UNIVERSITY OF CALIFORNIA, BERKELEY SPACE SCIENCES LABORATORY

Digitally signed by Michael Raffanti
Date: 2015.06.02 16:33:31 PDT

Science Data Management Plan (SDMP)

Document #: ICN-SCI-001 Rev. C

Release Date: 05/05/15

ICON Mission

CDRL: OP-02

Document author: Scott England

Abstract: This document defines the plan for the Ionospheric Connection Explorer (ICON) data processing, data management, and data archiving system, which is effective after the data is delivered to the Science Operations Center (SOC) at UC Berkeley from the Mission Operations Center (MOC) at Berkeley. This plan describes how the mission will meet the science requirements that address the preparation, distribution, and archiving of processed science data for the general community.

APPROVALS

PI: _____

_____

Thomas Immel, ICON Principal Investigator Revision/Date

Manager: _____

_____

William Craig, ICON Project Manager Revision/Date

Systems: _____

_____

Ellen Taylor, ICON Systems Engineering Lead Revision/Date

Science: _____

_____

Scott England, ICON Project Scientist Revision/Date

Operations: _____

_____

Manfred Bester, ICON Operations Manager Revision/Date

SDC: _____

_____

Tori Fae, ICON Science Data Center Lead Revision/Date

DOCUMENT REVISION RECORD

Rev.	Date	Description of Change
DRAFT	09/11/13	Draft Release
-	10/12/13	Preliminary Release
	10/14/13	Data volumes and file format information included
	10/31/13	Updated to be consistent with driving Level 2 requirements
	01/06/14	Updates from Douglas Rowland, to define delivery schedule of data to SPDF (section 3.7), version controlling of the data and documentation (Section 3.10) and a new discussion of Metadata (Section 3.6). Updated Guiding L2 requirements to match newest revision of the L2 science requirements.
	02/20/14	Updated descriptions of Level 3 data products (3.4 removed and 3.10 description updated) to make them consistent with the Data Product Information Document
	3/6/14	Changed document title from Science data Management Plan to Science Data Management Plan to agree with CDRL document title.
	3/13/14	Updated Data Level Definitions to be consistent with NASA standard and Data Product Information Document. These changes were made in Sections 3.2, 3.3, 3.5, 3.6, 3.7 and 3.8. Table 1 that listed all data products with descriptions was removed from Section 3.2 as this is now replaced by the DPID, which is also more complete. The data flow diagram in Section 3.5 was updated to highlight both the Level 4 data and the processing of ancillary data products, which are now defined in the DPID.
	4/15/14	Updated document numbers for the other ICON documents referred to herein.
	5/13/14	Updates as per suggestions from NASA HQ: revised description of SOC / MOC components of the MOS in Section 3.3. Added additional names to the distribution list. Removed ‘shall’ statements as this is not a requirements document. Several typos.
A	6/30/14	Updates in response to input from SPDF. Updates to Figure 2, and MOS overview in Section 3.3 for clarity of responsibility, reference to MOU in Section 3.4 removed, Figure 3 updated to clarify public interface, Section 3.8.3 now clearly states Level 0 data are delivered to SPDF, description of metadata revised as per SPDF’s inputs. Updated Figure 1 for clarity – no change to content.
B	08/11/14	Updated Section 2 to be consistent with IVM A/B. Updated Pipeline figure to be consistent with Ops PDR presentation material.
C	10/15/14	Comments received from GSFC. Updates to Table of contents to include Table 1, added IVM-B to Section 2, clarified that Level 4 data are delivered to SPDF, clarified that all data will be in netCDF format with SPASE metadata, identified data quality flagging as a Science Team task

		in Section 3.3. Note that the plan for the public website is in flux, but no changes were made at this time.
D	05/04/15	Multiple updates to make this document consistent with the SOC-SPDF ICD which has been negotiated with NASA SPDF. Replaced all references to the SOS with science team (following new description in other documents including SOC-Science Team ICD), updated data volumes to reflect new telemetry rate, updated references to metadata standard, updated mission system interfaces to match SOC Internal ICD. Revised Section 3.4 to describe on the science data system interfaces – not the entire mission system interfaces, which include commands to be sent to the spacecraft etc.

DISTRIBUTION LIST

NAME, ROLE	ORG	EMAIL
Thomas Immel, PI	UCB	immel@ssl.berkeley.edu
Ellen Taylor, PSE	UCB	ertaylor@ssl.berkeley.edu
William Craig, PM	UCB	craig@ssl.berkeley.edu
Manfred Bester, Ops Manager	UCB	mbester@ssl.berkeley.edu
Douglas Rowland, Mission Scientist	GSFC	douglas.e.rowland@nasa.gov
Elsayed Talaat, Program Scientist	HQ	Elsayed.r.talaat@nasa.gov
Voleak Roeum, Program Executive	HQ	vroeum@nasa.gov
Jeffrey Klenzing, Dept. Mission Scientist	GSFC	jeffrey.klenzing@nasa.gov

Table of Contents

List of Figures & Tables

1 Introduction

1.1 Purpose and Scope

This document defines the plan for the Ionospheric Connection Explorer (ICON) data processing, data management, and data archiving system, which is effective after the data is delivered to the Science Operations Center (SOC) at UC Berkeley from the Mission Operations Center (MOC) at Berkeley. This plan describes how the mission will meet the science requirements that address the preparation, distribution, and archiving of processed science data for the public. For further details of the interface between the MOC and SOC, the reader is directed to the ICON MOC-SOC ICD.

Expected readers of this document are members of the ICON Project Office, program managers at NASA HQ, members of the ICON development and operations team, and ICON data users.

Readers are reminded that this document does not serve as the original source for technical information such as instrument specifications and telemetry formats, or high-level agreements. These technical and/or programmatic issues are defined and documented in the original documents maintained by the ICON Project Office and NASA HQ.

1.2 Relationship to Other Documents

This document summarizes and refers readers to information contained in several other documents. The information contained in this document is intended to be generally accurate; however, some documents referred to by this document are evolving works, and those documents supersede this document in case of a conflict.

Other documents referred to by this document:

- ICON L2 Science Requirements Document (ICN_SCI_003)
- ICON Data Product and Interface Document (DPID) (ICN_SCI_004)
- ICON L3 Mission Operations System Requirements Document (ICN_SYS_013)
- ICON L4 SOC Requirements Document (ICN_SYS_013)
- ICON L4 MOC Requirements Document (ICN_OPS_003)
- ICON Flight Operations Plan Vol. 3 - Mission Operations Processes (ICN_OPS_001V3)
- ICON MOC-SOC ICD (ICN_ICD_008V2)
- ICON SOC Internal ICD (ICN_ICD_008V3)
- ICON SOC-Science Team ICD (ICN_ICD_008V4)
- ICON SOC to Space Physics Data Facility (SPDF) ICD (ICN_ICD_008V5)
- ICON SOC to SOC public website ICD (ICN_ICD_008V6)
- ICON Radio Frequency ICD (RFICD; ICN_ICD_007)

1.3 Key Definitions

To Be Confirmed (**TBC**): an attribute or parameter value believed to be known with reasonable confidence and stated, but subject to further minor (5-10%) refinement.

To Be Resolved or Reviewed (**TBR**): an attribute or parameter value that is used in the requirement statement as a reasonable placeholder, but known to be in dispute until a disposition can be reached by the affected parties. May also be designated by the use of brackets [ ] around the attribute.

To Be Determined (**TBD**): an attribute or parameter value unknown at the time of writing.

1.4 Document Layout

A brief overview of the mission is given in Section 2. The data management plan is described in Section 3.

2 Mission Overview

A detailed overview of the ICON mission is given in the Mission Plan. Only a brief summary is provided here, since the primary focus of this document is the plan for processing, managing, and archiving the data after it has been delivered from the MOC to the SOC. For further details of the interface between the MOC and SOC, the reader is directed to the ICON MOC-SOC ICD (ICN-ICD-008V2). For details of the ICON telemetry and data handling inside the MOC, the reader is directed to the ICON MOC internal ICD (ICN-ICD-008V1) and the ICON Flight Operations Plan Vol. 3 - Mission Operations Processes.

ICON will explore the boundary between Earth and space to understand the physical connection between our atmosphere and our space environment. Recent NASA missions have shown how dramatically variable the region of space near Earth is, where ionized plasma and neutral gas collide and interact. This region, the ionosphere, has long been known to respond to space weather drivers from the sun, but in this century the energy and momentum of our own atmosphere have newly emerged as regularly having effects of equal or greater magnitude. ICON will weigh the impacts of these two drivers as they exert change on our space environment.

Though the solar inputs are now well quantified, the drivers of ionospheric variability originating from lower atmosphere regions are not. ICON is a single spacecraft mission that measures and analyzes these drivers. ICON presents a revolutionary concept of combining remote optical imaging and *in situ* measurements of the plasma at points where these are tied together by Earth's magnetic field. With these measurements, ICON will simultaneously retrieve all of the properties of the system that both influence and result from the dynamical and chemical coupling of the atmosphere and ionosphere. With this approach, ICON will be unique in its capability to quantify each driver and pinpoint the real cause of the variability. ICON will give us the ability to explain how energy and momentum from the lower atmosphere propagate into the space environment, and how these drivers interact with solar and magnetospheric effects during the extreme conditions of solar-driven magnetic storms.

The ICON payload consists of four instruments. The first is a pair of imaging interferometers (MIGHTI), which measure the Doppler shift of two emission lines from atomic oxygen (5577 and 6300 Å) in order to determine neutral wind speeds over a range of altitudes. In addition MIGHTI also measures the brightness of 3 components of the O_2 emission band around 7620 Å in order to determine neutral temperatures over a range of altitudes. The second is a spectrographic imager in the Far UV range (FUV) that measures the brightness of the emission from atomic oxygen at 1356 Å and from molecular nitrogen in the LBH band (near 1550 Å), from which both the daytime thermospheric composition and nighttime ionospheric O^+ density can be determined over a range of altitudes. The third is an imaging spectrometer in the Extreme UV range (EUV) that measures the brightness of two emission lines from atomic oxygen at 834 and 617 Å in order to determine the daytime O^+ density over a range of altitudes. Finally, a pair of ion velocity meters (IVM-A/B) measure the 3 dimensional vector motion of O^+ ions *in situ* (these are pointed in opposite directions and only one operates at a time). The instrument suite is designed to work together so that the remote sensing instruments (EUV, FUV, MIGHTI) are viewing a volume that is magnetically connected to the spacecraft when the spacecraft is at the geomagnetic equator, allowing ICON to compare their measurements with the local ion drift measurements of IVM, after suitable mapping of IVM measurements along the magnetic field line.

3 Data Management Plan

This data management plan addresses the requirements related to data processing, management, and archiving listed in the ICON L2 Science Requirements Document. These requirements are repeated below, for reference. The data management plan describes the elements of the ICON Mission System dedicated to meeting these requirements, including the Mission System interface with the SPDF, which is the designated data archive for the ICON mission.

3.1 Driving Requirements

L2-SCI-11a: ICON shall have a Science data Management Plan. This plan will address the total activity associated with the flow of science data, from acquisition, through processing, data product generation and validation, to archiving and preservation. This plan will address the science analysis software development, utilization, and ownership.
L2-SCI-11b: ICON data services shall be governed by an open data policy, using standard formats suitable for use in a range of available software environments, including IDL, Matlab and Python.
L2-SCI-11c: The Science data Management Plan shall be formally approved no later than the Project's Critical Design Review. A draft version of this document shall be released no later than the Project’s Preliminary Design Review.
L2-SCI-11d: The ICON Mission Operations Center shall be responsible for collecting engineering and ancillary data necessary to validate and calibrate the scientific data prior to making the data publicly available. The ICON data products will be produced by the Science Operations Center and will be calibrated in physical units or dimensions. The ICON observatory will be designed to support the necessary instrument calibration maneuvers and modes.

L2-SCI-11e: All data shall be archived in standard format for NASA missions (netCDF, CDR CDF or FITS where appropriate), along with the ICON software packages defined in the Science Data Management Plan at the Space Physics Data Facility (SPDF).
L2-SCI-11f: The first delivery of archival data shall take place no more than 6 months after the Initial Orbit Checkout is completed, and subsequent deliveries shall occur at least every 3 months thereafter.
L2-SCI-11g: The ICON SOC shall produce a comprehensive set of calibrated Level 1 data products from the ICON measurements collected by the observatory over the duration of the operational mission. The ICON SOC shall be designed produce and validate the first set of each Level 1 data product within 3 months of initial orbit checkout. The ICON SOC shall be designed to produce and calibrate all subsequent Level 1 data products within 30 days of the corresponding observations.
L2-SCI-11h: ICON shall produce a comprehensive set of inverted Level 2 data products from the Level 1 data products over the duration of the operational mission. These Level 2 data products shall be in geophysical units. All Level 2 data products shall be produced by the SOC and validated within 30 days of the corresponding Level 1 data products.
L2-SCI-11i: ICON shall produce a comprehensive set of Level 3 and Level 4 data products over the duration of the operational mission. These Level 3 data products include high-level scientific analyses of the Level 2 data products and the Level 4 data products include numerical models of the ionosphere and thermosphere. All Level 3 data products shall be produced by the SOC and ICON Science Team and validated within 60 days of the corresponding Level 2 data products.
L2-SCI-11j: All ICON Data Products will be accessible to the broader scientific community immediately upon their validation, with no proprietary period.

3.2 Definition of Data Products

Level 0 data are raw CCSDS telemetry packets processed by the MOC into 24-hour packet files per CCSDS APID, along with data quality statistics. Spacecraft and instrument housekeeping data, instrument science, ephemeris, observatory attitude and orbital state vector data are included. A full description of Level 0 data processing and data formats is provided in the Mission System ICD. See the Instrument Operations Handbook and the RFICD for definitions of the ICON CCSDS packet data.

A full description of the Level 1 through Level 4 data products is provided in the Data Product and Interface Document (DPID). All Level 1 ICON data products are produced from the Level 0 instrument files. Level 2, 3 and 4 data products are produced from either Level 1 or Level 2 data products. Figure 1 shows this flow of the ICON data products, which are color-coded by instrument or model.

Level 1 – All Level 1 data are calibrated instrument observations, in physical units (such as brightnesses in Rayleighs, rather than instrument count rates). These data are at the same instrument resolution as the Level 0 data. Level 1 data are in instrument coordinates, and are not generally georeferenced.

Level 2 –Level 2 data are inverted geophysical parameters (plasma motions in m/s, ionospheric densities in m^{-3} etc) specified at locations in geographic coordinates. These are at the highest possible resolution as determined by the instrument and inversion algorithm constraints.

Level 3 – Geophysical variables mapped on uniform space-time grids, at a common sampling rate.

Level 4 – Level 4 data are outputs from numerical models.

Ancillary data - These are georeferencing parameters, coordinates, the observatory state vector, calibration coefficients, housekeeping data etc that are used, along with the ICON observations to produce the Level 1 and higher level data products.

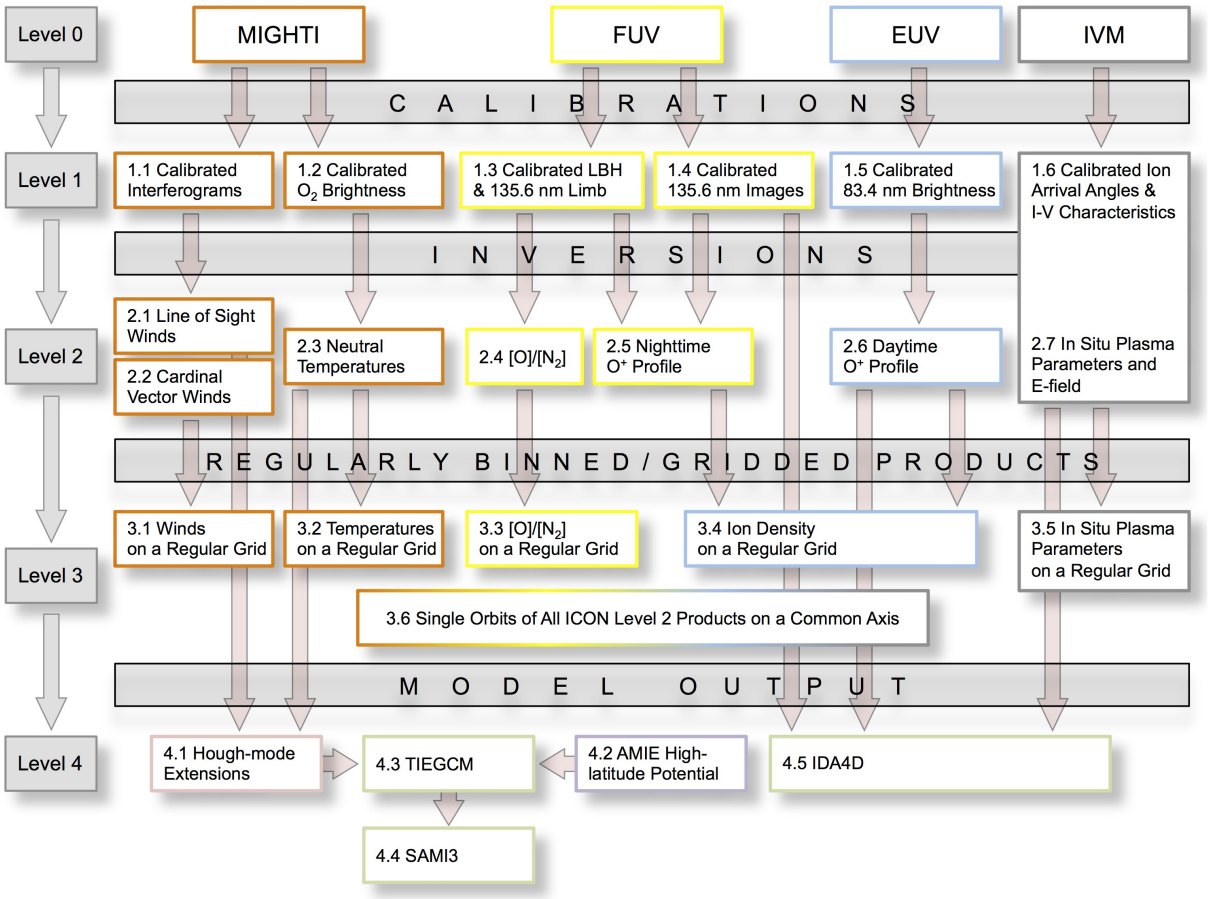

Figure 1 Flow of the ICON data products.

3.3 Mission System Overview

The Mission Operations System (MOS) architecture consists of the organizations and teams required to develop and conduct flight operations, the corresponding Ground Data System elements, and the necessary tracking networks and resources. This architecture includes the Mission Operations Center, ground stations, and the Science Operations Center, shown in Figure 1. The MOC provides the processes and functionality necessary to communicate with the SC, to plan the desired events, build and check Observatory command sequences, to command the Observatory, to receive and process telemetry and housekeeping data from the Observatory, and to analyze and submit the returned science and engineering data for archival. The MOC also interacts with the Science Operations Center (SOC) to exchange science data products and the ground stations. The SOC plans science operations, calibration maneuvers etc. and communicates these requests to the MOC. The SOC also includes the Science Data Center (SDC), which processes the science data and delivers it to the NASA archive (SPDF) and the public. The science data center uses analysis software and calibration files provided by the science team.

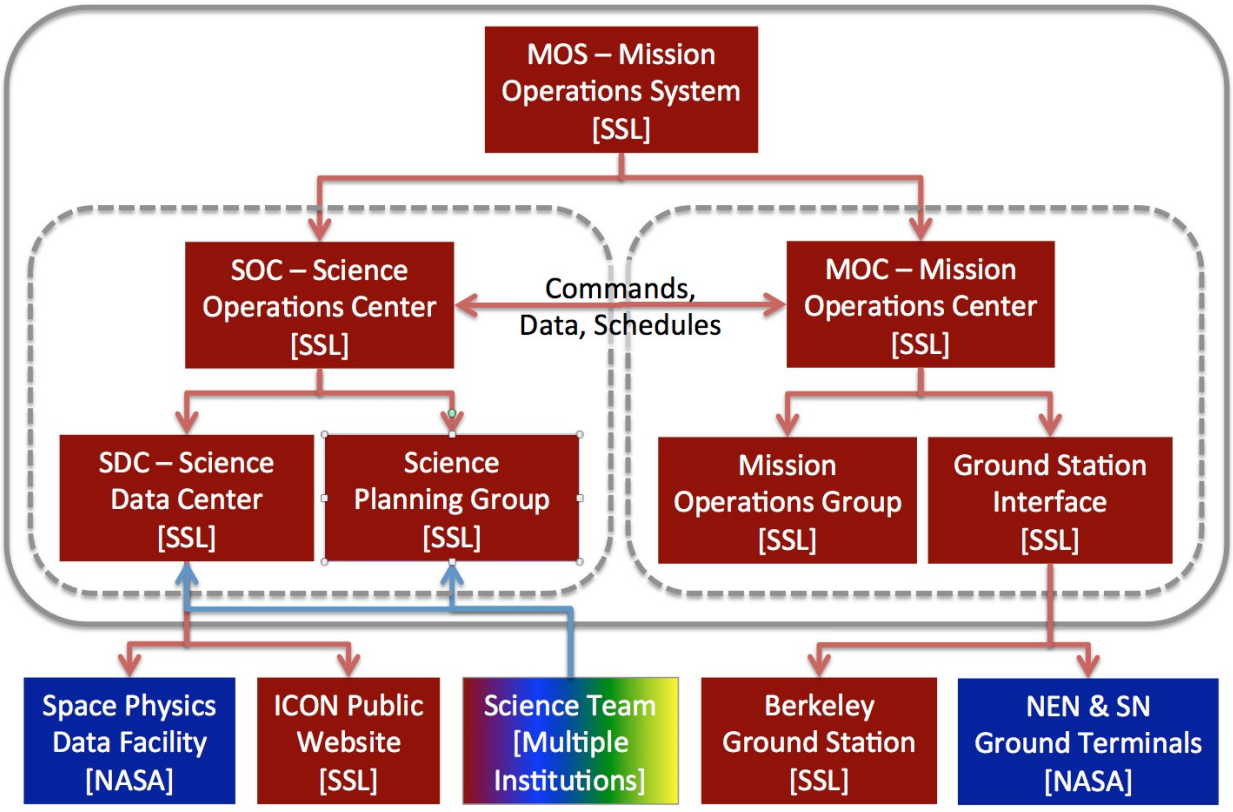

Figure 2 Mission System functional decomposition.

The MOC performs all day-to-day command, control, and monitoring of the spacecraft. The MOC responsibilities include tools to perform mission planning, ground station scheduling, CCSDS-formatted command generation, real-time spacecraft and instrument commanding, telemetry processing and archiving, spacecraft and instrument health and safety monitoring,

engineering trend analysis, orbit prediction and propagation, and data delivery to the SDC for scientific analysis and archive.

The SOC is the central hub for ICON science and instrument operations, and data processing and archiving. It is divided into two main parts – the Science Planning Group and the Science Data Center (SDC). The Science Planning Group will be responsible for turning observation plans into instrument commands for delivery to the MOC for uplink to the spacecraft, for monitoring instrument health and safety, and for processing and delivering science data products to the SPDF, the ICON science team and the public. The SDC will receive Level 0 science data and housekeeping data from the MOC over an intranet interface. The SDC will be responsible for generating and validating Level 1 and higher data products and will make those products available for the science team use on a local web server. SDC will also maintain an active archive, calibration files, and catalog accessible by the science team. The respective instrument teams, with SDC oversight, will be responsible for providing updates to the calibration files that the SDC will maintain. The SDC will maintain the interface with the SPDF by which science products, calibration files and documentation will be archived. The SOC also has a key interface with the Science Team, which provides data processing algorithms, software, instrument calibration files and observing plans to the SOC as well as the output of the AMIE model that forms Data Product 4.2. The SOC ingests these science team-generated products and ensures that they are made available to the MOC and the SPDF in appropriate formats.

The functions of the ICON SOC may be summarized as follows:

- Generate the observation and calibration schedule and predicted ephemeris, with input from the Science Team, and forward these to the MOC in a timely manner.
- Deliver instrument commands for uplink to the MOC as needed.
- Receive Level 0, Real-time, and S/C engineering and orbit data from the MOC.
- Produce all Ancillary data products identified in the DPID.
- Produce all Level 1 data files, using algorithms/software provided by the Science Team.
- Ingest and maintain algorithms, software, and calibration data from the science team, incorporate into data processing pipeline software.
- Produce all Level 2 data files, using algorithms/software provided by the Science Team.
- Maintain the calibration file repository and distribute to the Science Team for any required analysis.
- Produce all Level 3 Data Products using algorithms/software provided by the Science Team.
- Produce all Level 4 Data Products except for 4.2 using algorithms/software provided by the Science Team.
- Interface with the Science Team for the receipt of Data Product 4.2 from ASTRA.
- Distribute all data to Science Team.

- Make all ICON Level 1 through 4 data available to the public via a web interface.
- Deliver Level 0 through Level 4 data, metadata and corresponding documentation to the SPDF permanent archive, using the delivery methods, formats and schedule agreed in the SPDF-SOC ICD.
- Deliver all calibration data to SPDF archive, using the delivery methods, formats and schedule agreed in the SPDF-SOC ICD. Maintain an active archive of all levels of data produced by science team, the software for generating them, and the associated calibration files.
- Maintain a catalog of all available data and corresponding observing modes.
- Provide version control of the all Level 1 through 4 ICON data product software, data products and associated documents and metadata.
- Provide the information to the science data pipeline needed to flag potentially low-quality data based on the definitive ephermis (e.g. flag IVM data during magnetorquer actuation).

The Science Team is involved in the production of software, maintenance of calibration files, observation plans, and ICON instrument data analysis and software required for the Mission System to perform its functions. The Science Team will:

- Develop the science data processing algorithms and calibration files that will be maintained and operated at the SDC to generate the Level 1 and higher level data products;
- Generate the instrument command requests that will be ingested by the Science Planning Group and transmitted to the MOC for validation and uplink to the observatory;
- Define and develop the monitoring and verification tools necessary to verify that the performance of the instrument system meets the science requirements;
- Examine the science data and diagnostic plots produced by the SDC and validate the data, or flag it for further review;
- Produce the metadata associated with the ICON data products and distribute the accompanying documentation for all data products;
- Produce the Science Data Users Guide and deliver it to the SPDF.

The responsibilities of, and requirements levied on the MOS are described in detail in three requirements documents:

- ICON L3 Mission Operations System Requirements Document (ICN_SYS_012).
- ICON L4 SOC Requirements Document (ICN_SYS_013)
- ICON L4 MOC Requirements Document (ICN_OPS_003).

3.4 Science Data System Interfaces

The science data system interfaces are described in detail in three documents: the MOC-SOC ICD (ICN-ICD-008V2), the SOC-Science Team ICD (ICN-ICD_008V4), and the SOC-SPDF ICD (ICN-ICD-008V5). Together, these three ICDs capture the details of the data management system that will be built to execute this data management plan, and to meet the requirements laid out in Section 3.1. The contents of these ICDs are discussed here at a high level; the reader is referred to the original documents for specific details.

The MOC-SOC ICD also describes in detail the delivery of the Level 0, ephemeris and other data files from the MOC to the SDC, the format of these files, and the responsibilities for archiving these data.

The SOC-Science Team ICD describes the delivery of data files, new science processing algorithms, calibration files and new documentation from the science team to the SDC. This ICD describes the version control, archiving and distribution of all of these files.

The SOC-SPDF ICD describes in detail the data delivery interface between the SOC and the SPDF, the protocols for delivering data products over this interface, the products and documents to be delivered, and the schedule for their delivery.

3.5 Data Processing Flow, Data Volume and Data Distribution

Figure 3 illustrates the data processing flow, from receipt of raw telemetry at the MOC to release of calibrated data and software to the community. Items colored green are algorithms/software routines requiring input from the Science Team. All ICON data are shown in yellow. Items colored in purple are processes requiring input from the Science Team that is not automated (calibration of instruments, verification of data). All science data processing software used in the pipeline are provided to the SOC by the science team.

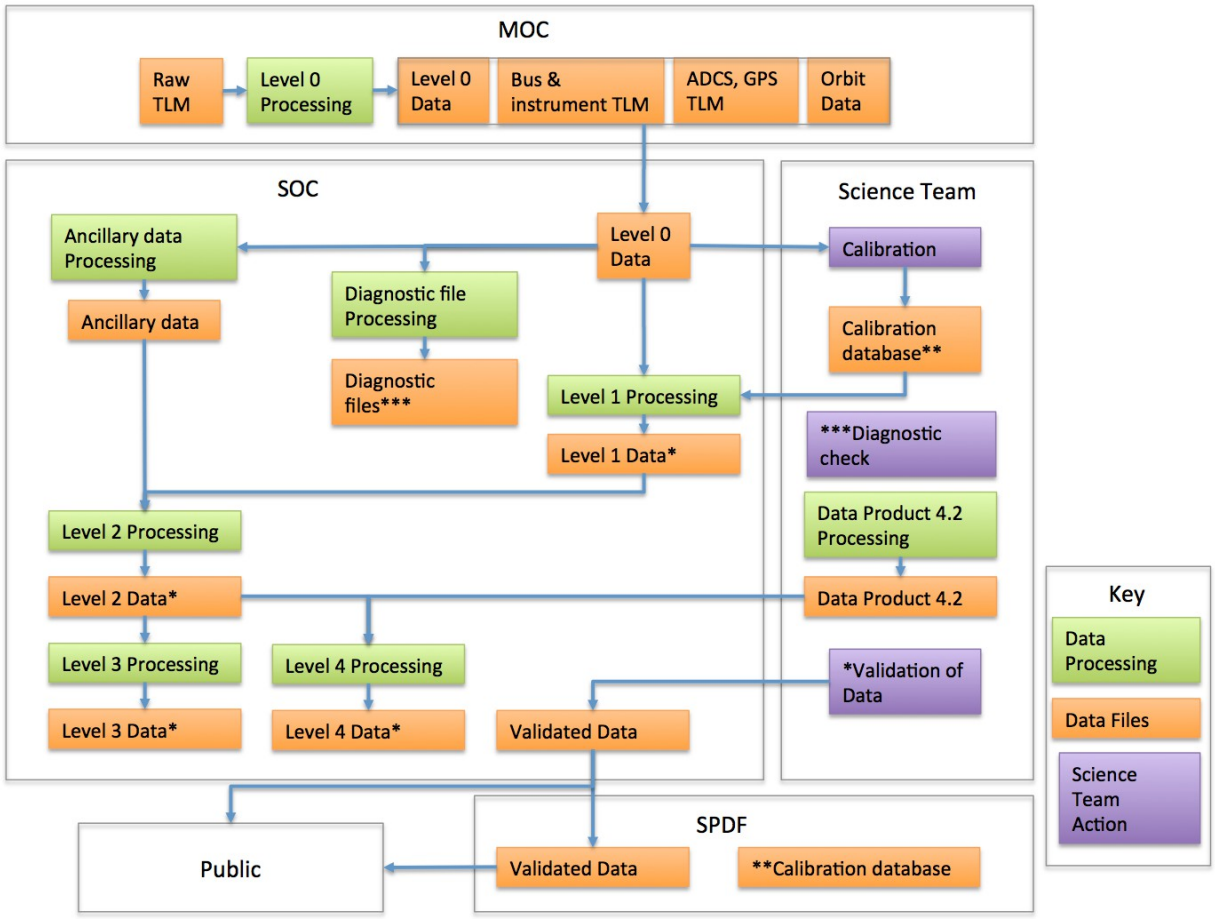

Figure 3 ICON Data Processing Flow. The * and ** identify data flow not shown by arrows, as described in the text below.

The MOC receives the raw telemetry from the ground stations. An automated process discovers gaps in the received telemetry data and automatically replays these on either the same (ongoing) pass or in the next pass. The details of this system are described in the MOC Internal ICD.

The MOC will make a subset of the L0 and ephemeris archive available to the SDC over the UC Berkeley SSL internal network. The MOC-SOC ICD provides more details of the Level 0 data interface between the MOC and the SDC.

Receipt of new ephemeris and Level 0 data at the SDC will trigger an ingest process at the SDC, where the data are added to the ephemeris and Level 0 data inventory. This also triggers the production of the ancillary data products required for processing the ICON science data, as defined in the DPID.

Starting immediately after the ingest process has completed, the new data are converted to Level 1 through Level 3 data, and added to the archive of data at the SDC. This is done using software developed by the Science Team, which is described in the Data Product Information Document.

The incorporation of these into the pipeline is described in the SOC Development plan, and their testing is described in the SOC Verification Plan.

The Level 4 data products consist of the output of physics-based and assimilative models. These data products are produced by the SDC (Data Product 4.1 and 4.3 through 4.5) and ASTRA (Data Product 4.2). All Level 4 data products are aggregated by the SDC and provided to the SPDF for archiving and distribution. The interface between the SOC and ASTRA is described in the ICON SOC to Science Team ICD.

The ICON project requires a verification phase to calibrate the instruments and validate the processing software, which starts after the observatory turn-on and checkout. The steps performed during this period to produce the first set of calibrated and validated data products, and deliver these to SPDF are shown in Figure 3. Data access during the verification and checkout period, is restricted to the ICON Team members, however once any data product has been verified, it will be made available to the public along with its corresponding metadata and documentation via the SOC website immediately (there is no proprietary period). The timeline for data production and delivery to SPDF and the public is described in Section 3.7.

During normal operation, the ICON data will be verified by the Science Team prior to release to SPDF, the science community and the public, as described in the SOC-SPDF ICD and SOC-Public Website ICD. Following their validation, all Level 1 through Level 4 data are available through the SDC website in addition to the SPDF data portal with no proprietary period. Data, calibration files and documentation/algorithms important to the correct long-term use of these data will be delivered to SPDF as these become public.

3.6 Data Formats, Metadata and Documentation

3.6.1 Data Formats

Level 0 data are raw CCSDS telemetry packets processed by the MOC into 24-hour packet files per CCSDS APID, along with data quality statistics. Spacecraft and instrument housekeeping data, instrument science, ephemeris and star tracker data are included. A full description of Level 0 data processing and data formats is provided in the Mission System ICD. See the Instrument Operations Handbook and the RFICD for definitions of the ICON CCSDS packet data.

All Level 1 through 4 data files produced by the SOC and ASTRA shall be in classic netCDF 4 format, using software issued by the Unidata program of the University Corporation for Atmospheric Research (UCAR) respectively. For Level 1 and 2 data, there will be 1 file per orbit. For Level 3, there will be 1 file per half precession cycle (~28 days). For Level 4, the duration of each file is model specific. Start and stop times for the data contained within each file will be contained within the metadata for that file.

All calibration data files produced by the SOC shall be in CDF or classic netCDF 4 format, and will follow ISTP metadata and file naming conventions conform to SPDF CALDB standard to the extent judged reasonable by the ICON team.

The Science Team will produce a Science Data Users Guide of how to analyze the ICON data and make it available to the SPDF.

Data format and organization will be negotiated with SPDF and follow SPDF guidelines to the extent judged reasonable by the ICON team.

3.6.2 Metadata

All ICON data will be accompanied by the relevant metadata required to interpret these data. Metadata are generated by the SOC and all metadata associated with each data product are detailed in the Data Product and Interface Document (DPID). Examples of such metadata include, the ICON observatory location in the geocentric reference frame, the UT of the observations, the version of the IGRF model of the Earth's magnetic field used for magnetic coordinate transformations, the model of the Earth's geoid used for translation of limb data to the geocentric reference frame and the identification of the calibration file used in producing that data product and data version number. The format of these metadata and naming conventions will generally conform to the ISTP standard.

Metadata format and organization will be negotiated with SPDF and follow SPDF guidelines to the extent judged reasonable by the ICON team.

3.6.3 Data Volume

The expected daily, uncompressed data volumes are:

- Level 0: 800 Mbytes (based on an average science data rate of 80 kpbs, engineering data rate of 8 kbps and a CCSDS overhead of 1.85 %)
- Level 1: 1 Gbytes
- Level 2: 700 Mbytes
- Level 3, 350 Mbytes
- Level 4
    - Data Product 4.1 (HMEs) 1 Mbyte
    - Data Product 4.2 (AMIE): 100 Mbytes
    - Data Product 4.3 (TIEGCM): 17 Gbytes
    - Data Product 4.4 (SAMI3): 8.6 Gbytes
    - Data Product 4.5 (IDA4D): 2 Gbytes

The total expected volume of data is 22.3 TB (uncompressed) over the 2 years of the prime mission. Data will be compressed by variable within the netCDF files. An accurate estimate of the compression ratio will be determined from simulated data.

3.6.4 File Names

The names for all ICON data files will generally follow the ISTP conventions. All ICON data files will contain the version number in the filename, and metadata.

3.6.5 Documentation

Every data product will be accompanied by a supporting document that describes the ICON team member responsible for that data product, its inputs (both data and metadata), key processing steps, any calibration steps taken (if any), key assumptions made or models used, and output (both data and metadata), including the format of that output. Each of these documents are described in detail in the Data Product and Interface Document (DPID). These documents will be available through either the ICON SOC or SPDF.

3.7 Data and Software Processing, Verification and Delivery Schedule

The ICON project requires a verification phase to calibrate the instruments and validate the processing software, which starts after the observatory turn-on and checkout. The steps performed during this period to produce the first set of calibrated and validated data products, and deliver these to SPDF are shown in Table 1. Data access during the verification and checkout period, is restricted to the ICON Team members, however once any data product has been verified, it will be made available to the public along with its corresponding metadata and documentation via the SOC website immediately (there is no proprietary period). As several days or months worth of lower-level data products are required to produce some of the higher-level data products, the duration of this verification and check-out period varies by data product level. For Level 0 and Level 1 data products, this phase shall be no more than 3 months. For Level 2 data products, this phase shall be no more than 4 months. For Level 3 and 4 data products, this phase shall be no more than 6 months.

Data, calibration files and documentation/algorithms important to the correct long-term use of these data will be delivered to SPDF as these become public, with the first regular delivery of data to occur 6 months after initial orbit check-out and subsequent deliveries as data are verified by the Science Team and made public. Minimum data delivery frequency will be once per month. The first official release of ICON data and calibration files by SPDF will coincide with the first official release of these data by the ICON project.

A preliminary delivery of example data files to test format, compatibility and the SFTP connection will occur no later than launch minus 12 months.

Step	Predecessor	Time Allowed	Cumulative time since IOC	Comment
1 TLM downlink	N/A	N/A	~5 hours	6 downlinks/day
2 TLM to L0 processing at MOC	1	~1 hour	~6 hours	
3 Transfer of L0 from MOC to SOC	2	~1 hour	~7 hours	MOC / SOC co-located

4 Generation of First In-Flight Calibration Files	3	28 days	29 days	28 days allows for full moon for EUV
5 L0 to Calibrated L1 processing at SOC	4	1 day	30 days	
6 Science Team Validates L1 data	5	60 days	90 days	Tasks performed iteratively. Meets L2-SCI-11g.
7 L1 data reprocessed if required	6			
8 L1 to L2 processing at SOC	6, 7	10 days	100 days	CPU intensive, thus 10 days are allocated.
9 Science Team Validates L2 data	8	20 days	120 days	Will commence as soon as L1 available. Meets L2-SCI-11h.
10 L2 data reprocessed if required	9			
11 First Delivery of Calibrated and Validated ICON observations to SPDF	6, 9	1 day with 59 days margin	180 days	Meets L2-SCI-11f.
12 L2 to L3 processing at SOC	8, 9	1 day	121 days	
13 L2 to L4 processing at SOC	8, 9	30 days	150	CPU intensive, thus 30 days are allocated
14 Data Product 4.2 processing at ASTRA	1	30 days	30 days	Requires external data, thus 30 days
15 Data Product 4.2 data transfer from ASTRA to SOC	14	1 day	31 days	
16 Science Team Validates L3&4 data	12, 15	30 days	180 days	Meets L2-SCI-11i.
17 L3&4 data reprocessed if required	16			
18 Second Delivery to SPDF	11, 17	1 day with 29 days margin	210 days	Meets L2-SCI-11f.

Table 1: *Steps & Cumulative Time Since IOC to Produce First Version Data Products Delivery to SPDF*

Data, calibration files and documentation/algorithms important to the correct long-term use of these data will be delivered to SPDF as these become public, with the first regular delivery of data to occur 6 months after initial orbit check-out and subsequent deliveries as data are verified by the Science Team and made public. Minimum data delivery frequency will be once per month. The first official release of ICON data and calibration files by SPDF will coincide with the first official release of these data by the ICON project.

Following the initial validation process described above, the ICON data will be processed in a routine fashion, with new releases of data made as they are validated by the science team. To ensure routine processing occurs both in a timely manner and that all outputs can be validated routinely, the task of validating the data will be delegated to the ICON Tohban or Duty Officer. The Tohban is a member of the Science Team who will be assigned for a two-week rotation to inform the SOC that each data product has been verified and is ready for release. The steps performed during this routine processing of data are described in Table 2.

Step	Predecessor	Update Frequency	Comment
1 TLM downlink	N/A	5 / day	5 downlinks/day
2 TLM to L0 processing at MOC	1	5 / day	
3 Transfer of L0 from MOC to SOC	2	5 / day	MOC/SOC are co-located
4 Update of In-Flight Calibration Files	3	1 / 28 days	1 update per full moon for EUV
5 L0 to calibrated L1 processing at SOC	4	1 / day	
6 Tohban Validates L1 data	5	1 / day	Tasks performed iteratively. Meets L2-SCI-11g.
7 L1 data reprocessed if required	6		
8 L1 to L2 processing at SOC	6, 7	1 / day	
9 Tohban Validates L2 data	8	1 / day	Will commence soon as L1 available. Meets L2-SCI-11h
10 L2 data reprocessed if required	9		

11 L2 to L3 processing at SOC	8, 9	1 / day	
12 L4 processing at SOC	8, 9, 10	1 / 27 days	27 days of data required for 4.1
13 Data Product 4.2 processing at ASTRA	1	1 / 7 days	ASTRA will review external data inputs weekly
14 Data Product 4.2 transfer from ASTRA to SOC	13	1 / 7 days	
15 Tohban Validates L3&4 data	10, 14	1 / day	Meets L2-SCI-11i.
16 L3&4 data reprocessed if required	15		
17 Deliver Calibrated / Validated data to SPDF	6, 9, 15	1 / month (min)	Meets L2-SCI-11f.

Table 2: *Steps Required & Frequency of Updates for Validated of ICON Data Products Delivered to SPDF*

3.6 Data Archiving and Preservation

3.8.1 SOC Active Archive

The SOC archives all Level 0 through Level 4 data products, calibration data, ancillary data, metadata, data processing software and associated documentation for the life of the mission, as per the SOC L4 requirements document. These requirements state, among other things, that the SOC data archives shall be backed up on a regular basis with no more than 2 days between backups, and that the data archive backups shall be stored in a physically different location than the archive.

3.8.2 SPDF Archive

All ICON Level 0 through Level 4 data, as well as the ancillary data and calibration data required for producing the ICON data products, and associated documentation will be archived at the SPDF according to established SPDF archiving practices.

3.7 Version Control for ICON Data, Software and Documentation

All ICON data, the software to produce it and the accompanying documentation that describes the data will be version controlled at UC Berkeley using the most appropriate software for each task (described below). Updates to any of the ICON data analysis software or documentation can be requested by any member of the Science Team using an ICON Ground Systems Configuration Change Request. Access to each of the version-control systems described below will be restricted to the relevant members of the Science Team. Users of the ICON data outside of the ICON team will generally only access the most recent data and accompanying files. To allow users of the ICON data outside of the ICON team to be notified of to the ICON data processing software and data products, a change log will be maintained at the SOC and posted on the SOC website that provides a brief description of the updates to the ICON data products, software and documents. Interested parties will be able to subscribe to an email distribution list to receive updates of this change log as they are made. This change log file will be made available to SPDF.

3.9.1 ICON Data

Updates to the ICON data associated with reprocessing of any Level 1 through Level 4 data (e.g. in response to an updated calibration file or processing software change) are communicated to the ICON science team members responsible for those data products and any higher level products associated with those (as defined in the DPID) via a daily email update. Once any new data file has been produced, subsequent reprocessing of any higher-level products associated with that file occurs automatically at the SOC in response to the production of that lower-level data file (for example, production of a new Level 1 product automatically starts the update of the associated Level 2 through Level 4 data products). A complete log of all updates is maintained in a release notes file at the SOC website. Only the newest versions of data files are visible to users of the SOC data. All old versions of data files are archived offline at the SOC throughout the duration of the ICON mission.

Notification of any update to the list of files on the SFTP site will be provided to SPDF via a file on the SFTP site, which will contain the name and directories of files updated in plain text.
All ICON data files will contain the version number in the filename, and metadata.

3.9.2 Documentation

All documentation supporting the ICON data products will be held under configuration control using a library within the ICON Subversion (SVN) system that is used to version control ICON systems engineering documents. SVN (Subversion®; http://subversion.apache.org) is an open source version control system that manages files and directories and tracks changes made to them. SVN allows older versions of data to be recovered and provides a history of how data has changed over time, including information on who made the changes when. The ICON SOC and SPDF will disseminate the newest approved version of any supporting documentation.