 7.2.1 (a) Provide secure data storage.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems..
SR5.3 Data shall be electronically secure.
[6] Section 7.2.1 (a) Provide secure data storage.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
UR1.9 GSA users may select sets of data supersets for retrieval
UR1.11 Retrieve proprietary data
SR5.4 GSA staff shall respect proprietary periods.
[6] Section 7.2.1 (b) Respect proprietary periods.
SR5.5 GSA operations staff shall support GSA users.
UR1.22 The GSA staff must provide support to GSA users
SR5.6 GSA operations staff shall support ongoing development of Gemini.
[6] Section 7.2.1 (a) Provide secure data storage.
SR5.7 GSA operations staff shall incorporate evolving technologies into the GSA.
[6] Section 7.2.1 (a) Provide secure data storage.
SR5.8 The GSA operations staff should add new features to the GSA.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
[6] Section 7.2.2 (h) Provide data mining capabilities.
SR6.1 GSA development cost.
[6] Section 7.2 Requirements for Science Archive.
SR6.2 GSA operating cost.
[6] Section 7.2 Requirements for Science Archive.
[6] Section 8.7 (a) (6) A firm proposal for operating the Science Archive once it is complete.

TABLE 2. Software requirement to source cross reference

SR7.1 Gemini shall collect instrument configuration descriptors.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (f) Ensure calibration material is available.
UR1.1 The archive data must be fully described
SR7.2 Gemini shall collect telescope configuration descriptors.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (f) Ensure calibration material is available.
UR1.1 The archive data must be fully described
SR7.3 Gemini shall record weather information.
[6] Section 7.2.1 (e) Provide atmospheric monitoring data.
UR1.1 The archive data must be fully described
SR7.4 Gemini shall collect accurate WCS information.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR1.1 The archive data must be fully described
SR7.5 Gemini shall make proposal information available to the GSA.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR1.1 The archive data must be fully described
UR1.14 Proposal information for public data must be available
UR1.15 Proposal information must be searchable
SR7.5.1 Gemini shall provide passwords for executed proposals.
[6] Section 7.2.1 (b) Respect proprietary periods.
UR1.11 Retrieve proprietary data
SR7.5.2 Gemini shall provide information to link proposals with datasets.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR1.14 Proposal information for public data must be available
UR1.16 Search for data based on proposal
SR7.6 Gemini shall provide a data quality assessment of each dataset.
[6] Section 7.2.1 (j) Provide data quality assessment for each dataset.
UR1.1 The archive data must be fully described
UR2.16 Data quality indicator must be one of the data descriptors
SR7.7 Gemini should track publications associated with data.
UR2.13 Publications should be linked to the data
SR7.8 Gemini shall provide an electronic observing log.
[6] Section 7.2.1 (d) Provide electronic observing log.
SR7.9 Supply executed science programs to the GSA.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR2.19 View and retrieve the science program
SR7.10 Gemini shall provide a general description of the telescope.
UR1.6 A general description of the GSA must be available to users
SR7.10.1 Gemini shall provide a description of the instruments.
UR1.7 A description of the instruments must be available
SR7.10.2 Gemini shall provide a description of Gemini facilities.
UR1.8 A description of Gemini facilities must be available
SR7.11 Gemini shall provide descriptions of data descriptors.
UR1.2 There should be meta data describing the data descriptors

TABLE 2. Software requirement to source cross reference

SR7.12 Gemini shall provide raw science data to the GSA.
[6] Section 7.2.1 (g) Archive all science data.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
SR7.13 Gemini shall ensure that calibration data is available.
[6] Section 7.2.1 (f) Ensure calibration material is available.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
UR1.10 Calibration data must be easily available
SR7.13.1 Gemini shall provide information to identify calibration data.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (f) Ensure calibration material is available.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR1.1 The archive data must be fully described
UR1.10 Calibration data must be easily available
UR2.2 Automatic re-processing of data should be an option
SR7.14 Gemini should provide data processing recipes.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR2.2 Automatic re-processing of data should be an option
UR2.4 Users should be able to choose one or more processing recipes
UR2.6 GSA users may preview data for selected data supersets
SR7.14.1 Gemini should provide descriptions of data processing recipes.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
UR2.3 Automatic re-processing inputs and options should be documented
UR2.4 Users should be able to choose one or more processing recipes
UR2.5 Information about re-processing recipes should be available
SR7.15 Gemini shall provide a release date for each dataset.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
UR1.1 The archive data must be fully described
UR1.11 Retrieve proprietary data
SR7.15.1 Gemini shall inform the GSA when release dates change.
[6] Section 7.2.1 (b) Respect proprietary periods.
[6] Section 7.2.1 (c) Provide access to data descriptors.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
UR1.1 The archive data must be fully described
UR1.11 Retrieve proprietary data
SR7.15.2 Inform the GSA which Gemini staff are permitted to access proprietary data.
UR1.11 Retrieve proprietary data
Email from F.G.?

TABLE 2. Software requirement to source cross reference

SR7.16 Gemini shall provide support for GSA users.
UR1.22 The GSA staff must provide support to GSA users
SR7.17 Gemini shall provide version information.
[6] Section 7.2.1 (c) Provide access to data descriptors.
UR1.1 The archive data must be fully described
SR7.18 Gemini shall provide test data.
[6] Section 7.2.1 (g) Archive all science data.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.12 Select data supersets based on object parameter
UR1.13 Display object parameters
UR1.21 The GSA should have an object catalogue
SR8.2 A unified astronomical archive catalogue should be created.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
[6] Section 7.2.2 (h) Provide data mining capabilities.
UR1.17 Support unification of astronomical catalogues
UR2.1 Support multi archive searches
UR3.1 Other astronomy archives must participate in unification efforts

TABLE 3. Source to software requirement cross reference

[6] Section 7.2 Requirements for Science Archive.
SR6.1 GSA development cost.
SR6.2 GSA operating cost.
[6] Section 7.2.1 (a) Provide secure data storage.
SR1.1 The GSA shall archive all data.
SR3.4.2 Only authorized users shall be able to retrieve proprietary data.
SR3.4.6 Authorized Gemini staff can retrieve proprietary data.
SR5.2 Data shall be physically secure.
SR5.3 Data shall be electronically secure.
SR5.6 GSA operations staff shall support ongoing development of Gemini.
SR5.7 GSA operations staff shall incorporate evolving technologies into the GSA.
[6] Section 7.2.1 (b) Respect proprietary periods.
SR1.2 The GSA shall respect proprietary periods assigned by Gemini.
SR1.3.1 Data processing shall respect Gemini policy on proprietary periods.
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR1.4.2 Remove previews when release date is changed.
SR2.4 Proposal information should be associated with data supersets.
SR2.5 The catalogue should include image descriptors derived from the data.
SR2.5.1 Remove derived image descriptors for proprietary data.
SR2.6 Release date is a data descriptor.

TABLE 3. Source to software requirement cross reference

SR2.7 The GSA catalogue should contain associated datasets. SR2.14.3 The object catalogue should include object descriptors derived from data. SR2.14.4 Remove derived objects for proprietary data. SR3.3.5 View proposal information associated with a data superset. SR3.4.2 Only authorized users shall be able to retrieve proprietary data. SR3.4.4 Data processing shall not use proprietary data as input. SR3.4.6 Authorized Gemini staff can retrieve proprietary data. SR3.4.7 GSA users can retrieve calibration data. SR3.4.9 It shall be possible to retrieve science programs. SR3.4.10 It shall be possible to retrieve proposal information. SR5.2 Data shall be physically secure. SR5.3 Data shall be electronically secure. SR5.4 GSA staff shall respect proprietary periods. SR7.5.1 Gemini shall provide passwords for executed proposals. SR7.15 Gemini shall provide a release date for each dataset. SR7.15.1 Gemini shall inform the GSA when release dates change.
[6] Section 7.2.1 (c) Provide access to data descriptors. SR2.1 The GSA catalogue shall contain all data descriptors. SR2.11 Proposal id shall be a descriptor. SR2.15 Publications should be linked to data supersets. SR3.1 The GSA shall have an interactive user interface. SR3.3 View descriptors for a subset of the catalogue. SR3.3.11 Display science programs. SR3.4.9 It shall be possible to retrieve science programs. SR7.1 Gemini shall collect instrument configuration descriptors. SR7.2 Gemini shall collect telescope configuration descriptors. SR7.4 Gemini shall collect accurate WCS information. SR7.5 Gemini shall make proposal information available to the GSA. SR7.5.2 Gemini shall provide information to link proposals with datasets. SR7.9 Supply executed science programs to the GSA. SR7.13.1 Gemini shall provide information to identify calibration data. SR7.15 Gemini shall provide a release date for each dataset. SR7.15.1 Gemini shall inform the GSA when release dates change. SR7.17 Gemini shall provide version information.
[6] Section 7.2.1 (d) Provide electronic observing log. SR2.10 The electronic observing log shall be associated with the catalogue. SR3.3.9 GSA users shall be able to view the electronic observing log. SR3.4.11 It shall be possible to retrieve the electronic observing log. SR7.8 Gemini shall provide an electronic observing log.
[6] Section 7.2.1 (e) Provide atmospheric monitoring data. SR2.2 The catalogue shall include environmental information. SR3.3.10 View environmental data. SR3.4.8 Retrieve environmental data. SR7.3 Gemini shall record weather information.

TABLE 3. Source to software requirement cross reference

[6] Section 7.2.1 (f) Ensure calibration material is available.
SR2.13 Calibration data shall be associated with science data.
SR3.3.9 GSA users shall be able to view the electronic observing log.
SR3.4.7 GSA users can retrieve calibration data.
SR3.4.11 It shall be possible to retrieve the electronic observing log.
SR7.1 Gemini shall collect instrument configuration descriptors.
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.13.1 Gemini shall provide information to identify calibration data.
[6] Section 7.2.1 (g) Archive all science data.
SR1.1 The GSA shall archive all data.
SR4.1 Maximum and average requirements for catalogue ingest.
SR4.2 Maximum and average requirements for raw data ingest.
SR5.6 GSA operations staff shall support ongoing development of Gemini.
SR7.12 Gemini shall provide raw science data to the GSA.
SR7.18 Gemini shall provide test data.
[6] Section 7.2.1 (i) Provide quick-look reduction pipelines for each archived instrument mode.
SR1.3 The GSA shall be able to automatically process data.
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR2.5 The catalogue should include image descriptors derived from the data.
SR2.13 Calibration data shall be associated with science data.
SR2.14.3 The object catalogue should include object descriptors derived from data.
SR3.4.3 GSA users should be able to choose to have data processed.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.14 Gemini should provide data processing recipes.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
[6] Section 7.2.1 (j) Provide data quality assessment for each dataset.
SR2.8 Data quality assessment shall be in the catalogue.
SR7.6 Gemini shall provide a data quality assessment of each dataset.
[6] Section 7.2.1 (k) Provide data storage, catalogue, calibration frames and pipeline processing.
SR1.1 The GSA shall archive all data.
SR1.3 The GSA shall be able to automatically process data.
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR2.5 The catalogue should include image descriptors derived from the data.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.13 Calibration data shall be associated with science data.
SR3.4.7 GSA users can retrieve calibration data.
SR7.12 Gemini shall provide raw science data to the GSA.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.14 Gemini should provide data processing recipes.

TABLE 3. Source to software requirement cross reference

[6] Section 7.2.1 (l) Provide a mirrorable catalogue.
SR1.4.2 Remove previews when release date is changed.
SR1.4.3 Preview should be mirrorable.
SR2.1 The GSA catalogue shall contain all data descriptors.
SR2.3 The catalogue shall include accurate WCS information.
SR2.6 Release date is a data descriptor.
SR2.11 Proposal id shall be a descriptor.
SR2.12 The GSA catalogue shall be mirrorable.
SR2.14 The GSA should have an object catalogue.
SR2.14.2 The GSA object catalogue should be up to date.
SR2.14.3 The object catalogue should include object descriptors derived from data.
SR2.14.4 Remove derived objects for proprietary data.
SR2.15 Publications should be linked to data supersets.
[6] Section 7.2.1 (m) Provide catalogue search capabilities.
SR2.3 The catalogue shall include accurate WCS information.
SR3.2 Select subsets of the GSA catalogue.
SR3.6 Proposal information should be searchable.
SR3.7 Preview display for each science data superset.
[6] Section 7.2.1 (n) Provide retrieval and delivery systems.
SR3.4 GSA users shall be able to select data for retrieval.
SR3.4.1 Users shall retrieve data via Internet, EXABYTE, or CD-ROM.
SR4.6 Data for Internet retrieval should be available promptly.
SR5.2 Data shall be physically secure.
SR5.3 Data shall be electronically secure.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.15.1 Gemini shall inform the GSA when release dates change.
[6] Section 7.2.2 (a) Provide advanced search capabilities.
SR2.3 The catalogue shall include accurate WCS information.
SR2.5 The catalogue should include image descriptors derived from the data.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.14 The GSA should have an object catalogue.
SR2.14.1 Object parameters should be extracted from object catalogues.
SR2.14.2 The GSA object catalogue should be up to date.
SR2.14.3 The object catalogue should include object descriptors derived from data.
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
SR3.3.13 Save a set of data supersets to a file.
SR3.2.3 Support cross catalogue searches.
SR3.6 Proposal information should be searchable.
SR5.1 Participate in catalogue unification efforts.
SR5.8 The GSA operations staff should add new features to the GSA.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
SR8.2 A unified astronomical archive catalogue should be created.

TABLE 3. Source to software requirement cross reference

[6] Section 7.2.2 (b) Provide on-the-fly re-calibration.
SR1.3 The GSA shall be able to automatically process data.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.13 Calibration data shall be associated with science data.
SR3.4.3 GSA users should be able to choose to have data processed.
SR4.6 Data for Internet retrieval should be available promptly.
SR5.8 The GSA operations staff should add new features to the GSA.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.14 Gemini should provide data processing recipes.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
[6] Section 7.2.2 (c) Provide cross-referencing to other archives.
SR2.3 The catalogue shall include accurate WCS information.
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
SR3.2.3 Support cross catalogue searches.
SR3.3.4 View descriptors of objects in a data superset field of view.
SR4.5 Access to the GSA catalogue must be interactive.
SR5.1 Participate in catalogue unification efforts.
SR5.8 The GSA operations staff should add new features to the GSA.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
SR8.2 A unified astronomical archive catalogue should be created.
[6] Section 7.2.2 (g) Provide visualization and plotting tools.
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR3.3.8 Plot a subset of data supersets in spatial coordinates.
SR3.7 Preview display for each science data superset.
SR3.3.13 Save a set of data supersets to a file.
SR3.3.14 Load a set of data supersets from a file.
SR5.8 The GSA operations staff should add new features to the GSA.
[6] Section 7.2.2 (h) Provide data mining capabilities.
SR1.3 The GSA shall be able to automatically process data.
SR2.3 The catalogue shall include accurate WCS information.
SR2.13 Calibration data shall be associated with science data.
SR2.14 The GSA should have an object catalogue.
SR2.14.1 Object parameters should be extracted from object catalogues.
SR2.14.2 The GSA object catalogue should be up to date.
SR2.14.3 The object catalogue should include object descriptors derived from data.
SR2.14.4 Remove derived objects for proprietary data.
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
SR5.1 Participate in catalogue unification efforts.
SR5.8 The GSA operations staff should add new features to the GSA.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
SR8.2 A unified astronomical archive catalogue should be created.
[6] Section 8.7 (a) (6) A firm proposal for operating the Science Archive once it is complete.
SR6.2 GSA operating cost.

TABLE 3. Source to software requirement cross reference

UR1.1 The archive data must be fully described
SR1.5 Collect information needed to track of publications based on archive data.
SR2.1 The GSA catalogue shall contain all data descriptors.
SR2.2 The catalogue shall include environmental information.
SR2.3 The catalogue shall include accurate WCS information.
SR2.6 Release date is a data descriptor.
SR2.10 The electronic observing log shall be associated with the catalogue.
SR2.13 Calibration data shall be associated with science data.
SR2.15 Publications should be linked to data supersets.
SR2.16 The GSA shall store Gemini hardware and software versions.
SR3.3 View descriptors for a subset of the catalogue.
SR3.3.12 Display the members of association data supersets.
SR7.1 Gemini shall collect instrument configuration descriptors.
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.3 Gemini shall record weather information.
SR7.4 Gemini shall collect accurate WCS information.
SR7.5 Gemini shall make proposal information available to the GSA.
SR7.6 Gemini shall provide a data quality assessment of each dataset.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.15 Gemini shall provide a release date for each dataset.
SR7.15.1 Gemini shall inform the GSA when release dates change.
SR7.17 Gemini shall provide version information.
UR1.2 There should be meta data describing the data descriptors
SR3.3.1 There shall be on-line help describing all catalogue data descriptors.
SR7.11 Gemini shall provide descriptions of data descriptors.
UR1.3 Select sets of the data supersets in the GSA
SR2.7 The GSA catalogue should contain associated datasets.
SR3.2 Select subsets of the GSA catalogue.
SR4.5 Access to the GSA catalogue must be interactive.
UR1.4 Select sets of data supersets based on coordinate
SR2.3 The catalogue shall include accurate WCS information.
SR2.7 The GSA catalogue should contain associated datasets.
SR3.2.1 Select subsets of the GSA catalogue based on coordinate region search.
SR4.5 Access to the GSA catalogue must be interactive.
UR1.5 The GSA must display data superset descriptors to the user
SR2.7 The GSA catalogue should contain associated datasets.
SR3.3 View descriptors for a subset of the catalogue.
SR4.5 Access to the GSA catalogue must be interactive.
UR1.6 A general description of the GSA must be available to users
SR3.5 A general description of the content of the GSA shall be available.
SR7.10 Gemini shall provide a general description of the telescope.
UR1.7 A description of the instruments must be available
SR3.5.1 A description of the Gemini instruments shall be available.
SR7.10.1 Gemini shall provide a description of the instruments.

TABLE 3. Source to software requirement cross reference

UR1.8 A description of Gemini facilities must be available
SR3.5.2 A description of the Gemini facilities shall be available.
SR7.10.2 Gemini shall provide a description of Gemini facilities.
UR1.9 GSA users may select sets of data supersets for retrieval
SR2.7 The GSA catalogue should contain associated datasets.
SR3.2 Select subsets of the GSA catalogue.
SR3.4 GSA users shall be able to select data for retrieval.
SR3.4.1 Users shall retrieve data via Internet, EXABYTE, or CD-ROM.
SR4.3 Maximum and average requirements for data retrieval by users.
SR4.6 Data for Internet retrieval should be available promptly.
SR5.3 Data shall be electronically secure.
UR1.10 Calibration data must be easily available
SR2.13 Calibration data shall be associated with science data.
SR3.3.2 GSA users can view a list of calibration data supersets.
SR3.4.7 GSA users can retrieve calibration data.
SR3.4.8 Retrieve environmental data.
SR3.4.11 It shall be possible to retrieve the electronic observing log.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.13.1 Gemini shall provide information to identify calibration data.
UR1.11 Retrieve proprietary data
SR3.4.2 Only authorized users shall be able to retrieve proprietary data.
SR5.3 Data shall be electronically secure.
SR7.5.1 Gemini shall provide passwords for executed proposals.
SR7.15 Gemini shall provide a release date for each dataset.
SR7.15.1 Gemini shall inform the GSA when release dates change.
SR7.15.2 Inform the GSA which Gemini staff are permitted to access proprietary data.
UR1.12 Select data supersets based on object parameter
SR2.3 The catalogue shall include accurate WCS information.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.14 The GSA should have an object catalogue.
SR2.14.1 Object parameters should be extracted from object catalogues.
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
SR4.5 Access to the GSA catalogue must be interactive.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
UR1.13 Display object parameters
SR1.3 The GSA shall be able to automatically process data.
SR2.14 The GSA should have an object catalogue.
SR2.14.1 Object parameters should be extracted from object catalogues.
SR2.14.2 The GSA object catalogue should be up to date.
SR2.14.3 The object catalogue should include object descriptors derived from data.
SR3.3.1 There shall be on-line help describing all catalogue data descriptors.
SR3.3.4 View descriptors of objects in a data superset field of view.
SR8.1 GSA needs access to SIMBAD and NED catalogues.

TABLE 3. Source to software requirement cross reference

UR1.14 Proposal information for public data must be available
SR2.4 Proposal information should be associated with data supersets.
SR2.11 Proposal id shall be a descriptor.
SR3.3.5 View proposal information associated with a data superset.
SR3.4.10 It shall be possible to retrieve proposal information.
SR7.5 Gemini shall make proposal information available to the GSA.
SR7.5.2 Gemini shall provide information to link proposals with datasets.
UR1.15 Proposal information must be searchable
SR2.4 Proposal information should be associated with data supersets.
SR3.6 Proposal information should be searchable.
SR7.5 Gemini shall make proposal information available to the GSA.
UR1.16 Search for data based on proposal
SR2.11 Proposal id shall be a descriptor.
SR3.6 Proposal information should be searchable.
SR3.6.1 Find all data supersets associated with proposals.
SR7.5.2 Gemini shall provide information to link proposals with datasets.
UR1.17 Support unification of astronomical catalogues
SR2.3 The catalogue shall include accurate WCS information.
SR3.2.3 Support cross catalogue searches.
SR5.1 Participate in catalogue unification efforts.
SR8.2 A unified astronomical archive catalogue should be created.
UR1.19 Data requested by users should be available quickly
SR1.3.2 Data processing shall proceed promptly.
SR4.3 Maximum and average requirements for data retrieval by users.
SR4.6 Data for Internet retrieval should be available promptly.
UR1.20 The GSA should be available 24 hours a day
SR1.3.2 Data processing shall proceed promptly.
SR4.6 Data for Internet retrieval should be available promptly.
UR1.21 The GSA should have an object catalogue
SR2.3 The catalogue shall include accurate WCS information.
SR2.12 The GSA catalogue shall be mirrorable.
SR2.14 The GSA should have an object catalogue.
SR3.2.2 GSA users should be able to select data supersets based on object descriptor.
SR3.3.1 There shall be on-line help describing all catalogue data descriptors.
SR3.3.4 View descriptors of objects in a data superset field of view.
SR3.3.7 Sort the displayed objects.
SR8.1 GSA needs access to SIMBAD and NED catalogues.
UR1.22 The GSA staff must provide support to GSA users
SR5.5 GSA operations staff shall support GSA users.
SR7.16 Gemini shall provide support for GSA users.
UR2.1 Support multi archive searches
SR5.1 Participate in catalogue unification efforts.
SR8.2 A unified astronomical archive catalogue should be created.

TABLE 3. Source to software requirement cross reference

UR2.2 Automatic re-processing of data should be an option
SR1.2 The GSA shall respect proprietary periods assigned by Gemini.
SR1.3 The GSA shall be able to automatically process data.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.13 Calibration data shall be associated with science data.
SR3.4.2 Only authorized users shall be able to retrieve proprietary data.
SR3.4.3 GSA users should be able to choose to have data processed.
SR3.4.4 Data processing shall not use proprietary data as input.
SR4.3 Maximum and average requirements for data retrieval by users.
SR4.6 Data for Internet retrieval should be available promptly.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.14 Gemini should provide data processing recipes.
UR2.3 Automatic re-processing inputs and options should be documented
SR1.3 The GSA shall be able to automatically process data.
SR3.4.5 Processing recipes shall be documented.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
UR2.4 Users should be able to choose one or more processing recipes
SR1.3 The GSA shall be able to automatically process data.
SR3.4.3 GSA users should be able to choose to have data processed.
SR3.4.4 Data processing shall not use proprietary data as input.
SR3.4.5 Processing recipes shall be documented.
SR4.6 Data for Internet retrieval should be available promptly.
SR7.14 Gemini should provide data processing recipes.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
UR2.5 Information about re-processing recipes should be available
SR3.4.5 Processing recipes shall be documented.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
UR2.6 GSA users may preview data for selected data supersets
SR1.3 The GSA shall be able to automatically process data.
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.12 The GSA catalogue shall be mirrorable.
SR2.13 Calibration data shall be associated with science data.
SR3.7 Preview display for each science data superset.
SR4.4 Preview should be display promptly.
SR7.14 Gemini should provide data processing recipes.
UR2.7 Preview must be display promptly
SR1.4 Preview data should be generated for each dataset.
SR1.4.1 Generate previews for associations.
SR4.4 Preview should be display promptly.
UR2.8 Sort the displayed set of data supersets
SR3.3.6 Sort the displayed subset of the catalogue.
SR3.3.7 Sort the displayed objects.

TABLE 3. Source to software requirement cross reference

UR2.9 GSA catalogue needs descriptors derived from the data
SR1.3 The GSA shall be able to automatically process data.
SR2.5 The catalogue should include image descriptors derived from the data.
SR2.5.1 Remove derived image descriptors for proprietary data.
SR2.13 Calibration data shall be associated with science data.
SR3.2 Select subsets of the GSA catalogue.
SR3.3 View descriptors for a subset of the catalogue.
SR3.3.1 There shall be on-line help describing all catalogue data descriptors.
UR2.10 The catalogue should contain information on associated data
SR1.3 The GSA shall be able to automatically process data.
SR1.4.2 Remove previews when release date is changed.
SR1.5 Collect information needed to track of publications based on archive data.
SR2.6 Release date is a data descriptor.
SR2.7 The GSA catalogue should contain associated datasets.
SR2.12 The GSA catalogue shall be mirrorable.
SR2.13 Calibration data shall be associated with science data.
SR2.14 The GSA should have an object catalogue.
SR2.15 Publications should be linked to data supersets.
SR3.2 Select subsets of the GSA catalogue.
SR3.3 View descriptors for a subset of the catalogue.
SR3.3.3 View the list of publications associated with a data superset.
SR3.3.4 View descriptors of objects in a data superset field of view.
SR3.3.5 View proposal information associated with a data superset.
SR3.3.9 GSA users shall be able to view the electronic observing log.
SR3.3.10 View environmental data.
SR3.3.11 Display science programs.
SR3.3.12 Display the members of association data supersets.
SR3.4.15 Retrieve associated data supersets.
SR3.4.16 Provide data processing recipes for associations.
SR3.7 Preview display for each science data superset.
UR2.11 A single query needs to address any or all instruments
SR3.2 Select subsets of the GSA catalogue.
UR2.12 Access to the GSA catalogue must be interactive
SR3.1 The GSA shall have an interactive user interface.
SR3.3 View descriptors for a subset of the catalogue.
SR3.4 GSA users shall be able to select data for retrieval.
SR4.5 Access to the GSA catalogue must be interactive.
UR2.13 Publications should be linked to the data
SR2.12 The GSA catalogue shall be mirrorable.
SR2.15 Publications should be linked to data supersets.
SR3.3.3 View the list of publications associated with a data superset.
SR7.7 Gemini should track publications associated with data.
UR2.14 It should be possible to plot the fields of view of a set of data supersets
SR2.3 The catalogue shall include accurate WCS information.
SR3.3.8 Plot a subset of data supersets in spatial coordinates.

TABLE 3. Source to software requirement cross reference

UR2.15 Select a set of data supersets based on spatial density
SR2.3 The catalogue shall include accurate WCS information.
SR3.3.13 Save a set of data supersets to a file.
SR3.3.14 Load a set of data supersets from a file.
UR2.16 Data quality indicator must be one of the data descriptors
SR2.8 Data quality assessment shall be in the catalogue.
SR7.6 Gemini shall provide a data quality assessment of each dataset.
UR2.17 Search based on a list of coordinates
SR2.3 The catalogue shall include accurate WCS information.
SR2.7 The GSA catalogue should contain associated datasets.
SR3.2.1 Select subsets of the GSA catalogue based on coordinate region search.
SR4.5 Access to the GSA catalogue must be interactive.
UR2.18 Monitor progress of data requests
SR3.4.12 It should be possible for users to monitor the progress of data requests.
UR2.19 View and retrieve the science program
SR2.9 The science program shall be associated with the catalogue data.
SR3.3.11 Display science programs.
SR3.4.9 It shall be possible to retrieve science programs.
SR7.9 Supply executed science programs to the GSA.
UR2.20 The content of user requests should be documented
SR3.4.14 Document user data requests.
UR2.21 The GSA should support a batch mode interface
SR3.2.4 The GSA should support a batch mode interface.
SR3.4.13 Batch mode requests.
UR2.22 The GSA should be accessible to users with poor internet connections
SR3.2.4 The GSA should support a batch mode interface.
SR3.4.13 Batch mode requests.
UR3.1 Other astronomy archives must participate in unification efforts
SR5.1 Participate in catalogue unification efforts.
SR8.2 A unified astronomical archive catalogue should be created.

2. GSA Software Requirements

Each of the software requirements has the following information:

- A unique software requirement identifier. This identifier is in the form SR1.1, and after the acceptance of the FPRD, will remain fixed for each requirement throughout the GSA development.
- A “Level” value indicates whether the requirement is a fundamental requirement, a key requirement, or an advanced requirement. Values for the Level indicator, taken from [9] are:

  Fundamental — These requirements must be met in order for data to be fully archievable. Data collected before these requirements are met will not be fully archievable.

  Key — These requirements are necessary in order to have a basically functional archive. An archive that meets all of these requirements will be able to ingest data into the archive,

allow users to identify data in the archive, and will allow users to retrieve data from the archive.

Advanced — These are requirements for the GSA to actively manipulate the archived data in order to extract "value added" information.

- A "necessity" value, which indicates the importance of the requirement to the functionality of the GSA. Necessity has values of:

  Mandatory — The functionality must be in place in order for the GSA to function correctly.

  Note that there are cases where there are mandatory requirements which only make sense if other non-mandatory requirements are implemented. An example of this is the generation of a derived object catalogue (SR2.14), which is not mandatory, but deletion of objects in the object catalogue, where were derived from proprietary data (SR2.14.4) is mandatory. This should be interpreted as meaning that if the non-mandatory requirement is implemented, the mandatory requirement shall also be implemented.

  Very desirable — The functionality will make the GSA much more useful, but the archive will still be usable without the functionality.

  Desirable — The functionality will make the GSA somewhat more usable.

  Optional — The functionality is useful, but the cost of implementation is expected to be high, and/or the benefit is marginal.

- A short description of the requirement.
- A complete description of the requirement.
- A description of how the requirement will be verified.

2.1 Functional Requirements

This section lists the functionality required from the GSA in order to allow the scenarios and science cases listed in [8] to be completed. This section is divided into three sub-sections:

- Section 2.1.1 on page 40 lists general requirements
- Section 2.1.2 on page 43 lists the requirements for the GSA catalogue
- Section 2.1.3 on page 49 list requirements specific to the user interface

2.1.1 General Requirements

This section lists the requirements on the GSA which are neither catalogue or user interface specific.

SR1.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: The GSA shall archive all data.

Description: The GSA shall archive all data supplied from the Gemini telescopes. The GSA shall include a data store for all science and calibration data supplied by the Gemini Telescopes. The GSA shall include a catalogue of all data in the GSA data store. The GSA catalogue is fully described in Section 2.1.2 on page 43. The data input rates to be supported by the GSA design are detailed in Section 2.2 on page 58. The exact nature of the GSA to Gemini interface will be documented in ICD TBD.

Verification: A test should be created which simulates reception of data from Gemini, and then verifies that all received data was ingested into the GSA.

SR1.2 **Level:** Key **Necessity:** Mandatory

Short Description: The GSA shall respect proprietary periods assigned by Gemini.

Description: GSA shall allow Gemini to set a public release date for each dataset. Datasets shall be considered proprietary until the public release date. The GSA shall allow Gemini to change the public release date at any time. Any data derived from proprietary data shall be considered proprietary, except where Gemini has created data derived from proprietary data and indicated that the derived data is public (e.g. a sky flats derived from science frames could fall into this category). Gemini shall set and modify public release dates using the interface documented in ICD TBD.

Verification: Attempting to retrieve data, preview, or view derive descriptors for data which is proprietary shall fail, except in those cases where users have authorization to retrieve proprietary data.

SR1.3 **Level:** Advanced **Necessity:** Mandatory

Short Description: The GSA shall be able to automatically process data.

Description: The GSA shall be able to automatically process data. The processing is to be used for preview generation (see SR1.4), optional generation of derived data descriptors (see SR2.5), optional generation of derived object descriptors for the object catalogue (see SR2.14.3), optional generation of automatically re-calibrated data for users (see SR3.4.3), and optional generation of associated data products for users (see SR3.4.3). Automatic processing shall be triggered by various events:

- Data being received by the GSA.
- Data becoming public.
- Data requests by users.
- Manually triggered by GSA staff when new data processing recipes become available for preview generation, derived data superset generation or derived object descriptors generation.

The automatic processing system shall log the inputs and options used for each processing step.

The automatic processing system shall have a priority based queue which will allow user requests to be given priority over other tasks.

The processing recipes will be developed by GSA staff, initially based on the data quality evaluation recipes developed and used by Gemini staff.

Verification: There will be tests which show data processing can be initiated when data is received by the GSA. There will be tests which show data processing can be initiated when data becomes public. There will be tests which show data processing being initiated by user requests. There will be tests which show data processing being initiated by new recipes being added to the suite of recipes to be automatically executed.

SR1.3.1 **Level:** Key **Necessity:** Mandatory

Short Description: Data processing shall respect Gemini policy on proprietary periods.

Description: The GSA data processing shall not allow GSA users access to proprietary data (as defined SR1.2) unless the GSA user has authority to access all proprietary data used as input to the data processing. See SR3.4.2 for a description of how a user would access proprietary data.

Verification: A test which simulates an unauthorized user requesting a processed data product which is generated from proprietary data will demonstrate that the GSA will not deliver the processed data product.

SR1.3.2 **Level:** Advanced **Necessity:** Mandatory

Short Description: Data processing shall proceed promptly.

Description: Data processing required to satisfy user requests shall be given priority over the other data processing tasks, and shall proceed promptly and automatically. Data processing required to satisfy user requests shall not be blocked by other data processing, including by the data processing for other user requests. Data processing for a small user request shall not be delayed by large requests. This can be achieved by implementing a priority queue for processing requests, where smaller requests get higher priority, and by processing several data requests in parallel.

Verification: A test scenario will be created where the GSA data processing system has a large number of low priority tasks queued for processing, and a variety of user requests are submitted. The user requests should include a mixture of large and small requests, and requests that do and do not require data processing. The tests will demonstrate that user requests are processed before the low priority tasks, and large user requests do not block smaller requests.

SR1.4 **Level:** Advanced **Necessity:** Very desirable

Short Description: Preview data should be generated for each dataset.

Description: The data for each science dataset in the GSA should be processed, compressed and stored. The preview data should be available for display as described in SR3.7. This compressed preview should contain enough information to allow an astronomer viewing the preview to judge the quality. In order to keep preview display times reasonable, and to keep preview storage space within practical limits, lossy compression may be used on preview data, and data which doesn't contribute to the evaluation of the dataset may be omitted. Initially, preview data processing recipes should be based on the data processing recipes supplied by Gemini, however GSA staff can extend this data processing as necessary. Preview data shall only be generated after all data used to generate the preview is public, as defined in SR1.2.

There may be cases where preview images are available for associated data supersets (see SR1.4.1), and where in the judgment of GSA scientists the preview data for individual datasets have no value (e.g. infrared images). In these cases, the GSA may choose to intentionally omit previews for individual datasets. The decision about which datasets should have previews will have to be encapsulated into rules which will be automatically executed by the data processing system used to generate previews.

Verification: A test will be created which shows previews being created as datasets become public. There will be tests which count the number of public datasets with preview data, and the number of public datasets without preview data for comparison. Note that since this isn't a mandatory requirement, it is sufficient if most, but not all science datasets have previews.

SR1.4.1 **Level:** Advanced **Necessity:** Desirable

Short Description: Generate previews for associations.

Description: As the GSA identifies data associations as described in SR2.7, each science association should have a preview generated, similar to the previews for individual data supersets described in SR1.4. The previews should contain enough information to allow an astronomer viewing the preview to judge the quality of the data. In order to keep preview display times reasonable, and to keep preview storage space within practical limits, lossy compression may be used on preview data, and data which doesn't contribute to the evaluation of the association may be omitted. Preview data shall only be generated after all data used to generate the preview is public, as defined in SR1.2.

Verification: A test will be created which shows previews being created as associations become public. There will be tests which count the number of public associations with preview data, and

the number of public associations without preview data for comparison. Note that since this isn't a mandatory requirement, it is sufficient if most, but not all science associations have previews.

SR1.4.2 **Level:** Key **Necessity:** Mandatory

Short Description: Remove previews when release date is changed.

Description: If previews for data supersets are generated, as described in SR1.4 and SR1.4.1, and if Gemini changes the release date of public data making it proprietary again as defined in SR1.2, then any preview data generated from the now proprietary data shall be removed from the GSA. This removal shall be completed within one working day after the GSA has been notified of the change in the release date.

Only previews derived from a single program need be removed from the catalogue. Previews of associations derived from several programs (e.g. by co-adding images) need not be removed.

Verification: A test will be created where the release dates are changed for several public data supersets with preview data. The test will verify that in those cases where the datasets become proprietary, the preview data is removed from the GSA.

SR1.4.3 **Level:** Key **Necessity:** Desirable

Short Description: Preview should be mirrorable.

Description: It should be possible for the primary GSA site to mirror previews generated as described in SR1.4 and SR1.4.1 to other sites as part of the catalogue mirroring requirement described in SR2.12. The preview mirroring shall include removal of proprietary previews, as described in SR1.4.2.

Verification: Either:

- If a commercial system is used to store preview data, document that the system has a mirroring capability or option which will propagate data creation, deletion, and updates to a mirror site.

or:

- Create a test where the GSA preview data is mirrored to a second system. The test will demonstrate that preview creation, deletion, and updates are mirrored to the second system.

SR1.5 **Level:** Fundamental **Necessity:** Optional

Short Description: Collect information needed to track of publications based on archive data.

Description: The GSA should have the infrastructure required to keep track of publications based on data retrieved from the GSA. This infrastructure will allow Gemini to find publications based on archive data (see also SR7.7). The exact information to be collect shall be determined in cooperation with Gemini.

Verification: There will be a demonstration which shows that if a GSA user follows the TBD requirements for citing the GSA as a data source for publications, the data used to produce the publication can be identified based on the information in the citation.

2.1.2 Catalogue Requirements

The GSA shall have a catalogue of data supersets, which shall contain all available data descriptors for each data superset. All of the descriptors which must be searchable by (see SR3.2), and displayable to GSA users (see SR3.3) shall be included. The catalogue shall be stored so as to facilitate efficient searches as described in SR3.2, allowing users to have interactive access to the catalogue content.

SR2.1 **Level:** Key **Necessity:** Mandatory

Short Description: The GSA catalogue shall contain all data descriptors.

Description: All data descriptors provided by Gemini shall be stored in the GSA catalogue. This shall include enough information to completely describe the state of the optical path while each dataset or data superset was observed. The GSA shall allow the set of descriptors to change over time, as the telescope, instruments and facilities evolve.

Verification: Visual comparison of ingested datasets with the data available from the GSA user interface will show that all descriptors are stored in the GSA catalogue. A test will be created which demonstrates that the GSA catalogue ingest process will notify GSA staff when any descriptors received from Gemini cannot be ingested into the catalogue.

SR2.2 **Level:** Key **Necessity:** Mandatory

Short Description: The catalogue shall include environmental information.

Description: The GSA catalogue shall include a record of environmental information recorded during the observation of each dataset. This shall include both weather data recorded by the observatory weather station, and internal data collected by environmental sensors with the telescope enclosure. This shall include all data provided by Gemini, not simply a sample at the beginning and/or end of the observation.

Verification: A test will be created showing environmental data received from Gemini being ingested into the GSA. Another test will verify that all environmental data received can be extracted from the GSA catalogue.

SR2.3 **Level:** Key **Necessity:** Mandatory

Short Description: The catalogue shall include accurate WCS information.

Description: The GSA catalogue shall include accurate WCS information (as supplied by Gemini, see SR7.4) describing each data superset, including the field of view, and wavelength (in the case of spectral observations). The WCS information shall allow the GSA to determine if an object with a known position is in the field of view, to the accuracy of the coordinates supplied by Gemini. The WCS information shall be stored so as to allow correlation between the GSA catalogue and other observation and object catalogues.

Verification: A test will be created that shows that data supersets can be found by searching for a point which is in the field of view, and that data supersets will not be found if the point is not in the field of view.

SR2.4 **Level:** Key **Necessity:** Very desirable

Short Description: Proposal information should be associated with data supersets.

Description: The public information from proposals should be stored in the GSA catalogue, linked to data superset and should be available to archive users. Proposal information should include keywords, science category, and primary target object class associated with the proposal. Any proposal information which is proprietary, as determined by Gemini, shall not be available to archive users, and should not be stored in the GSA catalogue. The association between proposals and data supersets should be two way in that it should be possible to find: a) all data supersets associated with a proposal, b) find the proposals associated with any data superset. If associated data supersets are created (see SR2.7), then there may be more than one proposal associated with an association.

Verification: A test will be created showing that using the user interface, proposal information can be identified and displayed for a selection of datasets.

SR2.5 **Level:** Advanced **Necessity:** Desirable

Short Description: The catalogue should include image descriptors derived from the data.

Description: The GSA catalogue should contain image descriptors derived from the data (for example signal to noise ratio, and signal to noise ratio of associated data). It should be possible to search and display these derived descriptors as described in SR3.2 and SR3.3. The derived descriptors should be automatically calculated after the data becomes public, as defined in SR1.2.

Verification: A test will be created showing derived descriptors being automatically processed as data supersets become public. A test will be created demonstrating that a user can search and display the derived image descriptors.

SR2.5.1 **Level:** Advanced **Necessity:** Mandatory

Short Description: Remove derived image descriptors for proprietary data.

Description: If image descriptors are derived from raw data as described in SR2.5, derived descriptors shall be removed from the catalogue if the release date of any data used to calculate the descriptors is changed to make the data proprietary again, as defined in SR1.2. This removal shall be completed within one working day after the GSA has been notified of the change in the release date. (Note this is similar to SR1.4.2).

Only image parameters derived from a single program need be removed from the catalogue. Parameters derived from several programs (e.g. by co-adding images) need not be removed.

Verification: A test will be created which modifies the release date of public datasets which have derived descriptors. The test will ensure that the derived descriptors are removed from the GSA catalogue.

SR2.6 **Level:** Key **Necessity:** Mandatory

Short Description: Release date is a data descriptor.

Description: The release date for each data superset shall be considered a data descriptor stored in the GSA catalogue, and should be searchable and viewable by archive users as described in SR3.2 and SR3.3. Any updates made to the release date shall be reflected in the GSA catalogue, and in any mirror catalogues (see SR2.12). The release date for associations shall be calculated as described in SR1.2.

Verification: Demonstrate that release date is a descriptor available for searching and display through the GSA user interface.

SR2.7 **Level:** Advanced **Necessity:** Optional

Short Description: The GSA catalogue should contain associated datasets.

Description: Initially association data supersets should be sets of datasets which overlap spacially, and which can be stacked to create data supersets with a higher S/N ratio. It must be possible to broaden the definition of associations as part of the GSA development or operations, to incorporate other ways of combining datasets (e.g. mosaiced images). Associations should appear in the catalogue subset view described in SR3.3, they should be retrievable as described in SR3.4, and they should have preview images as described in SR1.4.1. Each association should have a public release date which is the latest public release date of all of the data which are used to create the association, as defined in SR1.2.

Verification: Show that associated datasets can be searched and displayed through the GSA user interface.

SR2.8 **Level:** Key **Necessity:** Mandatory

Short Description: Data quality assessment shall be in the catalogue.

Description: The results of the data quality assessment done by Gemini (see SR7.6) shall be a data descriptor stored in the GSA catalogue.

Verification: Show that data quality can be searched and displayed with the GSA user interface.

SR2.9 **Level:** Key **Necessity:** Mandatory

Short Description: The science program shall be associated with the catalogue data.

Description: The catalogue shall link science programs to the datasets collected for the science program. It shall be possible to follow the link in both directions, to find datasets associated with a science program, or to find the science program associated with a dataset. If associated data supersets are created, as described in SR2.7, then the catalogue shall link the associations with all science programs used to collect the members of the association. More than one science program may be linked with a single association.

Verification: Demonstrate that the GSA user interface can display science program associated with data supersets, and that the GSA user interface can locate data supersets associated with a science program.

SR2.10 **Level:** Key **Necessity:** Mandatory

Short Description: The electronic observing log shall be associated with the catalogue.

Description: The content of the electronic observing log, as supplied by Gemini (see SR7.8) shall be stored on-line, and linked to the data supersets in the GSA catalogue. For association data supersets, it shall be possible to link the data superset to the observing logs of all of the member datasets.

Verification: Demonstrate that the GSA user interface can display the electronic observing log associated with both datasets and associated data supersets. Create a test which shows an electronic observing log as received from Gemini being ingested into the GSA catalogue, and verifies that all parameters in the electronic observing log are stored in the GSA catalogue.

SR2.11 **Level:** Key **Necessity:** Mandatory

Short Description: Proposal id shall be a descriptor.

Description: The proposal ID assigned by the Gemini phase I proposal process shall be a descriptor for each data superset in the Gemini catalogue (see SR7.5). If association data supersets are created as described in SR2.7, then all of the proposal identifiers of the member datasets will also be assigned to the associated data superset.

Verification: Demonstrate that proposal id is searchable and viewable for both datasets and associated data supersets.

SR2.12 **Level:** Key **Necessity:** Mandatory

Short Description: The GSA catalogue shall be mirrorable.

Description: It shall be possible to automatically mirror the GSA catalogue to other sites. The data being mirrored shall include all data superset descriptors, preview images, publication data (SR2.15), proposal data (SR2.4), science program data (SR2.9), and if it exists, the object catalogue (SR2.14). The mirrored data does not need to include the complete raw data store. Nominally, the data transfer for to the mirror sites shall be done over the Internet. All insertions, deletions and updates to the GSA catalogue shall be reflected in the mirror site catalogue.

Verification: Either:

- If a commercial system is used to store preview data, document that the system has a mirroring capability or option which will propagate data creation, deletion, and updates to a mirror site.

or:

- Create a test where the GSA preview data is mirrored to a second system. The test will demonstrate that preview creation, deletion, and updates are mirrored to the second system.

SR2.13 **Level:** Key **Necessity:** Mandatory

Short Description: Calibration data shall be associated with science data.

Description: For each data superset, it shall be possible to determine which calibration data are appropriate to recommend to users, to use when processing data to be delivered for users, to use by the GSA when processing data for data mining, and to use when generating previews. Calibration data shall include the calibration data used by Gemini for processing to do data quality evaluation, standard calibration products collected by Gemini (possibly after the science dataset is observed), and any calibration collected at the request of the PI. The selection of calibration data shall be based on the values of data superset descriptors, rules documented by Gemini in the instrument, telescope, and facility manuals (see SR7.13.1), and rules implemented by the Gemini data processing system.

Note that the calibration data initially identified by Gemini may evolve after the dataset is collected, as new calibration data become available.

Verification: Demonstrate that for any science dataset, the list of associated calibration files can be display by the GSA user interface.

SR2.14 **Level:** Advanced **Necessity:** Desirable

Short Description: The GSA should have an object catalogue.

Description: Known objects in the field of view should be associated with each data superset. Information about these objects should include (where known and applicable): object type, limiting magnitude, distance, reddening parameter, and redshift.

Verification: Demonstrate that for any science data superset in the GSA catalogue, the GSA user interface can display a list of astronomical objects in the field of view.

SR2.14.1 **Level:** Advanced **Necessity:** Desirable

Short Description: Object parameters should be extracted from object catalogues.

Description: Object parameters should be extracted from existing external object catalogues such as SIMBAD and NED, based on the field of view from the data superset WCS descriptors. The object descriptors retrieved from these external catalogues should be inserted into the GSA object catalogue. The object information determined in this way is not subject to proprietary restrictions, and should be generated automatically when datasets are received by the GSA. The goal should be to have the objects associated with a data superset in the object catalogue within 24? hours of the GSA receiving the descriptors for the data superset. The object catalogue should be mirrorable, as described in SR2.12.

Verification: Create a test which demonstrates object parameters being added to the object catalogue as new datasets are ingested into the GSA. The test should independently query SIMBAD and NED and compare the results to those ingested into the catalogue to ensure that all objects have been detected.

SR2.14.2 **Level:** Advanced **Necessity:** Desirable

Short Description: The GSA object catalogue should be up to date.

Description: If the object information described in SR2.14.1 is taken from dynamic catalogues like SIMBAD and NED, the object catalogue should be updated to ensure changes in the source catalogues are reflected in the Gemini object catalogue. This could be done by periodically re-checking for objects in each data superset's field of view, by arranging to get notification of updates to the SIMBAD and NED catalogues, or by joining with the source catalogues "on the fly", as needed. The goal should be to have any new objects added to NED or SIMBAD associated with the appropriate data supersets within 30? days. The updates to the object catalogue should be mirrorable, as described in SR2.12.

Verification: Create a test which simulates an updated to the external catalogue. Compare the object catalogue before and after the update to demonstrate that the GSA object catalogue has been updated.

SR2.14.3 **Level:** Advanced **Necessity:** Desirable

Short Description: The object catalogue should include object descriptors derived from data.

Description: Object descriptors should be automatically derived by processing the data for each data superset. The object descriptors should be inserted into the GSA object catalogue. Initially the object descriptors should include those produced by SExtractor [11] and Galfit [3] when run on calibrated science data, however this list should expand during the lifetime of the GSA, as part of the GSA operations. Object extraction should be done automatically. Object extraction shall only occur after all data used as input to the extraction recipes is public, as defined in SR1.2. The derived object descriptors should be mirrorable, as described in SR2.12.

Verification: Demonstrate that the GSA user interface can search and display descriptors derived from the data. Create a test which demonstrates object descriptors being created after data supersets become public.

SR2.14.4 **Level:** Advanced **Necessity:** Mandatory

Short Description: Remove derived objects for proprietary data.

Description: If object parameters are derived from data supersets as described in SR2.14.3, and if the release date of an data superset is changed to make public data proprietary as defined in SR1.2, then any object descriptors derived from the now proprietary data shall be removed from the GSA object catalogue within one working day. The removed object descriptors shall also be removed from any mirrored copies of the GSA object catalogue (see SR2.12).

Only object parameters derived from a single program need be removed from the catalogue. Parameters derived from several programs (e.g. by co-adding images) need not be removed.

The potential problems associated with removing data from external catalogues will have to be reviewed before the GSA enters into any catalogue sharing agreements with other organizations (e.g. SIMBAD or IVO).

Verification: Create a test which modifies the release date of a public data superset, making it proprietary. The test should show derived object descriptors being removed from the object tables.

SR2.15 **Level:** Key **Necessity:** Optional

Short Description: Publications should be linked to data supersets.

Description: If Gemini tracks publications associated with Gemini data (see SR7.7), then the GSA should link publications based on Gemini data to the data supersets used in the publication.

Publication information stored in the GSA catalogue should include a complete citation of the publication. The publication information should be mirrorable, as described in SR2.12.

Verification: Demonstrate that the GSA user interface can be used to display the list of publications associated with any data superset.

SR2.16 **Level:** Key **Necessity:** Mandatory

Short Description: The GSA shall store Gemini hardware and software versions.

Description: The GSA shall store the version of the instrument hardware and software, telescope hardware and software, and the versions of the hardware and software of any Gemini facilities (e.g. ALTAIR) which affect the nature of the data (see also SR7.17).

Verification: Demonstrate that version information is available through the GSA user interface.

2.1.3 User Interface Requirements

This section lists all of the user interface specific requirements for the GSA.

SR3.1 **Level:** Key **Necessity:** Mandatory

Short Description: The GSA shall have an interactive user interface.

Description: The GSA user interface shall allow GSA users to interact with the GSA catalogue from their home institutes. The interaction between the GSA and GSA users shall be via the Internet.

Verification: Demonstrate that the GSA user interface can be accessed over the Internet, and that the features described in the requirements in this section (those that are implemented) can be used interactively.

SR3.2 **Level:** Key **Necessity:** Mandatory

Short Description: Select subsets of the GSA catalogue.

Description: A GSA user shall be able to select a subset of the data supersets in the GSA catalogue by specifying values for data descriptors. For each of a user selected subset of the descriptors, the GSA user shall be able to specify the required values in one of the following ways:

- A set of 1 or more specific values.
- A range of values.

The GSA user shall be able to choose to specify values for any of the descriptors supplied by Gemini, or any of the descriptors derived from the data by the archive centre.

Verification: Demonstrate using the GSA interface to identify a subset of the catalogue based on descriptor values, and that all data supersets in the returned subset do in fact meet the search criteria.

SR3.2.1 **Level:** Key **Necessity:** Mandatory

Short Description: Select subsets of the GSA catalogue based on coordinate region search.

Description: A GSA user shall be able to select the subset of the data supersets in the GSA catalogue which have a field of view which overlaps any one of a user specified set of 1 or more search regions. Each search region shall be specified by a central coordinate (the users choice of ra., dec., and epoch, or galactic coordinates), and a search radius (in seconds of arc). This search must work in conjunction with SR3.2 in that a coordinate search combined with a search based on descriptor values will return the intersection of the two subsets of the GSA catalogue.

Verification: Demonstrate using the GSA interface to identify a subset of the catalogue base on coordinate region, and that all data supersets in the identified subset do in fact fall within the desired coordinate region.

SR3.2.2 **Level:** Advanced **Necessity:** Desirable

Short Description: GSA users should be able to select data supersets based on object descriptor.

Description: The GSA user interface should allow users to select a subset of the data supersets in the GSA catalogue based on the values of object descriptors in the GSA object catalogue (see SR2.14). The resulting subset of the GSA observation catalogue should be all data supersets with objects matching the search criteria in the field of view. This selection process will work in conjuction SR3.2 and SR3.2.1, allowing the user to get the union between the subsets of the GSA catalogue. The users options for specifying object descriptor values are the same as the options for specifying data superset descriptor values, described in SR3.2.

Verification: Demonstrate using the GSA user interface to identify a subset of the catalogue based on object descriptor value, and that all data supersets in the identified subset do in fact contain object with the specified parameters.

SR3.2.3 **Level:** Advanced **Necessity:** Optional

Short Description: Support cross catalogue searches.

Description: GSA users should be able to select subsets of the data supersets in the GSA catalogue based on a correlation with the content of other catalogues. These other catalogues may be the catalogues of other archives (e.g. HST, CFHT, VLT, etc.), they may be project specific catalogues (e.g. EMSS), or they may be general purpose catalogues (e.g., SIMBAD, NED). The types of queries supported should be like: “find data supersets in the GSA catalogue which have objects from catalogue *x* in the field of view, and were the value of descriptor *y* from catalogue x has a specified value” or “find data supersets in the GSA catalogue where the field of view overlaps an observation in catalogue *z*”. This search should work in conjuction with the GSA catalogue searching described in SR3.2, SR3.2.1, and SR3.2.2, producing a union of the subsets for the GSA user.

The only “join” parameter required to be supported is field of view, however it should be possible to qualify a query on data superset descriptor values in both the GSA catalogue and in the other catalogue.

The catalogues selected to be cross-referenced with the GSA will be selected by GSA science staff. Factors to be taken into account when selecting archives to join with the GSA should be:

- scientific usefulness.
- availability of other catalogues.
- cost to the GSA in both resources and money.

Verification: Demonstrate using the GSA user interface to do a cross archive search.

SR3.2.4 **Level:** Key **Necessity:** Desirable

Short Description: The GSA should support a batch mode interface.

Description: GSA users should have the option to use a non-interactive, batch mode user interface to the GSA. The batch mode may rely on the having an Internet and email connection to the GSA for non-interactive communication with the user, but should not use the Internet for direct interaction between the user and the GSA. The batch interface should allow selection of subsets of the data supersets in the GSA catalogue, as described in SR3.2, SR3.2.1, SR3.2.2, SR3.3.13, and SR3.2.3.

Verification: Create a test which uses the GSA batch interface to identify a subset of the GSA catalogue.

SR3.3 **Level:** Key **Necessity:** Mandatory

Short Description: View descriptors for a subset of the catalogue.

Description: A GSA user shall be able to view an interactive display of a user selected subset of the available data superset descriptors for a subset of the data supersets in the GSA catalogue. The descriptors shall be displayed in scientific units. The descriptors available for display are listed in Section 2.1.2 on page 43, and shall include all descriptors supplied by Gemini, and all descriptors derived from the data. The methods for selecting a subset of the data supersets in the GSA catalogue are described in SR3.2, SR3.2.1, SR3.2.2, SR3.3.13, SR3.2.3, and SR3.6.1.

Verification: Demonstrate viewing of descriptors for a sample of data supersets in the GSA catalogue.

SR3.3.1 **Level:** Key **Necessity:** Mandatory

Short Description: There shall be on-line help describing all catalogue data descriptors.

Description: Each of the descriptors required in Section 2.1.2 on page 43 shall have on-line help available through the GSA user interface, describing the purpose of the descriptor, its implications for the data content, and its scientific units. This on-line help will be available for all descriptors supplied by Gemini, all descriptors derived from the data by the GSA, and for all object descriptors in the GSA object catalogue.

Verification: Demonstrate displaying on-line help from the GSA user interface.

SR3.3.2 **Level:** Key **Necessity:** Mandatory

Short Description: GSA users can view a list of calibration data supersets.

Description: GSA users shall be able to view the complete list of calibration data supersets associated with any user selected science data superset. In cases where calibration data supersets are associations, both the associated data superset and the individual members shall be listed (e.g., where a sky flat is created from several science datasets, both the associated sky flat and the individual science frames shall be listed). The role each calibration file plays in the calibration process shall be displayed with the file.

Verification: Demonstrate that for any science dataset, the associated calibration files can be display by the GSA user interface.

SR3.3.3 **Level:** Key **Necessity:** Optional

Short Description: View the list of publications associated with a data superset.

Description: GSA users should be able to view the list of publications associated with any data superset or set of data supersets selected by the user. The publication data stored in the GSA catalogue is described in SR2.15.

Verification: Demonstrate that the GSA user interface can be used to display the list of publications associated with any data superset.

SR3.3.4 **Level:** Advanced **Necessity:** Desirable

Short Description: View descriptors of objects in a data superset field of view.

Description: A GSA user should be able to get a list of objects from the GSA object catalogue which are in the field of view of a user selected data superset. The user should be able to view a user selected subset of the object descriptors from the GSA object catalogue (see SR2.14). The object descriptors should be displayed in scientific units.

Verification: Demonstrate using the GSA interface to display a list of astronomical objects in the field of view of a data superset. Repeat the test for several different data supersets.

SR3.3.5 **Level:** Key **Necessity:** Very desirable

Short Description: View proposal information associated with a data superset.

Description: For any user selected data superset, GSA users should be able to display the public information from the associated proposal (described in SR2.4).

Verification: Use the GSA user interface to view proposal information for several different data supersets.

SR3.3.6 **Level:** Key **Necessity:** Optional

Short Description: Sort the displayed subset of the catalogue.

Description: Users should be able to sort the displayed set of data supersets described in SR3.3 based on the values of 1 or more user selected data descriptors. The sort can be on any combination of descriptors supplied by Gemini, and descriptors derived from the data by the GSA. Each sort descriptor can be sorted in ascending or descending order.

Verification: Demonstrate the user interface's ability to sort by data descriptor value.

SR3.3.7 **Level:** Key **Necessity:** Optional

Short Description: Sort the displayed objects.

Description: Users should be able to sort the displayed list of objects described in SR3.3.4 based on the values of a user selected set of 1 or more object descriptors. The sort can be done based on descriptors derived from data supersets, or on descriptors extracted from other object catalogues. Each sort descriptor can be sorted in ascending or descending order.

Verification: Demonstrate the user interface's ability to sort by object descriptor value.

SR3.3.8 **Level:** Advanced **Necessity:** Optional

Short Description: Plot a subset of data supersets in spatial coordinates.

Description: The GSA interface should provide a way for user to graphically display the fields of view of a subset of the GSA catalogue. The creation of the plot should be part of the catalogue display described in SR3.3. The coordinates of the plot can be chosen by the user to be either galactic coordinates, or ra. and dec.

Verification: Demonstrate by plotting a subset of the GSA catalogue.

SR3.3.9 **Level:** Key **Necessity:** Mandatory

Short Description: GSA users shall be able to view the electronic observing log.

Description: The content of the electronic observing log (see SR2.10) shall be available to GSA users. GSA users shall be able to specify the time period for the electronic observing log display by entering a start and end date and time. GSA users shall be able to specify the time period for the electronic observing log display as the observing periods of the data supersets in the currently displayed subset of the GSA catalogue (see SR3.2, SR3.2.1, SR3.2.2, SR3.3.13, and SR3.2.3 for a description of how subsets of the GSA catalogue shall be selected).

Verification: Demonstrate that the GSA user interface can display the electronic observing log associated with both datasets, and associated data supersets.

SR3.3.10 **Level:** Key **Necessity:** Mandatory

Short Description: View environmental data.

Description: GSA users shall be able to view a user selected subset of the environmental parameters stored in the GSA catalogue (see SR2.2). GSA users shall be able to specify the time period for the environmental data display by entering a start and end date and time. GSA users shall be able to specify the time period for the environmental data display as the observing periods of the data supersets in the currently displayed subset of the GSA catalogue (see SR3.2, SR3.2.1,

SR3.2.2, SR3.3.13, and SR3.2.3 for a description of how subsets of the GSA catalogue shall be selected).

Verification: Demonstrate that the GSA user interface can display the environmental data associated with a data superset, or selected by entering start and end times.

SR3.3.11 **Level:** Key **Necessity:** Mandatory

Short Description: Display science programs.

Description: It shall be possible for GSA users to display all science programs associated with a data superset, or all science programs associated with a proposal id. Ideally the science program display should be the same as that used by the Gemini Observing Tool.

Verification: Demonstrate that the GSA user interface can display the science programs associated with data supersets.

SR3.3.12 **Level:** Advanced **Necessity:** Mandatory

Short Description: Display the members of association data supersets.

Description: If the GSA catalogue contains associations, as described in SR2.7, it should be possible for GSA users to display the datasets which are members of associated data supersets. The member datasets shall be viewed using the same display as is used for the catalogue subset display described in SR3.3.

Verification: Demonstrate that the GSA user interface can display the member datasets for a selection of association data supersets.

SR3.3.13 **Level:** Key **Necessity:** Desirable

Short Description: Save a set of data supersets to a file.

Description: The GSA user interface should allow a user to extract the current subset of the GSA catalogue to a file on the users computer system. The file should contain the values of all descriptors the GSA has selected for viewing (see SR3.3). This feature will allow the user to manipulate the subset of data supersets using whatever tools the user has available on his/her computer.

Verification: Demonstrate using the GSA user interface to save a set of data supersets to a file. Show that all selected descriptors for all data supersets are saved.

SR3.3.14 **Level:** Key **Necessity:** Desirable

Short Description: Load a set of data supersets from a file.

Description: The GSA user interface should allow a user to select a subset of the GSA catalogue based on the content of a file on the users computer system. The file should containing a list of data superset identifiers which are to be members of the subset of data supersets. The file created as required in SR3.3.13 should be an acceptable format for the input file, allowing a GSA user to save data superset descriptors to a file, sub-sample the data supersets in the file using tools available on the users computer, and then load the modified set of data superset back into the GSA user interface for further manipulation or for data retrieval.

Note that only data superset identifiers should be read from the input file. If the file contains any descriptor values other than data superset identifier, they should be ignored.

Verification: Demonstrate using GSA interface to select a subset of the data supersets in the GSA catalogue, based on a list of data superset identifiers in a file on the users system.

SR3.4 **Level:** Key **Necessity:** Mandatory

Short Description: GSA users shall be able to select data for retrieval.

Description: GSA users shall be able to select either individual data supersets, or a complete subset of the data supersets to be retrieved to their own computer. This interface shall be part of

the catalogue display described in SR3.3. See SR3.2, SR3.2.1, SR3.2.2, SR3.3.13, and SR3.2.3 for a description of how subsets of the GSA catalogue shall be selected.

Verification: Demonstrate that the GSA user interface allows users to create data requests. Process test data requests and verify that the requested data is delivered to the user.

SR3.4.1 **Level:** Key **Necessity:** Mandatory

Short Description: Users shall retrieve data via Internet, EXABYTE, or CD-ROM.

Description: GSA users shall be able to select the delivery media type for their data request, from one of the types supported by the GSA. Although the media types supported for user retrieval must evolve over time as technologies change, the baseline for the GSA design shall include Internet (FTP), EXABYTE tape (tar format) and CD-ROM (ISO 9660 or UDF disk format). Data will be stored in FITS format files, or plain text (ASCII) files. For efficiency, and for the convenience of the user, data may be compressed with tar and/or gzip.

Verification: Demonstrate that the GSA user interface will allow users to select media type for data delivery. Process a data requests for each of the supported media, and demonstrate that the requested data can be read from the media.

SR3.4.2 **Level:** Key **Necessity:** Mandatory

Short Description: Only authorized users shall be able to retrieve proprietary data.

Description: Users shall be prevented from retrieving proprietary data, unless they have authorization to access the data for the proposal which generated the data. GSA users shall be able to retrieve proprietary data from the archive, provided they prove they have the required authorization. The method from proving authorization shall be based on the user supplying the passwords associated with the proprietary data in the request. SR1.2 describes how a data superset's proprietary state is determined.

Verification: Demonstrate attempting to retrieve proprietary data with the GSA user interface. The proprietary data will only be retrievable if appropriate authorization is provided.

SR3.4.3 **Level:** Advanced **Necessity:** Optional

Short Description: GSA users should be able to choose to have data processed.

Description: The GSA user interface should allow a user to select 0 or more recipes (described in SR1.3) to be executed on the data they have selected to be retrieved. The recipes will be selected from a set of recipes supported by the GSA.

Verification: Create a test which requests processed data. Show that the data supplied to the user contains the processed data requested.

SR3.4.4 **Level:** Advanced **Necessity:** Mandatory

Short Description: Data processing shall not use proprietary data as input.

Description: If GSA users can choose to have data re-processed (see SR3.4.3), then the re-processing recipes will not use proprietary data as inputs unless the user can provide appropriate authorization. Authorization will be done as described in SR3.4.2. SR1.2 describes how a data superset's proprietary state is determined.

Verification: Create a test which requests processed data, and where some of the inputs to the processing are proprietary. The request should not be processed unless the user has shown he/she has authority to access the proprietary data.

SR3.4.5 **Level:** Advanced **Necessity:** Mandatory

Short Description: Processing recipes shall be documented.

Description: If the automatic processing of retrieved data is selected by a GSA user, as described in SR3.4.3, then the data returned for the request shall include descriptions of the recipes executed, including the inputs and options passed to a recipe. The information shall be in the form of a log file associated with the processed data.

Verification: Create and submit a test data request which includes processed data. Show that the returned data contains a log file documenting the data processing done for the request.

SR3.4.6 **Level:** Key **Necessity:** Mandatory

Short Description: Authorized Gemini staff can retrieve proprietary data.

Description: Authorized Gemini staff shall be able to retrieve proprietary data. Authorization for Gemini staff shall be granted on a per-user basis. The GSA shall be able to add and remove authorizations for Gemini staff, at the request of Gemini.

Verification: Demonstrate an authorized user requesting proprietary data. Demonstrate the addition and removal of an authorized user.

SR3.4.7 **Level:** Key **Necessity:** Mandatory

Short Description: GSA users can retrieve calibration data.

Description: GSA users shall have the option to get all calibration data for data supersets being retrieved. The user shall get calibration data in an un-processed form, unless this conflicts with SR3.4.2 (for example if proprietary science data is merged into a public science flat, the proprietary science data shall not be retrieved by users without proper authorization). If un-processed calibration data cannot be delivered to a user because it is proprietary, the public processed calibration data shall be delivered instead.

Verification: Create and submit a test data request which requests science data supersets, and all associated calibration data. Demonstrate that all calibration data was included in the data to be supplied to the user. Demonstrate that the GSA user interface gives the user the option to get all calibration data supersets associated with the requested science data supersets.

SR3.4.8 **Level:** Key **Necessity:** Mandatory

Short Description: Retrieve environmental data.

Description: A GSA user shall have the option to retrieve all environmental data associated with a data request. The time periods covered by the environmental data shall include the time periods when the data being retrieved was observed.

Verification: Create and submit a data request which requests science data supersets and all associated environmental data. Demonstrate that the environmental data was included in the data to be supplied to the user. Demonstrate that the GSA user interface gives the user the option to get the environmental data associated with the requested science data supersets.

SR3.4.9 **Level:** Key **Necessity:** Mandatory

Short Description: It shall be possible to retrieve science programs.

Description: A GSA user shall have the option to retrieve science programs associated with the data supersets in a data request. Science programs shall include all science programs used to collect the data in the data request. The retrieved science program shall be in a form which can be loaded into the Gemini Observing Tool.

Verification: Create and submit a data request which requests science data supersets and all associated science programs. Demonstrate that the science programs are included in the data to

be supplied to the user. Demonstrate that the GSA user interface gives the user the option to get the science programs associated with the requested science data supersets.

SR3.4.10 **Level:** Key **Necessity:** Mandatory

Short Description: It shall be possible to retrieve proposal information.

Description: A GSA user shall have the option to retrieve the public part of executed Gemini proposals. The proposals retrieved will be all proposals associated with the data supersets in a data request.

Verification: Create and submit a data request which requests science data supersets and all associated proposals. Demonstrate that the proposals are included in the data to be supplied to the user. Demonstrate that the GSA user interface gives the user the option to get the proposals associated with the requested science data supersets.

SR3.4.11 **Level:** Key **Necessity:** Mandatory

Short Description: It shall be possible to retrieve the electronic observing log.

Description: A GSA user shall have the option to retrieve the electronic observing log. The periods covered by the retrieved electronic observing log shall be the periods when the data being retrieved was observed.

Verification: Create and submit a data request which requests science data supersets and the associated part of the electronic observing log. Demonstrate that the electronic observing log is included in the data to be supplied to the user. Demonstrate that the GSA user interface gives the user the option to get the electronic observing log associated with the requested science data supersets

SR3.4.12 **Level:** Key **Necessity:** Optional

Short Description: It should be possible for users to monitor the progress of data requests.

Description: After users have submitted a data request, it should be possible for the user to monitor the progress of the request through the GSA retrieval process. Information provided to the user should include the current state of the request (queued, processing, complete, waiting for a GSA operator).

Verification: Create and submit a data request. Demonstrate the user interface's ability to monitor the progress of the data request.

SR3.4.13 **Level:** Key **Necessity:** Optional

Short Description: Batch mode requests.

Description: It should be possible for the GSA to accept data requests in batch mode. Batch mode requests should not require an interactive connection to the GSA site. Batch mode requests should consist of a list of data superset identifiers, and information describing any data processing and optional types of data to be retrieved. All of the retrieval options available through the interactive user interface should also be available through the batch interface.

Verification: Create and submit a series of batch mode requests, demonstrating the ability to create requests with each of the options described in the requirement.

SR3.4.14 **Level:** Key **Necessity:** Mandatory

Short Description: Document user data requests.

Description: Each user data request shall include a file documenting how the request was created. If the user created the data request from a GSA catalogue subset created as described in SR3.2, SR3.2.1, SR3.2.2, SR3.3.13, and SR3.2.3 then the documentation shall include:

- A description of the qualifications used to create the catalogue subset.

- A flag indicating if the entire subset was selected to be retrieved, or if the user selected specific data supersets to be retrieved.
- A description of all processing options and optional data types selected by the user.

If the user created the data request using the batch interface described in SR3.4.13, then the entire content of the document used to submit the batch request shall be included in the request documentation.

Verification: Create and submit a user data request. Demonstrate that the resulting data includes documentation describing the request.

SR3.4.15 **Level:** Advanced **Necessity:** Optional

Short Description: Retrieve associated data supersets.

Description: GSA users should have the option to retrieve the individual datasets comprising an association. All of the requirements described in SR3.4 — SR3.4.14 should apply when retrieving association data supersets.

Verification: All of the verification for SR3.4 — SR3.4.14 should include or be repeated for requests for associated data supersets.

SR3.4.16 **Level:** Advanced **Necessity:** Optional

Short Description: Provide data processing recipes for associations.

Description: The GSA should try to provide processing recipes which will give GSA users the option to automatically combine the member datasets into a single result.

Verification: Demonstrate that the user interface is able to create a request for a processed data superset. Create a request for a processed data superset, and show that the processed result is created and delivered.

SR3.5 **Level:** Key **Necessity:** Mandatory

Short Description: A general description of the content of the GSA shall be available.

Description: The user interface shall be able to display a brief description of the GSA content. This shall include the wavelength coverage of the archive, the types of instruments included in the archive, and a general description of the Gemini telescopes. Note that this is somewhat different than the descriptions that might be available for current Gemini Telescope operations in that this description must include historical capabilities of the telescopes which may not be present in the descriptions of the currently operational telescopes. There shall be a way to access the instrument and facility descriptions described in SR3.5.1 and SR3.5.2 from the telescope description.

Verification: Use the GSA user interface to display the description of the GSA.

SR3.5.1 **Level:** Key **Necessity:** Mandatory

Short Description: A description of the Gemini instruments shall be available.

Description: The GSA user interface shall be able to display a description of the instrument used to collect any given data superset. This must include instrument capabilities, limitations, operating modes, and calibration procedures. The documentation shall describe the instrument as it was at the time of the observation, and may include the differences between various versions of the instrument.

Verification: Use the GSA user interface to display instrument descriptions for a selection of data supersets. The selected data supersets should include data supersets which use different versions of the instruments.

SR3.5.2 **Level:** Key **Necessity:** Mandatory

Short Description: A description of the Gemini facilities shall be available.

Description: The GSA user interface shall be able to display a description of any Gemini facilities (e.g. ALTAIR) that will effect the data for any given data superset. Enough information shall be available to allow GSA users to evaluate the content of the data superset, and to re-calibrate the data. The documentation must describe the facility as it was at the time of the observation, and may describe the differences between various versions of the facility.

Verification: Use the GSA interface to display facility descriptions for a selection of data supersets. The selected data supersets should include data supersets which use different versions of the facilities.

SR3.6 **Level:** Key **Necessity:** Very desirable

Short Description: Proposal information should be searchable.

Description: It should be possible to search for Gemini proposals based on keywords, and science category. This search should return a list of proposals, not a list of data supersets. The type of information available for display to the user for each proposal shall be determined by Gemini policy. The user should be able to view all public information for any proposal in the returned list of proposals. Only executed proposals should be available.

Verification: Use the GSA interface to search for and display proposals based on keywords and science category.

SR3.6.1 **Level:** Key **Necessity:** Very desirable

Short Description: Find all data supersets associated with proposals.

Description: Given the list of proposals described in SR3.6, it should be possible to find the subset of the data supersets in the GSA catalogue which contain data collected for any proposal selected by the user. The data supersets should be listed in the same was as described in SR3.3.

Verification: Use the GSA interface to find all data supersets associated with a proposal.

SR3.7 **Level:** Key **Necessity:** Very desirable

Short Description: Preview display for each science data superset.

Description: The user interface should provide a method for displaying the preview images described in SR1.4 and SR1.4.1.The interface to the preview should be part of the catalogue display described in SR3.3. Preview data shall not be available for proprietary data supersets, where the proprietary state of data supersets is defined in SR1.2.

Verification: Use the GSA user interface to display the preview data for a selection of science data supersets.

2.2 Data flow Requirements

This section describes the maximum data flow rates to be supported by the GSA design. The data flow rates from the Gemini Telescopes and to GSA users are both specified.

SR4.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Maximum and average requirements for catalogue ingest.

Description: The GSA shall be able to ingest a peak data rate of more than 5000 datasets per day, and a sustained average of more than 500 datasets per day. If peak data rates exceed the peak

GSA ingest rate, but not the average ingest rate, new datasets should be queued to be ingested as quickly as GSA resources permit.

Verification: Show that the GSA can ingest data at or above the peak rate, within one working day of the data being received. Show that the GSA can ingest data at or above the average sustained data rate, with the resources allocated in the GSA operations budget. Show (on paper) that the storage requirements required to support the sustained data rate can be afforded by the GSA operational budget.

SR4.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Maximum and average requirements for raw data ingest.

Description: The GSA shall be able to ingest raw data at a peak data rate of more than 40 Gigabytes per day, and a sustained average of more than 4 Gigabytes of per day. If peak data rates exceed the peak GSA ingest rate, but not the average ingest rate, new datasets should be queued to be ingested as quickly as GSA resources permit.

Verification: Show that the GSA can ingest data at or above the peak rate within one working day of the data being received. Show that the GSA can ingest data at or above the average sustained data rate, with resources allocated in the GSA operations budget. Show (on paper) that the storage requirements required to support the sustained data rate can be afforded by the GSA operational budget.

SR4.3 **Level:** Key **Necessity:** Mandatory

Short Description: Maximum and average requirements for data retrieval by users.

Description: The GSA design and software implementation shall be able to support peak user data retrieval of up to 1 Gigabyte per hour, with a sustained average of 5 Gigabytes per day. If data requests rates exceed the peak GSA data retrieval rates, but not the average data retrieval rates, new requests will be queued to be processed as quickly as GSA resources permit.

Note that this is the data rate which must be supported by the software design and implementation. Since data outflow to users is expected to start small and increase over time, the initial hardware configuration is not required to support this data rate. The “ramp” up to a hardware configuration which can support this data rate should occur over time and as required, funded by the GSA operating budget.

Verification: Show that the GSA can deliver data at or above the peak rate within one working day. Show that the GSA can deliver data at or above the average sustained with resources allocated in the GSA operations budget. The tests should be repeated with all of the data being delivered on each of the supported media (see SR3.4.1).

SR4.4 **Level:** Key **Necessity:** Very desirable

Short Description: Preview should be display promptly.

Description: Preview data should be displayed within 10 seconds of the user requesting the preview, 80 percent of the time. The typical Internet connection between CADC and the Gemini Hilo Base facility should be used to verify this requirement.

Verification: A Gemini staff member in Hilo will be able to run the user interface, and display preview for a selection of data supersets, timing the responses.

SR4.5 **Level:** Key **Necessity:** Mandatory

Short Description: Access to the GSA catalogue must be interactive.

Description: The response to queries of the GSA catalogue must be quick enough for the user interface to be used interactively. All of the following response requirements assume the typical Internet link between the CADC and the Gemini Hilo Base facility will be used to measure the

requirement. “Simple” queries which also return less than 100 data supersets shall take less than 10 seconds 80 percent of the time. Simple queries are defined as:

- Queries which qualify one indexed data superset descriptor (as described in SR3.2).
- Search by a single coordinate region (as described in SR3.2.1).

Note that it is impractical to impose meaningful response times for more complex queries, there are a large number of possible queries, with widely varying complexity.

Verification: A set of test queries should be identified, and supplied to a Gemini staff member in Hilo. The Gemini staff member can then execute the queries and time the responses.

SR4.6 **Level:** Key **Necessity:** Mandatory

Short Description: Data for Internet retrieval should be available promptly.

Description: When a GSA user requests un-processed data to be delivered via the Internet, the elapsed time between submission of the request and the notification that the data is available for down-load should be less than 10 minutes + 10 seconds per Mbyte of data requested, 80% of the time. It would be desirable to put a performance requirement on retrieving processed data, but the difficulty in quantifying the time it will take to do the processing makes this impossible.

Verification: A set of requests for un-processed data requests will be created, and the processing time for each of the requests will be measured.

2.3 Operational Requirements

These requirements must be incorporated into the ongoing operations of the GSA, but do not influence the initial design of the GSA software.

SR5.1 **Level:** Advanced **Necessity:** Very Desirable

Short Description: Participate in catalogue unification efforts.

Description: In order to support cross archives searches involving the GSA catalogue, the GSA should encourage and participate in efforts to create unified astronomical catalogues (e.g. the IVO initiative).

SR5.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Data shall be physically secure.

Description: The GSA shall provide secure data storage of all archived data received by the archive centre for the duration of the GSA. Physical access to the data shall be limited to GSA staff.

SR5.3 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Data shall be electronically secure.

Description: Proprietary data in the GSA shall be protected from electronic access. Outside users shall not be permitted to electronic access to proprietary data without going through the authorization mechanism described in SR3.4.2.

SR5.4 **Level:** Fundamental **Necessity:** Mandatory

Short Description: GSA staff shall respect proprietary periods.

Description: GSA staff shall not access the proprietary data stored in the GSA, except as their GSA duties require.

SR5.5 **Level:** Key **Necessity:** Mandatory

Short Description: GSA operations staff shall support GSA users.

Description: There shall be a method for GSA users to contact GSA staff, to obtain assistance with the use of the GSA. Types of support provided shall be:

- Assistance with using the GSA user interface.
- Assistance understanding the raw data in the GSA.
- Assistance in understanding the derived data created by the GSA pipe-line processing (i.e. previews, processed data retrieved by users, image parameters derived from the data, and object parameters derived from the data).
- Reporting bugs in the GSA user interface.
- Reporting problems with the raw data in the GSA.
- Reporting problems with the derived data in the GSA.

Support personnel shall include both astronomers familiar with the Gemini telescopes and the archive content, and software developers familiar with the GSA software infrastructure.

GSA user support requests shall be directed through the Gemini Help desk. These requests will then be directed to appropriate GSA or Gemini staff.

SR5.6 **Level:** Key **Necessity:** Mandatory

Short Description: GSA operations staff shall support ongoing development of Gemini.

Description: It is expected that the data produced by the Gemini telescopes will evolve over time as the telescopes are updated. The GSA operations staff shall be responsible for updating the GSA to support the evolution of the data produced by the Gemini telescopes. Gemini shall inform GSA staff when changes to the Gemini telescopes will cause changes to the data produced by the telescopes.

SR5.7 **Level:** Key **Necessity:** Mandatory

Short Description: GSA operations staff shall incorporate evolving technologies into the GSA.

Description: As part of the ongoing operations of the GSA, new technologies for data storage, data access, and data processing shall be incorporated into the GSA.

SR5.8 **Level:** Advanced **Necessity:** Optional

Short Description: The GSA operations staff should add new features to the GSA.

Description: The initial version of the GSA will not include all of the advanced features described in this document, and new ideas for archive functionality will no doubt be developed in the future. The GSA operations staff should pursue the implementation of these new and advanced features as GSA operations resources permit.

2.4 Resource Requirements

These requirements describe the resources expected to be allocated by Gemini to implement and operate the GSA. The GSA design shall keep the resources required for development and operations within these constraints.

SR6.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: GSA development cost.

Description: The implementation of the Conceptual design for the GSA shall cost less than USD 150,000.00.

SR6.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: GSA operating cost.

Description: The annual operating cost for the GSA shall be less than USD 175,000.00, with annual inflation adjustments which are tied to an inflation index to be specified in the GSA operations contract.

2.5 Requirements on Gemini

These requirements describe the functionality that shall be implemented by the Gemini project. Where data or information must be transferred on an ongoing basis between the Gemini telescopes and the GSA, the method of transfer will be described in ICD TBD.

SR7.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall collect instrument configuration descriptors.

Description: Gemini shall collect all data descriptors which describe the instrument configuration for each dataset and frame.

SR7.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall collect telescope configuration descriptors.

Description: Gemini shall collect all data descriptors which describe the telescope configuration for each dataset and frame, including the configuration of any facilities (e.g. adaptive optics facilities) which may effect the data.

SR7.3 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall record weather information.

Description: Gemini shall make a record of environmental information during each observation using the observatory weather station.

SR7.4 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall collect accurate WCS information.

Description: Gemini shall collect accurate WCS information describing each dataset, including the field of view, the start and end times, the wavelength being observed, and spectral resolution (if applicable).

SR7.5 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall make proposal information available to the GSA.

Description: Gemini shall make public proposal information available to the GSA. The proposal information shall include a proposal identifier, a science category, and a list of keywords which could be used to search for proposals by topic.

SR7.5.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide passwords for executed proposals.

Description: For every executed proposal, Gemini shall provide the GSA with a password. This password will be used by archive users to authenticate retrieval of proprietary data.

SR7.5.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide information to link proposals with datasets.

Description: Gemini will provide the information required to link each proposal with the datasets taken for that proposal.

SR7.6 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide a data quality assessment of each dataset.

Description: The data quality assessment shall indicate if the dataset meets Gemini quality standards, and if the dataset meets the standards required for the proposal.

SR7.7 **Level:** Fundamental **Necessity:** Optional

Short Description: Gemini should track publications associated with data.

Description: In order to provide the data required to support SR2.15, and SR3.3.3, Gemini should attempt to collect as much information as possible, linking publications to PI proposals and Gemini archive data requests. The publication information should be transmitted from Gemini to the GSA in the form of a link to the ADS (Astronomical Data Service), associated with one or more Gemini proposal identifiers. The transfer of publication information from Gemini to the GSA will be described in detail in ICD TBD.

SR7.8 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide an electronic observing log.

Description: Gemini shall create an electronic observing log recording the state of the telescope and instrument during each observation.

SR7.9 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Supply executed science programs to the GSA.

Description: Science programs executed by Gemini shall be provided to the GSA. The science programs shall be in a form which will allow them to be re-loaded into the Gemini Observing Tool.

SR7.10 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide a general description of the telescope.

Description: Gemini shall provide (electronically) the GSA with a general description of the capabilities of the Gemini Telescopes. Gemini shall provide updated descriptions of the Gemini Telescopes as the telescopes and their instruments evolve.

This may differ from the documentation that will (we assume) be available for current Gemini Telescope operations in that this description shall include historical capabilities of the telescopes which may not be present in the descriptions of the current operational telescopes. If different versions of the documents are created as the telescopes evolve, it shall be possible to associate the correct version of the documents with each dataset, using either observation date or telescope version information included with the dataset descriptors.

Note that using observation date to associate documentation with datasets assumes that there is a well defined date where a transition between versions occurs. This may not be true, especially in the case of software updates.

SR7.10.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide a description of the instruments.

Description: Gemini shall provide (electronically) the GSA with descriptions of the Gemini instruments. Gemini shall provide updates to the instrument descriptions as instruments evolve, and as new instruments are introduced.

This may differ from the documentation that will (we assume) be available for current Gemini Telescope operations in that this description shall include historical capabilities of the instruments which may not be present in the descriptions of the current operational instruments. If dif-

ferent versions of the documents are created as the instruments evolve, it shall be possible to associate the correct version of the documents with each dataset, using either observation date or instrument version information included with the dataset descriptors.

Note that using observation date to associate documentation with datasets assumes that there is a well defined date where a transition occurs. This may not be true, especially in the case of software updates.

SR7.10.2 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide a description of Gemini facilities.

Description: Gemini shall provide (electronically) the GSA with descriptions of Gemini facilities. Gemini shall provide updates to facility descriptions as facilities evolve, and as new facilities are introduced.

This may differ from the documentation that will (we assume) be available for current Gemini Telescope operations in that this description must include historical capabilities of the facilities which may not be present in the descriptions of the current operational facilities. If different versions of the documents are created as the facilities evolve, it shall be possible to associate the correct version of the documents with each dataset, using either observation date or telescope version information included with the dataset descriptors.

Note that using observation date to associate documentation with datasets assumes that there is a well defined date where a transition occurs. This may not be true, especially in the case of software updates.

SR7.11 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide descriptions of data descriptors.

Description: Gemini shall provide descriptions of the data descriptors collected as required in SR7.1, SR7.2, SR7.3, SR7.4, SR7.6, and SR7.8. These descriptions shall contain information at a level of detail appropriate for use in the on-line help system as described in SR2.1.

SR7.12 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide raw science data to the GSA.

Description: Gemini shall provide the GSA with all raw science and calibration data collected by the Gemini telescopes and which meets the Gemini data quality standards. Because of the proprietary period for raw data, the data need not be available promptly, and so weekly, or monthly deliveries of data by mail would be adequate.

SR7.13 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall ensure that calibration data is available.

Description: Gemini shall identify a basic calibration level for each instrument. For Queue and service mode observations, Gemini shall ensure that enough calibration data is collected to allow science data to be calibrated to the basic calibration level.

For classical mode observations, Gemini shall strongly encourage observers to collect the basic level of calibration data.

It is recognized that there will be situations where collection of the basic level of calibration data may not be practical due to the observing needs of a PI proposal. In these cases Gemini should collect as much of the basic calibration data as is practical.

SR7.13.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide information to identify calibration data.

Description: Gemini shall provide enough information to identify calibration data for each science dataset. This might be specific information supplied with each science dataset identifying a calibration dataset, it might be a set of rules which identify calibration datasets, or it might be information extracted from the science program. The information shall allow the GSA to find all calibration data required to automatically create preview data, and automatically re-calibration science data.

If the best calibration data to use changes over time, due to new calibrations becoming available, Gemini shall inform the GSA of the new set of “best” calibration date.

SR7.14 **Level:** Advanced **Necessity:** Very desirable

Short Description: Gemini should provide data processing recipes.

Description: Gemini should provide (electronically) the GSA with the data processing recipes used by Gemini to evaluate data quality. These recipes will be incorporated into the GSA reprocessing recipes described in SR3.4.3. GSA staff should be notified when new recipes are created or when existing recipes are updated.

SR7.14.1 **Level:** Advanced **Necessity:** Very desirable

Short Description: Gemini should provide descriptions of data processing recipes.

Description: Gemini should provide (electronically) the GSA with descriptions of the functionality of the data processing recipes described in SR7.14.1. These descriptions would be incorporated into the recipe descriptions available thought the GSA user interface.

SR7.15 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide a release date for each dataset.

Description: Gemini shall provide (electronically) the GSA with a public release date for each dataset.

SR7.15.1 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall inform the GSA when release dates change.

Description: Gemini shall notify the GSA whenever the release date for a dataset is changed.

SR7.15.2 **Level:** Key **Necessity:** Mandatory

Short Description: Inform the GSA which Gemini staff are permitted to access proprietary data.

Description: Gemini shall provide a list of Gemini staff permitted to access proprietary data. Gemini shall inform the GSA of any changes to the list of Gemini staff permitted to access proprietary data.

SR7.16 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide support for GSA users.

Description: GSA staff shall have the primary responsibility for supporting GSA users as described in SR5.5, however some questions regarding the GSA content may have to be directed to either Gemini Staff, or to the staff of one of the national project offices. Gemini shall have staff available to support GSA users.

SR7.17 **Level:** Fundamental **Necessity:** Mandatory

Short Description: Gemini shall provide version information.

Description: For each dataset, Gemini shall provide descriptors listing the versions of the telescope hardware and software in use when the data was collected. This shall include version information for any instrument or facility whose operation effects the data.

SR7.18 **Level:** Key **Necessity:** Mandatory

Short Description: Gemini shall provide test data.

Description: Gemini shall provide a variety of test data to be used to create a development and testing archive. This data should be supplied in the form described in ICD TBD. The GSA development team shall have permission to use the test data for GSA development purposes, however the data will not be accessible to the public. The schedule for providing the test data will be specified in the GSA development work scope.

2.6 External Requirements

There requirements must be satisfied outside the Gemini or GSA development or operations. Because of their external nature, it may not be possible to achieve these requirements.

SR8.1 **Level:** Advanced **Necessity:** Very desirable

Short Description: GSA needs access to SIMBAD and NED catalogues.

Description: In order to implement the object catalogue described in SR2.14, the source object databases (e.g. SIMBAD and NED) used as the basis of the object catalogue should be available to the GSA. These databases can also be used as resources for implementing other advanced features like cross catalogue searches and data mining.

SR8.2 **Level:** Advanced **Necessity:** Optional

Short Description: A unified astronomical archive catalogue should be created.

Description: In order to support correlation between the content of the GSA catalogue, and the content of archive catalogues for other telescopes, a unified catalogue of telescope archive catalogues must exist. This unified catalogue does not yet exist, and is beyond the scope of the GSA alone.

Created:March 21, 2000
Modified:November 21, 2000

Gemini Science Archive

Chapter 4 Effect of the GSA on Gemini Systems

1. Introduction

This chapter describes each of the Gemini systems which are involved with supplying the various inputs to the GSA, focusing on the aspects of the systems which must be in place in order for the GSA to function correctly.

The purpose of this chapter is to clearly identify the status of the GSA requirements on Gemini, identify the requirements which will require additional effort from Gemini staff, and to identify the Gemini components which will be involved in satisfying each requirement. The details of the formats and data transfer method will be described in ICD TBD.

Note that our intention is to include a complete list of requirements, even where we believe requirements are already implemented. The sub-sections of Section 2. list each of the Gemini components, and describe our assumptions about how the GSA requirements affect the component. As it stands, this chapter consists of assumptions made by the GSA development group. If any of these assumptions are incorrect, we would welcome input from Gemini staff.

Table 1 on page 67 lists the Gemini components which we believe support the GSA, and shows the requirements (taken from Chapter 3) which we believe affect with the components. Table 2 on page 69 lists each of the GSA software requirements from Chapter 3, showing which of the Gemini components we believe provide or should provide the functionality.

TABLE 1. Gemini system to GSA software requirement cross reference

Observing Tool
SR7.5 Gemini shall make proposal information available to the GSA.
SR7.5.1 Gemini shall provide passwords for executed proposals.
SR7.5.2 Gemini shall provide information to link proposals with datasets.
SR7.9 Supply executed science programs to the GSA.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.15 Gemini shall provide a release date for each dataset.
SR7.15.1 Gemini shall inform the GSA when release dates change.
Engineering Archive
SR7.1 Gemini shall collect instrument configuration descriptors.
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.3 Gemini shall record weather information.
SR7.4 Gemini shall collect accurate WCS information.
SR7.8 Gemini shall provide an electronic observing log.

TABLE 1. Gemini system to GSA software requirement cross reference

Observing Database
SR7.5 Gemini shall make proposal information available to the GSA.
SR7.5.1 Gemini shall provide passwords for executed proposals.
SR7.9 Supply executed science programs to the GSA.
SR7.13.1 Gemini shall provide information to identify calibration data.
Electronic Observing Log
SR7.8 Gemini shall provide an electronic observing log.
Gemini Operations Staff
SR7.6 Gemini shall provide a data quality assessment of each dataset.
SR7.7 Gemini should track publications associated with data.
SR7.10 Gemini shall provide a general description of the telescope.
SR7.10.1 Gemini shall provide a description of the instruments.
SR7.10.2 Gemini shall provide a description of Gemini facilities.
SR7.11 Gemini shall provide descriptions of data descriptors.
SR7.13 Gemini shall ensure that calibration data is available.
SR7.14 Gemini should provide data processing recipes.
SR7.14.1 Gemini should provide descriptions of data processing recipes.
SR7.15 Gemini shall provide a release date for each dataset.
SR7.15.1 Gemini shall inform the GSA when release dates change.
SR7.15.2 Inform the GSA which Gemini staff are permitted to access proprietary data.
SR7.16 Gemini shall provide support for GSA users.
Observatory Control System
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.3 Gemini shall record weather information.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.17 Gemini shall provide version information.
Data Handling System
SR7.6 Gemini shall provide a data quality assessment of each dataset.
SR7.12 Gemini shall provide raw science data to the GSA.
SR7.13 Gemini shall ensure that calibration data is available.
Telescope Control System
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.4 Gemini shall collect accurate WCS information.
SR7.8 Gemini shall provide an electronic observing log.
SR7.17 Gemini shall provide version information.
Instruments
SR7.1 Gemini shall collect instrument configuration descriptors.
SR7.4 Gemini shall collect accurate WCS information.

TABLE 1. Gemini system to GSA software requirement cross reference

SR7.8 Gemini shall provide an electronic observing log.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.17 Gemini shall provide version information.
Facilities
SR7.2 Gemini shall collect telescope configuration descriptors.
SR7.4 Gemini shall collect accurate WCS information.
SR7.8 Gemini shall provide an electronic observing log.
SR7.13.1 Gemini shall provide information to identify calibration data.
SR7.17 Gemini shall provide version information.

TABLE 2. GSA software requirement to Gemini system cross reference

Software requirement	Affected Gemini systems
SR7.1 Gemini shall collect instrument configuration descriptors.	Engineering Archive Instruments
SR7.2 Gemini shall collect telescope configuration descriptors.	Engineering Archive Observatory Control System Telescope Control System Facilities
SR7.3 Gemini shall record weather information.	Engineering Archive Observatory Control System
SR7.4 Gemini shall collect accurate WCS information.	Engineering Archive Telescope Control System Instruments Facilities
SR7.5 Gemini shall make proposal information available to the GSA.	Observing Tool Observing Database
SR7.5.1 Gemini shall provide passwords for executed proposals.	Observing Tool Observing Database
SR7.5.2 Gemini shall provide information to link proposals with datasets.	Observing Tool
SR7.6 Gemini shall provide a data quality assessment of each dataset.	Gemini Operations Staff Data Handling System
SR7.7 Gemini should track publications associated with data.	Gemini Operations Staff
SR7.8 Gemini shall provide an electronic observing log.	Electronic Observing Log Telescope Control System Instruments Facilities Engineering Archive
SR7.9 Supply executed science programs to the GSA.	Observing Tool Observing Database
SR7.10 Gemini shall provide a general description of the telescope.	Gemini Operations Staff

TABLE 2. GSA software requirement to Gemini system cross reference

Software requirement	Affected Gemini systems
SR7.10.1 Gemini shall provide a description of the instruments.	Gemini Operations Staff
SR7.10.2 Gemini shall provide a description of Gemini facilities.	Gemini Operations Staff
SR7.11 Gemini shall provide descriptions of data descriptors.	Gemini Operations Staff
SR7.12 Gemini shall provide raw science data to the GSA.	Data Handling System
SR7.13 Gemini shall ensure that calibration data is available.	Observing Tool Gemini Operations Staff Data Handling System
SR7.13.1 Gemini shall provide information to identify calibration data.	Observing Tool Observatory Control System Observing Database Instruments Facilities
SR7.14 Gemini should provide data processing recipes.	Gemini Operations Staff
SR7.14.1 Gemini should provide descriptions of data processing recipes.	Gemini Operations Staff
SR7.15 Gemini shall provide a release date for each dataset.	Observing Tool Gemini Operations Staff
SR7.15.1 Gemini shall inform the GSA when release dates change.	Observing Tool Gemini Operations Staff
SR7.15.2 Inform the GSA which Gemini staff are permitted to access proprietary data.	Gemini Operations Staff
SR7.16 Gemini shall provide support for GSA users.	Gemini Operations Staff
SR7.17 Gemini shall provide version information.	Observatory Control System Telescope Control System Instruments Facilities

2. Gemini Components

The following sections describe how we believe the GSA requirements on Gemini will affect the various components of Gemini telescope systems. The exact interface between Gemini and the GSA will be documented in ICD TBD, which will be created in consultation with Gemini. Wherever appropriate we have included what we think is the most reasonable way to implement the interface, however this is subject to change, and ICD TBD will be the definitive source for this information.

2.1 Gemini Observing Tool

The Gemini Observing Tool may cause proposal information, and science programs to be sent to the Gemini permanent store, and therefore to the GSA. The science programs will help the GSA, and GSA users to identify calibration data. The OT may already be sending proposals and science programs to the permanent store.

The OT may have a role in ensuring that a minimum level of calibration data is collected.

The OT may assign release dates to data, and may be used by Gemini staff when modifying release dates.

The Electronic Observing Log described in Section 2.4 on page 72 may be incorporated into the OT.

2.2 Gemini Engineering Archive

The Gemini Engineering Archive is a repository of information used by Gemini Staff when evaluating the performance of the telescopes. This repository will be a very useful source of information for the GSA.

The degree of interaction between the GSA and the GEA depends on GSA design. At a minimum, the GSE will be the source of telescope status, instrument status, facility status and weather information.

The requirements for the GEA described in [1] already require that the GEA collect all of the status and weather information required by the GSA. An additional requirement caused by the GSA is that this information be communicated to the GSA. The simplest way this data could be sent to the GSA would be to periodically create a data file containing the required information (defined in ICD TBD) and delivered to the GSA.

2.3 Gemini Observing Database

The Gemini observing database will be the source of various information required by the GSA:

- SR7.9 requires that Gemini provide executed science programs to the GSA. The science programs may also help the GSA identify calibration files as required in SR7.13.1.
- SR7.5 requires that Gemini provide public proposal information to the GSA. (This may come from a proposal database which is separate from the observing database.)
- SR7.5.1 requires that Gemini supply the password associated with proposals.

The simplest way for Gemini to transfer the science program to the GSA would be to create a file containing the science program (defined in ICD TBD) and sent to the GSA. This file should only be saved after the science program has been executed to ensure that the science program isn't changed before or during execution.

The simplest way for Gemini to transfer the proposals accepted for execution to the GSA would be to create a file (defined in ICD TBD) containing the public sections of each proposal to be sent to the GSA. The proposals can be sent to the GSA at any time after they have been accepted, but should be available no later than when the first data associated with the proposal arrives.

In order to support proprietary data retrieval by PIs, the proposal passwords must be available to the GSA when the first of the data collected for the proposal is received. It isn't clear what the best method of transferring the password is. The password is protected information, and so it should not be included as a dataset descriptor (we are assuming that the password must remain protected forever, since it could be used to access information which by Gemini policy is proprietary forever).

2.4 Electronic Observing Log

The EOL is an electronic log of comments from the Observer. The EOL has not yet been designed, and it may be incorporated into the Observing Tool. Gemini must determine the proprietary police for the content of the EOL.

2.5 Gemini Operations Staff

While the GSA puts many requirements on Gemini operations staff, we believe that most of these requirements are already requirements on Gemini, or they can be easily done.

2.5.1 Providing descriptions

Gemini staff are required to provide the GSA with descriptions of the meta data (FITS keywords) stored in the data files, descriptions of the telescopes, descriptions of the instruments, and descriptions of the Gemini Facilities. We assume similar documentation will be required to support telescope operations, and therefore there should only be modest additional effort required to supply this documentation to the GSA. Specifically:

- The data descriptor descriptions should be documented to allow Gemini users to interpret the data correctly. The GSA would use these same descriptions, in combination with the information in the DHS data dictionary, to document the data descriptors for archive users. The only additional requirement on Gemini staff is to supply these descriptions to the GSA as they are created or changed.
- The descriptions of the telescopes, instruments and facilities needed by the GSA should be nearly identical to the descriptions required to support proposal preparation and data reduction by Gemini queue mode observers. The only additional requirements on this documentation might be to maintain historical record of changes to the components being described (there are reasons to do this which have nothing to do with the GSA, and so this may not be a new requirement), and to provide the documents to GSA staff (in electronic form) when they are created or updated. Note that if the documentation used by Gemini meets all of the requirements of the GSA, this could be as simple as the GSA creating links to documentation available from a Gemini web server.

2.5.2 Providing data processing recipes and descriptions

The GSA has requirements to do automatic processing of data, much of which will be similar to the processing done by recipes in the Gemini On-Line Data Processing (OLDP) system, and by recipes supplied to users. If the GSA has access to the recipes and documentation produced by the Gemini staff, it will greatly simplify the creation of automatic re-processing recipes for the GSA. Gemini staff are not required to maintain backward compatibility with old data. Since recipes and documentation produced by Gemini are intended to be publically accessable, the only additional effort required from Gemini staff will be notifying the GSA when recipes are created or changed. The notification could be automated through the version control system used to store the recipes at Gemini.

Gemini Staff could also assist GSA staff producing recipe scripts with capabilities beyond the recipes supplied by Gemini. This would have a side effect of providing new recipes for use at Gemini. This assistance would be identical to that given to Gemini users when they are reducing their PI data.

2.5.3 Ensure calibration data is available

Science data in the GSA must have a minimal set of calibration data in order to be useful to archive researchers. In some cases, this calibration data may be more than the original observer

believes is necessary for his/her science program. In queue observing mode, Gemini should automatically acquire the additional calibration data. In clasical observing mode, Gemini should encourage observers to collect the

Gemini instrument scientists are the best people to define a reasonable basic calibration set for each instrument, balancing any extra effort required to collect the calibration data against the benefits of having the calibration data.

We expect that in most cases the observer will need all of the calibration data in the calibration set, so this requirement should have very little effect on telescope operations. A check should be added to the science program review process to ensure that a minimum calibration set is acquired.

In those cases where collection of all of the basic calibration data would interfere with the science goals of the PI science program, a best effort should be made to collect as much of the basic calibration data as possible.

2.5.4 Track publications associated with proposals

The GSA has a requirement to associate both PI and archive publications with observations, and we have assumed the task of identifying publications based on Gemini data will be done by Gemini staff. There will also be a requirement to communicate the list of publications to the GSA.

While the maintenance of a publication list could be an expensive task, possibly not justified by its benefits to the GSA alone, we feel that the benefits to Gemini as a whole may make this a worthwhile process. In particular, the publications based on Gemini data and their citations are the best measure of the effectiveness of Gemini in performing its primary task — producing good science. In addition to the usefulness to the GSA, the benefits to Gemini of a record of publications would include:

- documenting the scientific return on the past and continuing investment in the Gemini Telescopes made by the member countries.
- providing information to be used when planning upgrades to the Gemini Telescopes and their instrumentation. (The productivity of instruments, types of instruments, instrument modes, etc. could be measured.)
- providing Gemini management with the information required to answer the question “How does factor x relate to telescope productivity?”. (Factor x could be seeing conditions, seeing conditions vs. minimum seeing conditions permitted for the program, observing mode, observing strategy, etc.)
- provide time allocation committees with information that may help evaluate the merit of a proposal. (Determining if a given type of investigation has been effective in the past.)

The diversity of the benefits and cost of collection of the publication information are the reasons that the collection of the information is a requirement on Gemini staff, and not on the GSA operations staff.

Note that a “best effort” attempt at collecting a list of publications, which succeeded in identifying a subset of the publications would still provide many of the benefits described above, including much of the benefit to the GSA.

2.5.5 Provide a data quality assessment of science datasets

It is our understanding that Gemini already intends to do a data quality assessment for every science dataset, and so this requirement should require no additional effort from Gemini staff. Ideally, the data quality assessment should be a descriptor in the header of each dataset, however inserting a descriptor for a dataset after the dataset has been written to the Gemini permanent store may present a problem.

2.5.6 Provide release dates for datasets

Generally, release dates should be generated automatically by Gemini software, without any need for Gemini staff involvement. There will however be cases where it is necessary to modify release dates from their default values, either before or after observations are executed. This will involve a PI contacting Gemini staff to request a release date change, Gemini staff deciding if the request should be granted, and possibly entering a new release date into the system. The level of effort required to support this task will depend primarily on Gemini policy issues:

- The amount of consultation required within Gemini when making the decision to change a release date.
- The degree of encouragement given to PIs to try to get release date changes, based on the level of justification Gemini requires before changes are granted, and the past successes or failures of attempts to change release dates.

Note that these Gemini policy issues don't directly affect the GSA. The GSA needs to get notification that a release date has changed, but the process that led to the change is irrelevant to the GSA. ICD TBD shall specify how release dates and release date changes will be supplied to the GSA, including the level (e.g. dataset, proposal, or science program) at which Gemini will specify release dates and/or release date changes.

2.5.7 Gemini staff will support archive users

The GSA science staff have the primary responsibility for supporting archive users, however there will be some questions related to the data which GSA staff will be unable to answer. These questions will have to be referred to Gemini staff. The type of support required from Gemini staff would be identical to the support provide to PIs for interpreting their data. The level of effort from Gemini staff required to support GSA users should be minimal, the CADC experience with operating the HST, JCMT, and CFHT archives is that only a few questions per year must be referred to the Telescope operations staff.

Archive support questions could be directed to Gemini staff, or to support staff at the Gemini project offices of the member country most appropriate for the GSA user. This should be determined by Gemini policy.

2.6 Observatory Control System

Since the Gemini Observatory Control System is central to the control of the Gemini telescopes, it will be involved, at least indirectly in the collection of most of the information required for the GSA. It is our understanding that all of the information required by the GSA is already being collected and included in the dataset headers, with the possible exception of the version information for the various telescope subsystems.

2.7 DHS Storage Server Archive media

The DHS Storage Server is the component of the Gemini systems which creates permanent media for the PI, for the Gemini Engineering Archive, and for the Gemini Science Archive.

As it is currently implemented, the Storage Server makes two identical CD-ROMs containing all data sent to the DHS permanent store, which should include all science and calibration data produced by conforming Gemini instruments. One CD-ROM is intended to remain at the Gemini base facility as part of the Engineering Archive. The second copy is intended to be sent to the Gemini archive centre to be included into the GSA, and to serve as an off-site backup for the Gemini Permanent Store copy of the data. All of this functionality is currently available from the DHS with no additional effort required by Gemini.

If Gemini intends to deliver only data which meets Gemini data quality requirements to the GSA (as described in SR7.6, and SR7.12), then the DHS Storage Server will have to be modified to only copy a subset of the data saved to the Gemini Permanent store onto the archive CD-ROMs. In addition, since the GSA will not have a complete copy of the Gemini permanent store, alternate off-site backup of the permanent store may be necessary.

2.8 Telescope Control System

The telescope control system must provide a variety of information that will be incorporated into the GSA. This includes descriptor values, version information, and the information required to calculate accurate WCS information. We believe that all of this information, with the possible exception of the version information, is already being produced by the TCS and included in the dataset headers, and so the GSA should have very little effect on the TCS.

2.9 Gemini Instruments

In order to be fully re-usable, the data in the GSA must be fully documented. This includes recording the configuration and status of the instrument during each observation. It is our understanding that the Gemini telescopes already intend to collect this information and store it in the dataset headers.

2.10 Gemini facilities

In order to be fully re-usable, the data in the GSA must be fully documented. This includes recording the configuration and status of all facilities in use during each observation. It is our understanding that the Gemini telescopes already intend to collect this information and store it either in the dataset headers, or as separate calibration datasets. This information should therefore be available to the GSA with no additional effort from Gemini.

3. WCS coordinates

One piece of information which is considerably more important to the GSA than it is to PI users of Gemini data, is accurate WCS information. Several aspects of the GSA depend on accurate WCS information, and many Gemini subsystems must support calculation of WCS information. Any Gemini subsystem which effects the light beam during an observation may need to provide information to be used in calculating the WCS information.

It is our understanding that it is already a Gemini requirement to provide accurate WCS information for all observations, and so this should not be a new requirement.

4. Gemini Policy

The GSA requires a clear statement of several Gemini policies.

4.1 Distribution of proprietary data to PI

Our understanding is that proprietary data can be distributed to anyone with the proposal password, which will include the PI.

4.2 Distribution of proprietary data to non-PI.

Our understanding is that proprietary data can be distributed to anyone with the correct proposal password. The PI for the proposal will have the responsibility for distributing the password as he/she sees fit.

4.3 Proprietary periods for various classes of data

The following lists the various classes of data which may be subject to proprietary periods. Our understand of the current Gemini policy on public access to the data type is noted in parenthesis.

- Science data (proprietary until the public release date)
- Information derived from proprietary data, e.g. S/N, limiting magnitude, source catalogue (Generally proprietary until the public release date of the input data. There will be exceptions where Gemini staff derive information from proprietary data and designate the derived data as being public, for example, sky flats derived from science observations.)
- Telescope and instrument configuration data (always public)
- Proposal meta-data, e.g. title, abstract, keywords, etc. (always public)
- Proposal science case (never public)
- Calibration data (always public)
- Electronic observing log (?)
- Science program (?)

4.4 Gemini support for GSA users

Gemini will provide support for GSA users as described in Section 2.5.7 on page 74. Archive support questions could be directed to Gemini staff, or to support staff at the Gemini project offices of the member country most appropriate for the GSA user. This should be determined by Gemini policy. Note that from the GSA point of view, it would be simpler to rely on a single point of contact with Gemini staff, rather than having to deal with all seven project offices.

Created:May 17, 2000
Modified:November 30, 2000

Gemini Science Archive

Chapter 5 Review Comments and Responses

1. Final Review of the FPRD

The final review of this document was held at the Gemini North base facility in Hilo, Hawaii, October 30 and 31 2000. The review participants were:

- Inger Jørgensen, Jacques Peysson, Kim Gillies, Jean-René Roy, Ted von Hippel, Dayle Kotturi, Phil Puxley Gemini 8 Meter Telescopes Project.
- Felipe Richardson Comisión Nacional de Investigatión Científica y Tecnológica, Chile.
- Severin Gaudet, David Schade, Daniel Durand, Norman Hill, National Research Council of Canada.

The verbal and written comments from the review have been incorporated into the final release of this document. The only written comments follow.

1.1 Inger Jørgensen

We have discussed SR1.4.2, SR2.5.1, and SR2.14.4 internally and have decided that rather than removing the requirements totally from the FPRD, we would like a wording that specifies that only the derived data products that can be clearly/easily identified with a program id and release date on a one-to-one basis are required to be removed.

This would remove previews and catalogs originating from data where all the data have identical release dates and program id, but it would not remove previews and catalogs originating from, e.g., co-added images where the individual images had different release dates or came from different programs.

You are welcome to take the above paragraph and edit it to make it fit the style of the FPRD. As we discussed at our Hilo meeting, it should be possible to implement the removal of the derived data products identified as described above by attaching a program id and a release date to them.

2. Draft Review Comments

These are the comments from the review of the initial draft of the FPRD done on May 26 2000. Note that the document has changed significantly since these comments were written, and so it isn't always easy to relate the comments to the existing document.

Fred Gillett's comments on the comments of other reviews are noted with [fg-...].

2.1 Fred Gillett

2.1.1 **Comment:** Science Program - I would suggest that this be called the Observing Program instead, because I would usually think of the Science Program more as the scientific description or rationale for the observations, rather than the observations themselves.

Reply: The definition of “Science Program” is taken from the OT documentation. If this has changed to “Observing Program”, then we should modify our documents to match, but we should have definitions consistent with other Gemini documents.

2.1.2 **Comment:** UR1.10 It should be clear that the “calibration data” identified for each observation initially will be the standard calibration or special calibrations acquired at the time of the initial quality assessment of the observation. There may be improved calibrations, e.g. flat fields generated later that might also be associated with the observation after the fact.

Reply: We have made this clear in SR2.13, and SR7.13.1.

2.1.3 **Comment:** UR1.14 Gemini Policy is that the science case of a proposal is forever proprietary.

Reply: Agreed

2.1.4 **Comment:** SR7.7 This can only be on a best efforts basis - I see no way that we can be sure about all publications associated with Gemini observations. even beyond that, I would expect that what info we get will be publications associated with proposals, rather that specific observations.

Reply: Even a best effort will be useful to the GSA. We have made that clear in Section 2.5.4 on page 73 of chapter 4.

2.1.5 **Comment:** Section 7.2.1 Archive all Science data - This should be qualified to be all science data that meets Gemini Quality standards.

Reply: We have made this clear in SR7.12.

2.1.6 **Comment:** SR1.1 I suggest the following description. Description: The GSA must archive all science data delivered by Gemini.

Reply: Agreed.

2.1.7 **Comment:** UR1.18 We discussed this, and I think agreed that it will not be possible.

Reply: We have moved this requirement to the list of user requirements which will not be implemented (Table 1 on page 18).

SR3.4.1 It should be noted that this access should be consistent with proprietary restrictions

Reply: This is made clear in SR3.4.2.

2.1.8 **Comment:** SR3.5.1 It should be noted that this will change with time, and that all versions need to be archived and appropriate versions need to be associated with each observation.

Reply: We have made this clear in SR3.5.1 and SR3.5.2.

2.1.9 **Comment:** SR3.4.2 Need to add phrase noting that Gemini staff will have access to proprietary data for their observatory support duties.

Reply: This is made clear in SR3.4.6.

2.1.10 **Comment:** SR7.12 again, this need a phrase noting that it is raw science data that meets Gemini quality standards.

Reply: We have changed the wording.

2.1.11 **Comment:** SR7.7 Best efforts basis for linking publications to proposals. It is not real obvious to me how we can track publications based on Gemini archive data requests. At minimum we need to think about appropriate acknowledgement to be associated with use of Gemini Science archive data. The Board has established an acknowledgement, maybe we can use that for science archive data as well.

Reply: We will look at the acknowledgement required for PI publications. Archive publications should have a similar acknowledgement.

2.1.12 **Comment:** Chapter 4. We discussed this and the whole section will have to be rewritten along the lines of the concept Inger etal come up with.

Reply: This section has be re-organized.

2.1.13 **Comment:** Section 2.5.2 on page 72. The processing scripts made available to queue and classical users will also be made available to the GSA. I do not think there should be an asymmetry here.

Reply: The GSA would benefit from having access to all processing recipes developed by Gemini.

2.1.14 **Comment:** Section 4.1 on page 76. For lost or damaged data? Yes As a primary source of data for remote observers? Yes, after initial delivery As a distribution method to CoI's? Yes, via user name and password distributed as desired by PI Should include that Gemini staff have access to proprietary data as required for observatory support duties

Reply: Agreed, we have updated this section.

2.1.15 **Comment:** Section 4.1 on page 76. I think our thoughts were to have access controlled by the username/password approach. PI provides username and password to anyone he/she wants to have access to the data prior to public release.

Reply: Agreed, we have updated this section.

2.1.16 **Comment:** 3.3 Information derived from image data, e.g. S/N, source catalogue? yes, this is proprietary information other you have correct.

Reply: Agreed, we have updated this section.

2.2 Ted von Hippel

2.2.1 **Comment:** SR2.17 - is preview display useful for spectroscopy, e.g. GMOS?

Reply: Ideally, a GMOS preview might be a single FITS file containing:

- a flat/bias corrected, mosaiced, compressed image of the spectral mode image

- a compressed version of the image used to create the mask, with an overlay showing the slit positions
- FITS tables containing extracted spectra

Something like Skycat, with the addition of the ability to plot data from tables could then be used to browse the various components of the preview.

Practical considerations, like development time and money, and the size of the preview file may modify this somewhat:

- There is more than one preview product for each GMOS spectroscopic observation.
- The primary preview could be a flat/bias corrected, mosaiced, compressed image of the spectral mode image. This should be relatively easy to produce.
- A second preview product could be the mask creation image with the mask overlay. This should also be relatively easy to produce.
- Extracted spectra could be a third preview.

The processing pipeline for the spectral image and mask overlay previews should be relatively easy to produce, and with something like the Skycat "cut" feature, a GSA user could plot one row of pixels from the spectrum.

The extracted spectra preview is more difficult to produce, and could be a longer term goal for the GSA operations staff.

2.2.2 **Comment:** SR2.21 "A GSA user will be able to decimate data based on spatial coordinates." What is "decimate" supposed to mean in this context? To the Romans it meant kill 1 in 10 of your enemy's army. To modern Americans it has grown to mean "wipe out".

[fg - the meaning needs to be cleared up]

Reply: We have reworded this to "randomly sub-sample".

2.2.3 **Comment:** I don't understand what Gemini wants the GEA - GSA interface/communications to be, but it looks like we need to provide guidance in this area if it matters to us. [fg-our position is that there is not a direct GEA-GSA interface]

Reply: The details of the Gemini-GSA interface should be decided during the development of the Gemini-GSA ICD. No specific interfaces should be inferred from either the OCDD or the FPRD.

2.2.4 **Comment:** 2.7.4 - track publications - I agree with the authors of the FPRD that there are a lot of benefits to Gemini to track this information. And I don't think it will be that hard to do, although it will likely be incomplete. Gemini gets the publication info by 1) requiring all PIs to submit preprints of their papers (STScI does this and we would at least expect acknowledgement in any paper), or at least a reference once published, 2) supplement #1 with by searching the astronomical literature using ADS.

[fg-It is not real obvious to me that Gemini can commit to tracking all publications associated with a given observation or program, without devoting a lot of effort to it]

If we don't demand completeness I doubt this would be an excessive effort. The Canadian group could probably provide feedback on this as Liz Bryson does this for CFHT publications.

Reply: It is clear that even a mostly complete record of publications would greatly benefit Gemini as a whole, and should be evaluated in that context. Although important, the GSA use of such a record is not sufficient to justify the expense.

2.2.5 **Comment:** with regard to 3.1, I like the idea we discussed of a password given at the time of phase 2. The PI can then decide how to distribute this among the co-Is. I don't see a problem with a remote user obtaining their data from the GSA - in the end that might have as little delay in it as obtaining the data from Gemini. For the near term, Gemini ships data to the PI, with additional/ later recovery of data by the PI from the GSA as an option.

[fg- I think this is our current plan]

Reply: The password sounds like it should work ok, however this should be part of the Gemini-GSA ICD development process. Note that the password idea implies an all or nothing release of proprietary data for a proposal. Since there is no way to grant permission to retrieve selected observations (or to protect other proprietary information, like the proposal, science program, etc.), the only option a PI has for granting access to selected observations is to provide data to archive users directly.

2.2.6 **Comment:** for 3.2 it seems to me we already have the rule that either the PI or the Gemini Director can make proprietary data available/non-proprietary.

[fg-My understanding is that the proprietary period for scientific data will be nominally 18 months. Until that time expires, the PI is in control of who can access the data (e.g. by controlling who knows the password), except that Gemini Staff have access to all science data as necessary to carry out their support duties. Standard calibration data is not proprietary. The nominal proprietary period can be changed (shorter or longer) with approval of the Director at the time of proposal submission to Gemini.]

Reply: We have been assuming that generally public release dates will be provided by Gemini, for each dataset, which is sufficient for all of the cases mentioned. We have also assumed that there will be cases where release date will be changed (usually extended) after the science program is complete. Our experience with HST is that post execution (and even post public release) release date changes are a fact of life.

2.3 Inger Jørgensen

2.3.1 **Comment:** Glossary - "dataset" and "observation": The description of the term "dataset" is described as corresponding to the Gemini definition of an "observation". However, the result from an OBSERVE command may be several individual frames each with their own headers, e.g. NIRI frames. It is not necessarily one common header with a number of frames. I also think it will be confusing for the user if the definitions of "dataset" and "observation" do not agree between the Gemini OT and the GSA.

[fg - I have told Severin that we have to have consistent definition of terms between Gemini and the GSA. I do not have at hand the Gemini definitions of the relevant terms, but we need to agree on them, presumably at the time of the GSA video conference]

Reply: We have replaced “observation” with the term “data superset”, which is defined as either a single dataset as produced by the Gemini telescopes, or an association of datasets identified by the GSA.

2.3.2 **Comment:** Glossary - “Image quality”: I do not think that data that do not meet Gemini standards should be available in the GSA.

[fg - I think we (at least within Gemini) all agree on this, and that Gemini is the entity that makes the decision Re quality. I had a lot of trouble with this definition also, particularly since I normally think of image quality referring to a characterization of a PSF.]

As for data that meet Gemini standards but not the program requirements, then the statement that they should be used cautiously is misleading. The data may be fine for other scientific programs that the program for which they were obtained. Further, their quality in terms of image quality, sky transparency and sky background will be well-documented and available for the GSA user. [fg-yup]

Reply: We have renamed this to “Data quality” and updated the document.

Having low quality data available from the GSA isn’t a problem, as long as the user knows the data isn’t good, and has the option (possibly the default behaviour) to eliminate poor quality data from their results. Having low quality data is useful in those cases where the user is willing to deal with the problems.

2.3.3 **Comment:** SR1.11 + SR2.12: It is very ambitious (and maybe unrealistic) to provide object catalogs as part of the GSA. Is it the intent that the GSA would process, e.g., all co-added NIRI images with some object detection software and derive scientifically valuable parameters? Or is the intent just to provide cross references to already existing object catalog data?

[fg - I do not really know what the GSA crew intends in this area. In fact I do not know whether Gemini intends to do more that perhaps very basic co-adds as part of the pipeline processing. The GSA crew does recognize that this capability is of lower priority and may likely not be affordable. I certainly would be concerned if this evolved into a requirement on Gemini to do more than what is required for the pipeline data quality assessment as part of the deliverable to the GSA.

Reply: The requirement is for both a cross reference to existing catalogues (NED & SIMBAD), and extracted object catalogue (I don’t see this in the FPRD yet (it is incomplete after all), but it is in the OCDD). The cross referencing to NED & SIMBAD should be relatively straight forward (assuming accurate WCS information from Gemini). An extracted object catalogue is more of a challenge, and it will probably be something we implement for easy instruments/modes first, and work towards the hard end of the scale as the GSA development and operations budgets allow.

2.3.4 **Comment:** SR1.13: It needs to be made clear what is meant by making proposal information available for public data. Which sections of the proposal will be made public? SR1.13 and SR7.12 give the minimum requirements, but does not limit what may be made public. SR2.14 mentions that Gemini policy determines what type of proposal information is made public - what is Gemini policy on this topic?

[fg - Phil can comment better than I on this one. My understanding is that the abstract, instrument descriptors, coordinates? are public from Phase II onwards, and the scientific program descrip-

tion is never public, but would still be preserved in the GSA (I.e. available to username/password given to PI, and available to Gemini staff?)]

Reply: We have been assuming that all dataset header information and selected parts of the proposal would be public immediately as soon as the observation is complete. After a public release date, the dataset becomes public, but not the proprietary parts of the proposal.

Note that including the science program description (and science justification) in the GSA, available to users with the password, means that if a PI chooses to give their password to a GSA user for early access to their data, they also get access to this other information.

2.3.5 **Comment:** SR1.14: It is not clear to me how the field selection criteria can be described in a consistent way. I am assuming that the astronomer referred to in this requirement is the PI of the program. Is the PI supposed to describe the field in terms of “avoid galaxies bluer than some V-K”, “get as many galaxies redder than some V-K”, “observe only star no galaxies”, etc.?

Reply: Because this requirement is probably not possible, we have moved it to the list of things we won’t try to do.

2.3.6 **Comment:** SR1.15: With the GSA definition of an observation (smallest retrievable unit of data), I am wondering if the previews for each observation in the near-IR and mid-IR are useful for judging the suitability of the data. Will the GSA provide previews of co-added data - I guess so, cf. SR1.17.

[fg-for thermal IR observations, I assume that some level of co-addition and sky subtraction will be done by us as part of the pipeline processing, and maybe even as part of the observation itself. Do we need to think more about what is the smallest retrievable unit is?]

Reply: This is an interesting point. It may be that the primary unit (and minimum retrievable unit) of data in the GSA can’t correspond to any currently defined concept. For a simple optical imaging instrument, a dataset would be an adequate primary unit of data. For an IR instrument, it may be necessary to re-define the primary unit of data as a set of datasets, which have a single entry in the GSA catalogue, a single preview, and can only be retrieved as a single unit.

2.3.7 **Comment:** SR1.16: The spectral coverage and resolution will be readily available (to a certain accuracy) just from the instrument configuration. I also think this information should be publicly available for proprietary data, since it will aid the planning of other programs and help avoid duplicating observations with identical spectral coverage and resolution.

[fg-yup. I would guess that coord should also be available, prior to the end of the proprietary period for the data itself]

Reply: Agreed.

2.3.8 **Comment:** Section 2.2 (p45): Does the raw data written to CD-ROM by the DHS include data that does not meet Gemini standards? Will the raw data contain assessment of the quality before being written to the CD-ROM?

[fg-We need to discuss with Phil what happens to what data. My perspective at this point is that scientific data judged to be of Gemini quality (it may or may not be what the PI requested) would go onto some medium that would be sent to or made available to both the PI and the GSA. Data

that is not of Gemini quality could be marked and kept in the DHS store or thrown out (I do not much care - maybe there is useful and not useful reject data?), but it does NOT go to the GSA.]

Reply: Right now (speaking as a DHS developer), the DHS doesn't have the capability save a subset of data for the GSA which is different from the Gemini permanent store. This also implies that the GSA is an incomplete backup of the Gemini permanent store, raising the issue of a need for a complete backup.

2.3.9 **Comment:** Section 3.3 (p48): Data derived from the image data must also be proprietary. The GSA cannot derive object catalogs from the data and make these catalogs available prior to the end of the proprietary period.

[fg-I agree]

Only object catalogs that originate from cross referencing the field of view with existing catalogs should be available prior the end of the proprietary period. As mentioned above, I don't consider spectral coverage and resolution part of the derived data, since it will be available from the instrument configuration. In my opinion, the proposal science case (the scientific justification from the proposal) should not be made public. Only an abstract should be available, cf. my comments on the OCDD draft.

[fg-yup, I agree]

Reply: Agreed

2.3.10 **Comment:** Calibration data needs to be specified. There will be two classes of calibration data: (1) data obtained using the calibration unit, and (2) observations of standards stars.

[fg-Are you saying we should define two types of calibration data at a top level? It might even be more complicated than that. How about the case where sky background info is derived from the science observations themselves. Is that calibration data? What if it comes from some other science observation data set? Are the sky frames proprietary? Or should we be identifying some combined filtered, massaged science observations as a calibration observation that can be used by more than one program?? I think that has always been the intent of the queue concept, but we have to make sure that proprietary issues are adequately addressed].

Reply: The need for sky flats derived from science data raises a conflict with the previous comment (2.3.9), where we all agreed that data derived from proprietary data was proprietary. This is probably more of a problem for proprietary access to data for PIs, rather than users of public data, since all inputs to the sky flat will become public more or less at the same time.

2.4 Felipe Barrientos

Note that some comments were dealt with in the OCDD comments and responses.

2.4.1 **Comment:** Operating Costs. In section 5 you mention the operating costs should be less than USD$175k and it should be adjusted by the inflation in Canada. Since part of the GSA operating costs will used in Chile, the adjustments for the inflation in Chile should also be considered.

Reply: The important aspect of the effect of inflation is that the GSA should be designed for a fixed "real" operating budget, rather than a budget that is effectively shrinking due to inflation. We will remove the specific mention of Canadian inflation from SR6.2 and replace it with "infla-

tion in the host countries". We will also reword the requirement to make it clear that this is a GSA design requirement. The actual allowances for inflation will presumably be negotiated as part of the GSA operations contract, which is a can of worms we will cross when we get to it.