Quality Assurance
(Mike Freilich - Data Quality Control Panel)
- Working Definition (evolving)

QA = what can be done in an algorithm context.

e.g. spectral checks
spatial checks
temporal checks
Time period ~ 1 day

Validation - All other tests

Test versus: in-situ
models and
other products
[ocean → models to couple ocean codes,
SST, winds, ...]
comparisons with similar products
from other sources.

QA Draft Def

QA- procedure to identify/flag pixels to granules which "obviously do not conform to expected accuracy"

+

"performed within operational processing window"

Then use QA to:

- Monitor health of product
- USE QA of input products
- Delivered to IDS (extensive QA)
  General user (general QA)

QA PLAN - Submit between

(informal) Beta - Version 1 dates Jan 97

formal - Version 2 mid 97

outline/generic plan defined - 1/96
(from Data Quality Control Panel)

Draft plans submitted 6/96

Review workshop 9/96

Finalized Draft 12/96

PLAN INFO

Bob Lutz 301-286-7339

RLUTZ@LTPMAIL.GSFC
.NASA.GOV

Vets Quality Arrurance 13-2.1

Ed ...

Please find draft of an E-mail and a draft of the fourth version of the QA Plan ...Bob

DRAFT

To : QA Advisory Team
From : Bob Lutz / ESDIS Science Office
Re : Fourth draft of QA Plan
Date : May 1 1995

This mailing concerns the latest effort in regards to the specification of QA procedures and methodology for EOS products. As has been indicated within a previous E-mail, this issue has transitioned over to the ESDIS Science Office (H. K. Ramapriyan) with myself continuing to serve as the focal point of contact. Please find enclosed the fourth version of the QA Plan. We have tried to incorporate the suggestions (and to placate the objections) that were raised during reviews of the third draft (October 1994). It should be noted that there are several important differences between this and the third draft, which are summarized below.

- 1) Through discussions with the Chairman of the Data Quality Control Panel (Mike Freilich) and the EOS Validation Scientist (Dave Starr), a working definition of what constitutes QA is proposed. Using this definition, Step 4 (User QA) is dropped from the previous plan and Steps 2 (DAAC QA) and 3 (SCF QA) are now strongly limited. We realize that this definition may not please everyone, but we hope that all can "live" with it (Please see Sections 1.0 and 2.0 for a further discussion of this issue).

- 2) This draft also discusses a procedure to include suggestions and comments from other groups who may be generating QA within EOS and/or who would be interested in the subject of quality assurance. The interdisciplinary investigator (IDS) teams may generate QA data within their processing of higher level EOS products. In addition, these teams may have preferred formats and content of the QA that they would like to receive with the standard EOS products. These desires, if identified early enough, may be able to be accommodated within the operational QA

(Attached 13-2.1)

software of the science teams. Another group, the general science community may also be concerned about quality assurance, such as the content and the organization of the archived QA metadata. Early comments from this group will enhance the potential long-term usefullness of the archived QA data.

- 3) We are now proposing a schedule for the writing of the science team QA Plans that would generally coincide with the needs of QA information for the science team software deliveries and the ECS system releases. The development of these plans would be a three step iterative process: a preliminary scoping of roles and responsiblities for the Beta delivery, an informal (draft) QA Plan generated before Version 1 and a formal deliverable QA Plan due before Version 2. Schedules are also presented on how input from the IDS investigators (through their QA Plans) and the science community (through a panel chosen from interested participants) could be incorporated into the development of the general QA methodology.

- 4) The AHWGP may coordinate the creation of the above mentioned QA Plans, as well as coordinate the completion of those plans. The fourth draft, contained here, has a strawman QA Plan within it, that may or may not be used for this purpose.

- Again this office needs your feedback on this QA approach, especially to sections that pertain to your specifc area of interest, (Science teams and DAACs - Section 2, IDS teams - Section 3, and the User Community - Section 4). Please send your comments to Bob Lutz, rlutz@ltpmail.gsfc.nasa.gov (301-286-7339) by June 1 if possible. If the overall reaction from the group is positive, we will proceed as outlined here.

Thank you for your continued involvement.

Schedules

Introduction :

- Schedules are presented for the science teams, the IDS investigators and the user community. The schedules are basically a two step iterative process. During the first stage of the process, data is gathered (through QA Plans and questionnaires) independently from each group, outlining their QA procedures and needs. The data that is received will be compiled by this office, and then exchanged amongst the groups. A workshop will be convened with representatives from all groups participating in the formulation of a QA approach. The second stage will (hopefully) see each group somewhat modifying their individual plans to accommodate the wishes of others. A well thought out (and well represented) QA approach for EOS should be the result of this process.

A) Proposed Schedule for the Science Teams :

Introduction :

- It is realized that the development of QA methodology for the science teams will be an evolutionary process, as the algorithms mature and lessons are learned from the implementation of earlier versions of the software. We are therefore recommending a three stage iteration of these plans; requesting the science teams provide at first a general description of their QA methodology, then a draft of their QA approach, and then a formal QA Plan.

1) QA Procedural Plan for Beta Delivery

- At the present time, the ECS contractor (Hughes) has only a generic plan of how QA will be take place (in an operational sense) between the SCFs and their DAAC(s). Through discussions with Hughes personnel, it has been expressed that it would be beneficial for the development of the system, if some preliminary QA information, (such as the outlining of general roles and responsibilities of the DAACs and the SCFs and a realistic estimate of data rate flows between the DAACs and the SCFs), could be provided as soon as possible. It is realized though that at this time, some of the science teams may not be ready to provide such "realistic" scenarios and estimates. We request then that they would provide their "best guess" at this time, and note within their documentation where the uncertainties may exist.

Proposed schedule :

The general part of the strawman QA Plan would be completed by the science teams and the DAACs - 9/95
Possible inclusion of above information into ECS IR-1 release - 9/95

2) QA Plan for Version 1

- The Version 1 software delivery to the DAACs is now scheduled for 1/97. Quoting a statement from the science team working agreements "Programs shall demonstrate all major functional capabilites and a complete operator interface, including the generation of all needed messages using standard error and message services." We realize that within this version the science teams may not be able to provide details to specific content of their QA products, but it is anticipated that they will be able to indicate "in general" what QA elements they will include in their algorithm processing package. Also, it is envisioned that they should be able to incorporate the "hooks" needed to capture and process incoming QA streams from other science teams. It is our belief that it would be of great benefit to the science teams if they knew before Version 1, what the other science teams were doing in terms of QA (type and content), so that these "hooks" could be more realistically simulated. In addition, if possible suggested QA requirements indicated by the IDS investigators may be able to be accommodated within their software at this time (ie: within Version 1). Additional input from the science community may also be useful to the preliminary scoping of what portion of the operational QA will be archived.

We therefore are recommending that Draft QA plans be generated by the science teams between the Beta and Version 1 releases, with enough time left after the generation of such plans, for a workshop to be convened for an exchange of information. This workshop may be in conjunction with other common subject areas. After the workshop, "finalized" draft QA Plans may be generated.

Proposed schedule :

Format and content of "generic" Science Team QA Plans defined and- circulated to science teams - 1/96
Draft Science Team QA Plans submitted to ESDIS Science Office - 6/96
Draft Science Team QA Plans, Draft IDS Plans, User QA Panel comments circulated amongst all groups - 7/96
Workshop Convened to review the QA Approach - 9/96
"Finalized" Draft QA Plans -12/96

3) QA Plan for Version 2

- Version 2 is defined as "This version shall be a launch ready, complete, verified and operational software system". The approximate date of this release is 9/97. The finalized Draft QA Plans would be circulated amongst the user community panel and the IDS investigators for comment during the early part of 1997. In addition, it is anticipated that data dependent science teams will need to know several months in advance of this date, detailed QA characteristics from the other teams. With the feedback from the Draft QA plans as well as lessons learned from the implementation of Version 1, final QA Plans would be generated. These would be formal deliverables on the part of the science teams. A review would be done of these Plans to check for inconsistencies and incompatabilities amongst the science team Plans and also the IDS Plans.

We recommend Final QA plans be submitted to the ESDIS Science Office between the Version 1 and Version 2 timeframe, again with enough time alloted so that feedback could be incorporated within the Version 2 software.

Proposed schedule :

Final QA Plans submitted and circulated for review- 4/97
Final QA Plans reviewed and inconsistencies resolved - 6/97

B) Proposed Schedule for the IDS Teams

Introduction :

- The IDS teams would be notified, through their respective land, ocean and atmosphere division chairmen (or through a presentation at one of their joint meetings) that there is a desire within the project for their input to QA Procedures within EOS. Abbreviated QA Plans would be developed and completed by the IDS teams. A strawman plan is provided in the fourth draft. This would be coordinated by the AdHoc Working Group on Consumers.

Proposed Schedule :

Notification of IDS teams for their planned involvement - 6/95
Finalization of Generic IDS QA Plans - 12/95
Draft IDS QA Plans Completed - 6/96
Draft QA Plans, IDS Plans, User QA Panel comments circulated amongst all groups - 7/96
IDS representatives at Workshop - 9/96
"Finalized" Science Team Draft QA Plans circulated to IDS investigators for comment -1/97
Final IDS QA Plans completed - 4/97

C) Proposed Schedule for the Involvement of the User community :

Introduction :

- A panel should be formed of researchers that would be typical of the user science community. A possible source of these "type" of users may be members of the DAAC User Working Groups (UWGs), who are knowledgeable of EOS, yet most are not aligned with any specific instrument or product. They may also be able to recommend some graduate students for involvement in this panel, as many of the members of the UWGs are university folks. The AdHoc Working Group on Consumers has also been suggested as a source for members. After the panel is formed, a questionaire would be formulated, soliciting them for their comments on the proposed content and grouping (next item) of the archived metadata.

- The possible grouping of the products into categories for purposes of defining keyworks within the metadata should be coordinated by the ECS contractor. It may be possible to follow the methodology that they have undertaken to group the products in terms of data struture.

- There may be a possiblity that some of the suggestions regarding the content and organization of the QA within the metadata may impact developments within the ECS contractor's Release A. The proposed schedule would allow enough time for these comments to be incorporated.

Proposed Schedule :

Data products grouped by ECS- 12/95
User QA Panel formed - 12/95
Questionnaire prepared and given to User QA Panel -1/96
Questionnaires returned (User QA Panel comments)- 6/96
Draft Science Team QA Plans, Draft IDS Plans, User QA Panel comments circulated amongst all groups - 7/96
Possible input of User QA Panel to ECS Release A - 9/96
Representatives of Panel to attend Workshop - 9/96
"Finalized" Drafts of Sci. Team QA Plans provided - 12/96
Comments Due regarding Science Team Drafts - 2/97

DRAFT

Quality Assurance Procedures For EOS Products—Concepts, Implementation and Archival

1.0 Introduction

Quality control procedures of EOS products and the archival of quality control data within the product metadata are important aspects of EOS and EOSDIS. The EOS Data Quality Panel is presently proposing that quality control consist of four entities : calibration, monitoring, quality assurance and validation. This document describes the concepts and suggested implementation of quality assurance procedures for EOS products, as well as recommended archival guidelines. Quality assurance may be defined as a process whose objective is to identify and flag data products, at the granule or smaller level, which obviously and significantly do not conform to the expected accuracies for the particular product type (proposed definition by the EOS Data Quality Panel). Within an operational context, our office is recommending that in addition to the above definition, that quality assurance be also (and non-technically) defined as any quality control process that could be done (either by the software processing algorithms, DAACs or the SCFs) within the operational time-window of producing EOS products (ie : before archival at the DAACs).

A distinct seperation between quality assurance and validation borders almost on a "religious argument", with many researchers firmly believing in their positions and total agreement will not be be possible within the EOS community on the subject of this division. Some common agreement does pervail though. QA does not entail calibration aspects of quality control, which instrument control personnel will use to monitor the health of the instrument and to analyze instrument errors as they occur (defined as calibration). QA does include automated flagging within the science algorithm software and in addition "may" involve human or automated interpretation (ie : on-line checking at the DAACs and/or the SCFs) of these generated flags. Our office recommends that the latter part of

the previous sentence only pertain to procedures that may be done within the real-time window. Validation is defined as all other quality control processes; including long-term analyses (time series analysis, regional data analyses, detailed quality control flag analysis, etc ....) by the DAACs, SCFs and scientific researchers who are utilizing the data within their studies.

1.1 Scope

The overall purpose of this document is to develop a coordinated approach in the application of quality assurance methodology within the generation of EOS products, as well as to develop a rational method for archiving the statistics. It is realized that quality assurance products may be used by several "types" of users :

- 1) The science teams will use QA data for the monitoring of the "health" of their data products. It is conjectured that some of this data may be "internal" and not stored within the metadata of the product.

- 2) Data dependent science teams will need supplemental QA information from the other science teams (data providers) in order to process their own products. It is envisoned that some of this incoming QA will be operational; in otherwords QA that is generated by the data provider to monitor their (data producers') own product generation, but will not be stored within the metadata of the product.

- 3) Interdisciplinary teams who are generating higher level EOS products within the operational program. This class of scientists are defined as the funded researchers within the EOS program who will be responsible for generating level 3 or 4 products. It is envisoned that these scientists may need more extensive QA than the next class of users (the general science community), but less QA than the science teams generating the operational standard products.

- 4) The overall science community, which would represent scientists who would use EOS data for general research purposes, not for the generation of EOS products. Their needs for QA may be quite different than the above groups, in that QA may be principally used to "screen" data for its potential usefullness. It should be noted that there is a strong possibility of "overkill" in the archiving of QA within the metadata (ie : the saving of data that is neither wanted - or eventually used - by the general science community). This group may provide recommendations pertaining to the characteristics of the archived metadata (ie : what, how much, and in what resolution of the generated QA should be archived in the metadata).

To ascertain the requirements of each of these categories of users as well as to surface any potential conflicts and inconsistencies, this document outlines an approach for gathering the information needed for such an analysis.

1.2 Background

Generally, before the EOS/EOSDIS era (i.e., before 1990), detailed QA procedures had been incorporated into the operational processing algorithms after the launch of the satellite. This methodology was at many times ad hoc and incomplete. From a user's point of view, the organization and the content of such QA statistics within the archived data product, also left a lot to be desired. In addition, with the requirement that EOS products conform to an HDF standard, the appendage of QA (flags and generated statistics) into the product metadata may not be a trivial task, unless some forethought is given to the procedure and the space (within the file) necessary for the process. It is hoped that by defining a quality assurance approach early in the development of EOS and defining the needs of the general user community within this area, many such problems and shortcomings might be avoided.

1.3 Overview, strategy and document organization

This plan is divided into three sections.

- Section 2 discusses a three step methodology in which the EOS science teams, in conjunction with their associated DAACs, will ensure quality assurance of EOS standard data products.

- Section 3 describes two strawman QA Plans. The first plan is one in which the science teams, with their DAACs, would describe the operational methodology and the content of QA for their EOS standard products. This plan contains general and detailed sections that outline their intended QA procedures. Within the detailed section there is an opportunity for them to describe their projected QA requirements from other science teams, in terms of incoming EOS standard data products to their data stream. The Ad Hoc Working Group on Production (AHWGP) has agreed to coordinate the completion of these plans. The

final plans would be submitted to the ESDIS Science Office for coordination and review purposes. The second strawman plan is a shortened version of the above, in which the IDS teams would be requested to complete. This plan outlines their intended QA procedures, as well as their desired QA data needs from the science teams generating the EOS operational products.

In summary, completion of these plans would allow :

- A clarification of the respective roles of the DAACs and the science instrument teams with regard to QA. This would enable both entities to plan better in their development and resource allocation.

- The opportunity for the instrument teams and DAACs to modify their individual QA plans after surveying what other instrument teams are planning in this area (i.e., allow teams to learn from one another).

- The ability for data-dependent EOS instrument teams (ie : teams receiving EOS standard products from another team) to review how the received EOS-QA could be used in their processing algorithms.

- The opportunity for IDS teams to analyze and comment on the QA statistics that are intended to be generated and stored by the science teams and the possibility that these suggestions could be incorporated by the science teams within their processing streams.

- Section 4 presents a discussion of the development of a common methodology for the incorporation of QA results into the metadata. A hierarchal approach is recommended with the additional suggestion that it may be possible to define common QA statistics for "types" or groups of EOS products. This section will be highlighted for review by the general science community. While it is realized that this community can not lever QA requirements on the developers of the products, it is felt that this group should have "input" into topics such as :
  -the content of the subset of the operational QA that are archived
  -the organization of the QA data within the metadata.

A successful completion of this activity would :

- Enable the ECS contractor to plan ahead in the design of the QA metadata within the HDF data structure.

- An opportunity for the user community to comment on the types of QA that may be generated for a product before the QA is actually implemented.

2.0 Definition of the Quality Assurance Procedure (within the processing of standard EOS products)

Quality assurance of EOS products may consist of one or more of three possible steps:

- 1) Automated QA within the processing algorithm software.
- 2) QA performed by DAAC personnel, in consultation with the science teams, after the product is generated, but before it is archived
- 3) QA done by the SCFs on either complete or portions of the products, before archival.

All QA would be done within the operational time frame window (ie : in real time and before a product is archived at the DAAC). It is believed that this time period may be different from one product to another. The science teams and the DAACs would specify these operational times. It is recognized that some science teams are co-located with their DAACs, so that Steps 2 and 3 may be combined. For the purposes of a common methodology of incorporating the QA data within the metadata though, it is desirable to keep the Steps seperate. Furthermore, it is realized that some science teams may consider QA to be only the first step. Therefore, their QA Plan would only consist of the first step. Please keep in mind, that this office is only presenting the above as guidelines, not rules.

The QA process should be regarded as evolutionary in nature, in that the roles and responsibilities of the science teams and the DAACs may change as the algorithms become more robust and the system stabilizes. In addition, it should be realized that simply because a data product passes through a certain step in the QA process, this does not guarantee a stamp of approval by the processing entity. It indicates only that the data has passed through some certain predefined test. For example, if a product passes through the automated QA contained within Step 1 successfully, this will not imply that the SCF has "certified" its accuracy or correctness. It will mean only that the data has passed through a certain filter (for example boundary checking).

2.1 Possible Implementation of QA Methodology Within Each Processing Step

The quality assurance procedure is defined as a consisting of three steps, as defined above. All data products will be expected to pass through some form of Step 1 QA, with the possibility that portions

(or all) of the data products would be analyzed in Steps 2 and 3. As an overall concept, it is recommended that each QA step build upon what has preceded, examining a subset of the previous QA information. For example, DAAC QA (Step 2) may emphasize monitoring of the automated QA statistics (Step 1), and calling any questions to the instrument science team's attention. SCF QA (Step 3) may include analysis of automated and DAAC QA.

It should be noted that the following are only suggested scenerios.

- **2.1.1 Step 1: QA within the science team algorithm processing software.**

Implementation of QA analysis: Within the science team algorithm processing software, initial QA can be "built in" during the routine processing/generation of the data. The QA defined here should be written by science team personnel, incorporated within the science processing algorithms, and performed at the DAAC that is processing the data. This QA would be completely automated and be performed on the data as they are being processed.

QA analysis: QA performed at this step should catch large-scale algorithm or processing errors of the data products. As a minimum, the QA activities that should be performed within this step are the identification and flagging of missing data, and boundary checking. Simple statistical data may be generated at this step. For example in the case of image-type data, the mean and standard deviation of scan lines or scenes may be calculated. Ancillary QA products (secondary products) may also be derived at this step, such as standardized graphs and plots. It has also been suggested that QA flags could be organized along the lines of constraint thresholding, where thresholds would be determined for physical, algorithmic, and climatological constraints.

Storage/archival of QA analysis: QA statistics/flags may be incorporated within the metadata on a datum (point by point) or granular basis. Storage of QA by a datum basis could consume significant archive resources. It has been suggested that, in order to save computer storage space, some teams may opt to simply incorporate a flag within the metadata indicating that this specific data has been quality checked at this step. A user would then have to request the QA processing software/algorithms if interested in obtaining the actual QA statistics.

- **2.1.2 Step 2 : QA done at and by DAAC personnel after the EOS products are generated, but before archival.**

Implementation of QA analysis: QA flags or the generated products themselves may be "pulled" by the DAACs via the subscription service for QA analysis at the DAAC. In general, the DAACs' QA role would be to ensure that the data are generated within the quality specifications defined by the science teams. An additional role of the DAAC is to ensure the integrity of the data—i.e., that data are not corrupted in the transfer, archival, or retrieval process. Specifically, the role of the DAAC may involve the monitoring of QA statistics generated within the software mentioned in Step 1. As has been envisioned by a DAAC reviewer, an IST (Instrument Support Terminal)-like workstation may be set up that allows the DAAC operator to monitor the automated QA stream with dynamic graphical (plots or image-based) outputs. These secondary products may or may not be saved at the DAAC. In addition, the ECS contractor has suggested that a DAAC Quality Assurance Monitor (QAM) be assigned to each DAAC.

QA analysis: Criteria of "good" versus "bad" QA statistics would be defined by the science teams, with the DAACs alerting the science teams when the data indicate that there are problems in the data products. Large-scale error checking may also be done at this step. Through selective subsampling (for example, every eighth data point) or averaging, a "sanity check" can be done on the data product. In addition, the QA defined in this step may take the form of qualitatively comparing "yesterday's" (or the first complete set of older) data, with "today's" data. These QA procedures may take the form of analysis of visual or imaged data. This type of QA would involve consultation with the science teams (maybe a science team representative located at the DAAC), and it is envisioned that this type of QA would likely involve a mixture of a person and automation.

Storage/archival of QA analysis: Flags developed at this step may be decided to be consistent at all DAACs for the same level products (and maybe for all products). This would eliminate the need for the user to understand (for example) hundreds of quality flags for hundreds of EOS data products and parameters. If this is not possible, it is hoped that individual science teams could develop

consistent flags within the generation of their own science team data production. If this is not done, an online quality flag dictionary should be developed, so that users could easily interpret these flags.

The responsibilities of QA functions that may be done at the DAAC (Step 2) versus those performed by personnel at the SCF's (Step 3), need to be defined by the science teams and their respective DAACs. These roles will be discussed within the QA plans generated by the science teams.

- **2.2.3 Step 3: QA done at the SCFs by the science team members.**

Implementation of QA analysis: QA flags or the generated products themselves may be "pulled" by the SCFs via the subscription service for QA analysis at the SCF. This would require a transfer of some of the data processed at the DAACs to the instrument teams. Within the QA Plan, network transfer rates are requested.

QA analysis: Techniques such as visual analysis, subsampling or other statisitcal techniques may be used here.

Storage/archival of QA analysis: Where human analysis is performed, descriptive text may be generated (for example through visual analysis of the data). This descriptive text may not be stored within the HDF file, but elsewhere in an adjacent file that would be associated with the data product. With appropriate flagging within the product metadata, the user would be made aware of the existence of this additional information. The SCF would send these QA flags or statistics to the DAAC, via the ingest service, where they would be included within the product metadata.

3.0 QA Plans

Strawman QA plans are presented in this section. The first plan presented is
to be completed by the science teams, in consultation with their DAACs.
These plans will cover the QA procedures for the production of the EOS
standard products. The second plan is to be completed by the interdisciplinary
teams outlining their QA methodology and needs. Both are generic in nature,
with the intention here being that plans that are organized in the same
format will be easier to compare.

- Science team Plan :

The first part of the plan is general and would likely be the same for all products generated by that science team. It includes a general description of QA roles of the science teams and their DAACs. The latter part of the QA Plan is detailed and will probably vary from one product to another. The final section of this plan provides the opportunity for the science teams to indicate the QA statistics that they would desire from the other science teams. It is envisioned that the plans will be evolutionary in nature, with changing roles of processing entities as the system becomes more stable. It is hoped that this concept could be woven into the plan. The finalization and completion of these plans may be coordinated by the AHWGP and would be submitted to the ESDIS Project Science Office. There would be three stages of completion of these plans, corresponding to the software deliveries of the science teams.

- IDS teams plan :

These plans are an abbreviated form of the science team plans. This will most likely be coordinated by the Ad Hoc Working Group for Consumers and would be circulated amongst the IDS teams for completion.

3.1 - QA Plan for Science Teams and DAACs

- General :

1) The definition of the prelaunch QA process on simulated data.

2) General description of the responsibilities of the science teams and the DAACs within the complete procedure of the QA process. This high-level view of the QA process should also attempt to address the evolutionary nature of the QA process (i.e., how the roles of the science teams and the DAACs may change in time as the system stabilizes and the algorithms become more robust). A brief operational scenerio would also be very beneficial within this section.

3) The percentage of each data product that will be transferred between the DAAC and the SCF for QA purposes. An overall transfer rate (ie : for all products) between the science team and their DAAC(s) would be an alternate specification.

- Specific :

A) For each step in the envisioned QA process (this may be different for each product):

1) The overall methodology of the QA process (i.e., statistical, visual....).
2) The expected percentage of the data product that would be examined within this step.
3) All of the parameters/results generated from the QA process and how they should be interpeted. (i.e., types of flags, variables calculated, etc. ....).
4) Hardcopy data generated from the QA process and what portion of this type of data would be saved.
5) The parameters/results from 3 that are expected to be stored in the metadata. This could be all or a subset of the generated QA.
6) The response to the QA process
7) The expected timeframe for the QA process.
8) The resources needed/expected for the QA process. This would include computational, financial, and people-power requirements.

Also, a prioritization of the QA process if funding is limited.

B) Desired QA from other science teams generating EOS products (ie : data incoming from other science teams in the operational time window)

1) Name of Science team and Product

a)Desired QA statistics

b)Desired resolution of QA statistics (ie : by data point, granule ...)

3.2 - QA Plan (abbreviated) for IDS teams:

Please note : The purpose of the first three sections are only to place in context the last two parts.

- 1) General description of the QA procedure for generated products.

  Description of activities (if any envisoned) to be done by DAAC

  personnel.

- 2) The parameters/results generated from the QA process (i.e., types of flags, variables calculated, etc. ....).

- 3) Hardcopy data generated from the QA process and what portion of this type of data would be saved.

- 4) The parameters/results from 2 and 3 that are expected to be stored in the metadata.

- 5) Desired QA from the science teams generating the standard EOS

  products (ie : QA data incoming from the science teams)

  - Name of Science team and Product

    Desired QA statistics

    Desired resolution of QA statistics (ie : by data point, granule ...)

4.0 QA Metadata (Organization and content)

As discussed in the introductory section, QA is only one part of the total quality control process. It has been planned by the ECS contractor that "all" quality control related information would be stored in a common place within the metadata. Unfortunately, in ECS documentation this is referred to as QA metadata. Please note the difference between the limited definition of QA here and the much broader definition used by the contractor.

- QA organization :

It is recommended that a common approach be developed for the inclusion of QA results into the metadata. This would provide users with a consistent format in their understanding and interpretation of projectwide QA.

The incorporation of QA into the metadata would be hierarchical in nature. This would allow the user of the data the opportunity to examine, in more and more detail, QA aspects of the data. In addition, storage space would saved within each QA "region" of the HDF, so that all three levels (or steps) of QA could be accommodated in the same area.

The final methodology of how to include QA into the metadata must be adaptive enough to accommodate a changing QA data stream, because it is anticipated that QA procedures (statistics, flags, etc.....) will change during the life of the project.

- QA content :

It is envisioned that the majority of EOS standard products could be grouped into common "data types" (ie : image data, flux data, vertical profile data, point source data, etc ...). With each data type, certain general QA characteristics would be defined and key-words could be chosen. For example; within image data, a scans' line mean and standard deviation may be thought to be of importance for archival purposes. The determination of these data types and their related key words would be defined with help from the general science community.

An attempt should be made by the science teams and the DAACs to develop a common set of flags for the quality assurance process. If this is not possible, it is recommended that each science team develop a common QA flagging methodology for their own respective products.

4.1 Strawman Model for Inclusion of QA Into HDF Metadata

A strawman model is presented here only for discussion. It is applicable for data of a particular image-type format. It is envisioned that other "types" of data would have somewhat different formats, but all following the general conceptual idea of being hierarchal and grouping the three steps of QA results together.

The model presented has been adapted from the QA methodology developed by SeaWiFS. Other data products would need different models. (If a reviewer has a better model to illustrate the concept, please contact me.)

Part A presents a schematic diagram of the model, principally highlighting the organizational structure of the metadata. Part B provides specific details of how this model may be applied and presents suggested QA data for this data "type".

Part A : Schematic diagram of strawman model (organization)

1) Granule 1
.....1.1) QA metadata
.....QA metadata for granule 1 (3 levels)
.....QA metadata for scan line 1 (3 levels)
.....QA metadata for scan line 2 (3 levels)
.....QA metadata for scan line 3 (3 levels)
.
.
.
.
.....1.2) Data for granule 1

2) Granule 2
.....2.1) QA metadata
.....QA metadata for granule 2 (3 levels)
.....QA metadata for scan line 1 (3 levels)
.....QA metadata for scan line 2 (3 levels)
.....QA metadata for scan line 3 (3 levels)
.

.
.
.
.....2.2) Data for granule 2
.
.
.

Part B A Specific Example (QA content)

Please note that this is just one of many scenarios that may be developed. I have chosen this "type" of data because I am the most familiar with it.

At the granule level, one would find the QA metadata. (Corresponds to 1.1 in schematic). A granule (or a scene in this case) is composed of Y scan lines of X pixels each. There are Z granules in an orbit.This QA metadata would be composed of two parts: QA parameters associated with the granule and QA parameters associated the individual scan lines.

- **Granule QA parameters:** Within the associated QA metadata for a granule, one would find three groups of QA flags and parameters, corresponding to the three steps in the QA process. Suggested parameters and flags that may be found in this metadata are:

1) **Algorithm processing QA:**
Total number of scan lines
% of missing scan lines
% of filled scan lines
total number of pixels
% of missing pixels
% of filled pixels
% of out-of-bounds pixels
% of cloud-covered pixels*
Overall QA flag (criteria need to be determined),
Flag indicating that there is descriptive information elsewhere.

*Please note that it has been correctly pointed out by a reviewer that "cloudy" means different things in different parts of the spectrum. A

more useful term may be whether the data can be processed or not due to cloud contamination.

2) DAAC QA :
Was this granule examined (Yes/No flag)?
Overall QA flag for granule (criteria needs to be determined)
Flag indicating that there is descriptive information elsewhere.

3) SCF QA :
Was this granule examined (Yes/No flag)?
Overall QA flag for granule (criteria needs to be determined)
Flag indicating that there is descriptive information elsewhere.

The first scan line within this granule would now be examined.

- **Scan line QA parameters:** A scan line consists of X pixels. Within the metadata for a scan line, one would find three groups of QA flags and parameters, corresponding to the threesteps in the QA process.

Suggested parameters and flags that may be found in this metadata are:

1) Algorithm processing QA:
% of missing pixels
% of filled pixels
% of out-of-bounds pixels
% of cloud-covered pixels
Statistics: mean, standard deviation, minimum value, maximum value
Array with QA flagged pixel IDs and associated QA flags (whether pixel is out of bounds, filled, cloud covered, missing).

2) DAAC QA:
Format same as above

3) SCF QA:
Format same as above

After Y cycles (corresponding to the number of scan lines), one would find the data (X by Y pixels) (Corresponds to 1.2 in schematic).

As noted above, this data stream will be repeated Z times, corresponding to the number granules in an orbit