# Data Quality Assessment Database and a World-Wide Web Interface

C. Tullos, C. Cox, R. Jedrzejewski
October 30, 1997

**ABSTRACT**

*Contact scientists examine observation products and report problems and shortcomings in the data. We have developed a database to hold this information and a series of World-Wide Web pages to facilitate reporting directly into this database.*

## 1. Introduction

Up until August 1996, HST science data was routinely inspected by OPUS staff to look for instrumental or observatory anomalies. The presence of such anomalies, such as bad diodes, guide star acquisition problems, missing calibration files, was recorded in the PDQ (PODPS Data Quality) file and keywords populated in DADS to record the information in a form accessible to archive researchers using Starview.

On the recommendation of the Data Quality Project (Sparks et al. 1995), this function was transferred to the Contact Scientists in SSD. The main benefit of this was to enhance the end-to-end support provided by the Contact Scientists, since they are most familiar with the science goals and technical aspects of their proposals, and are thus most qualified to assess the quality of the data.

Rather than examining the data on a computer screen (as had been done in OPUS), the CS inspects the paper products and looks for certain anomalies. Depending on the nature and severity of the anomaly, the PI is contacted and the anomaly brought to their attention. However, this leaves a gap because there is no way to record this information for those who might retrieve the data from the archive.

For this reason, the Data Quality Coordination Committee (the descendant of the Data Quality Project) instigated the creation of a database for storing the CS Data Quality Assessments, and a WWW tool to facilitate the assessment process. The tool & database have been in routine use since early December 1996.

## 2. The DQ Database

### *Database Tables*

One table for each instrument contains information queried from the DADS database on a nightly basis. These data are observation parameters based on 'IPPPSSOOT' dataset-names, archive_class specifications, instrument name, and receipt times.The second set of database tables are the data quality assessments written as an end result of this web tool.

What follows below is a listing of all of these relations, the field names and associated data types:

### FOC_DADS

| receipt_time (datetime) | duration_time (float) | datasetname (varchar 21) | obset_id (varchar 21) | pep_id (int) |
|---|---|---|---|---|
| program_id (varchar 21) | start_time (datetime) | target_name (varchar 41) | optical_relay (char 3) | cam_mode (char 3) |
| smm_mode (char 3) | sht_mode (char 3) | px_formt (char 3) | filter1 (char 8) | filter2 (char 8) |
| filter3 (char 8) | filter4 (char 8) | line_num (char 15) | pi_name (varchar 55) | instrument (varchar 12) |

### FOC_DQ

| cs_name (char 24) | assessment_date (datetime) | datasetname (char 9) | poor_targ_pos (char 24) | hi_background (char 8) |
|---|---|---|---|---|
| nonlinear saturat (char 9) | eight_bit overflow (char 8) | psf_degrad (char 8) | misc (char 8) | text_comment (text) |

## FOS_DADS

| receipt_time (datetime) | duration_time (float) | datasetname (varchar 21) | obset_id (varchar 21) |
|---|---|---|---|
| pep_id (int) | program_id (varchar 21) | start_time (datetime) | target_name (varchar 41) |
| grndmode (char 18) | detector (char 5) | aper_id (char 3) | polar_id (char 1) |
| fgwa_id (char 3) | line_num (char 15) | pi_name (varchar 55) | instrument (varchar 12) |

## FOS_DQ

| cs_name (char 24) | assessment_date (datetime) | datasetname (char 9) | few_counts (char 4) |
|---|---|---|---|
| crowded_field (char 4) | guide_star (char 4) | corr_prob (char 4) | other_cause (char 4) |
| noisy_diode (char 4) | uncorr_dd (char 4) | unusual_cosm_ ray (char 4) | unusual_bgrd_ corr (char 4) |
| flt_fld_prob (char 4) | unusual_scat lt_corr (char 4) | flx_calib_not_ app (char 4) | unusual_calib_ spec (char 4) |
| text_comment (text) | | | |

## GHRS_DADS

| receipt_time (datetime) | duration_time (float) | datasetname (varchar 21) | obset_id (varchar 21) |
|---|---|---|---|
| pep_id (int) | program_id (varchar 21) | start_time (datetime) | target_name (varchar 41) |
| obsmode (char 18) | detector (int) | grating (char 9) | aperture (char 8) |
| maxwave (float) | minwave (float) | line_num (char 15) | pi_name (varchar 55) |
| instrument (varchar 12) | | | |

## GHRS_DQ

| cs_name (char 24) | assessment_date (datetime) | datasetname (char 9) | quality_flag (char 7) |
|---|---|---|---|
| targ_acq (char 9) | blemish (char 4) | deaddiode (char 4) | noisydiode (char 4) |
| flakydiode (char 4) | spikes (char 4) | missingdata (char 4) | filldata (char 4) |
| falselock (char 4) | trackerror (char 4) | carr_reset (char 4) | expo_short (char 4) |
| misc (char 4) | text_comment (text) | | |

## WFPC2_DADS

| receipt_time (datetime) | duration_time (float) | datasetname (varchar 21) | obset_id (varchar 21) |
|---|---|---|---|
| pep_id (int) | program_id (varchar 21) | start_time (datetime) | target_name (varchar 41) |
| filtnam1 (char 8) | filtnam2 (char 8) | mode (char 4) | imagetype (char 18) |
| line_num (char 15) | pi_name (varchar 55) | instrument (varchar 12) | |

## WF_DQ

| cs_name (char 24) | assessment_date (datetime) | datasetname (char 9) | bias_ripples1 (char 4) | bias_ripples2 (char 4) |
|---|---|---|---|---|
| bias_ripples3 (char 4) | bias_ripples4 (char 4) | horizontal smearing1 (char 4) | horizontal smearing2 (char 4) | horizontal smearing3 (char 4) |
| horizontal smearing4 (char 4) | residual_image1 (char 4) | residual_image2 (char 4) | residual_image3 (char 4) | residual_image4 (char 4) |
| contamination_worms1 (char 4) | contamination_worms2 (char 4) | contamination_worms3 (char 4) | contamination_worms4 (char 4) | stray_light1 (char 4) |
| stray_light2 (char 4) | stray_light3 (char 4) | stray_light4 (char 4) | dragon_breath1 (char 4) | dragon_breath2 (char 4) |
| dragon_breath3 (char 4) | dragon_breath4 (char 4) | ghosts1 (char 4) | ghosts2 (char 4) | ghosts3 (char 4) |
| ghosts4 (char 4) | ota_reflections1 (char 4) | ota_reflections2 (char 4) | ota_reflections3 (char 4) | ota_reflections4 (char 4) |
| moving targets1 (char 4) | moving targets2 (char 4) | moving targets3 (char 4) | moving_argets4 (char 4) | missing_data1 (char 4) |
| missing_data2 (char 4) | missing_data3 (char 4) | missing_data4 (char 4) | text_comment (text) | |

## 3. The WWW Tool

This web tool has been coded using Perl and SQL within a CGI scripting format. The database is a SYBASE database using SQL. The procedures and functionality of the tool allows for several approaches to the task at hand. We do, however, suggestion a particular method for assessing, but as the user gains experience in using the tool, several paths may be found leading to successful results

Using Netscape, enter the URL - http://www.stsci.edu/dqcc/webtool/dq_top.html. The **Assessor Web/Database Interface** page should come up.

Enter a contact scientist username, password, instrument, and proceed to the **Observation Selection Form**, making using of any single selection criteria. The program fully supports any combination of individual selection criteria.

There is an action- item choice, that of viewing the data or viewing and assessing the data. The "*View Data Only*" option allows the user to look at data in the '*InstrumentXXX_dads*' database, not the actual science data. Typically, a user would want to view and assess the data at the same time. At that point, click on '*Select Matching Observations*'. Once a database retrieval has occurred, the **Matching Observations** page appears as shown:

The **Matching Observations** Page gives more information. In the first column of every data row, the user is given the assessment status for each. Either a "YES" or "NO", indicating whether or not each dataset-name has been assessed. The next column tells the user whether or not the datasetname has been assessed as "OK". An "OK" for any datasetname indicates that there were no problems for any assessment consideration. The "OK" status functions as a default assigned assessment.

At this point, there are abundant choices on what can be done.

1. *View Existing Assessments* will display whatever assessments have been recorded in the database of the original selection criteria. This is simply a view of the database.

2. The *Assess Marked Datasets* choice requires the user to mark one or more datasetnames and in clicking on the "*Continue*", the program will lead to the assessment page. This page will be covered more fully later.

3. The *Select All To Assess* choice allows the user to mark all datasets as having a particular anomaly, making it unnecessary to click on every dataset to select it. Click on "*Continue*" to procede to the Assessment Page.

4. The *Assess Remainder* choice assumes at least one observation on the page has been assessed and will simply select the remainder of the observations (those not assessed) and "*Continue*" to the Assessment Page. This choice will often be used when the anomalous datasets have been assessed, and all that is required is to mark the rest as "OK".

5. The *Add comment* choice will allow the user to add a text comment to every datasetname which has come from the original query of the database. It will append to any existing comments. This is for those cases where one comment applies to all datasets that might have different individual assessments. The assessments are not changed by this choice, only a comment is added to the comment field.

For example, choosing "*Assess Marked Datasets*" and marking a few datasetnames, and clicking on "*Continue*", will cause the **Assessment** page to come up:

### *Assessment Page (Predefined Anomalies)*

The contact scientist has a pre-defined selection of known anomalies with each instrument. Therefore, the tool differs in selection options according to instrument.

The values recorded in each DQ table for these pre-defined anomalies is either "OK", indicating that there were no problems for a particular anomaly, and in most cases, the value of "YES" would indicate the existence of the anomaly. More descriptive values have been used instead of "YES" where warranted.

The following four pages shown are the **Assessment Pages** for each instrument. The pre-defined anomalies vary from instrument to instrument and are outlined here.
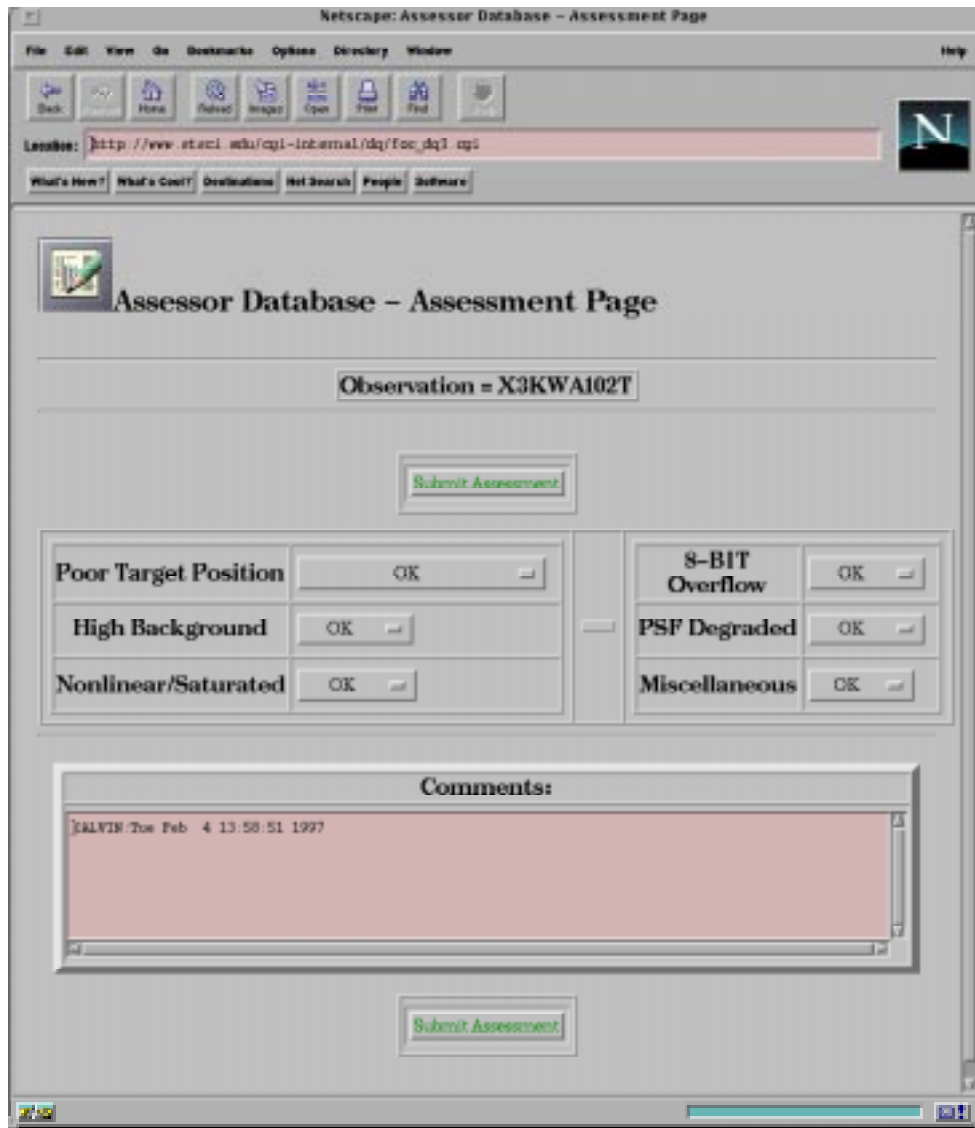
GHRS

| Quality Flag | Target Acquisition | Blemish | Dead Diode | Noisy Diode | Flaky Diode |
|---|---|---|---|---|---|
| Spikes | Missing Data | Fill Data | False Lock | Track Errors/ Loss of Lock | Carrousel Reset |
| Exposure Shortened | Miscellaneous | | | | |
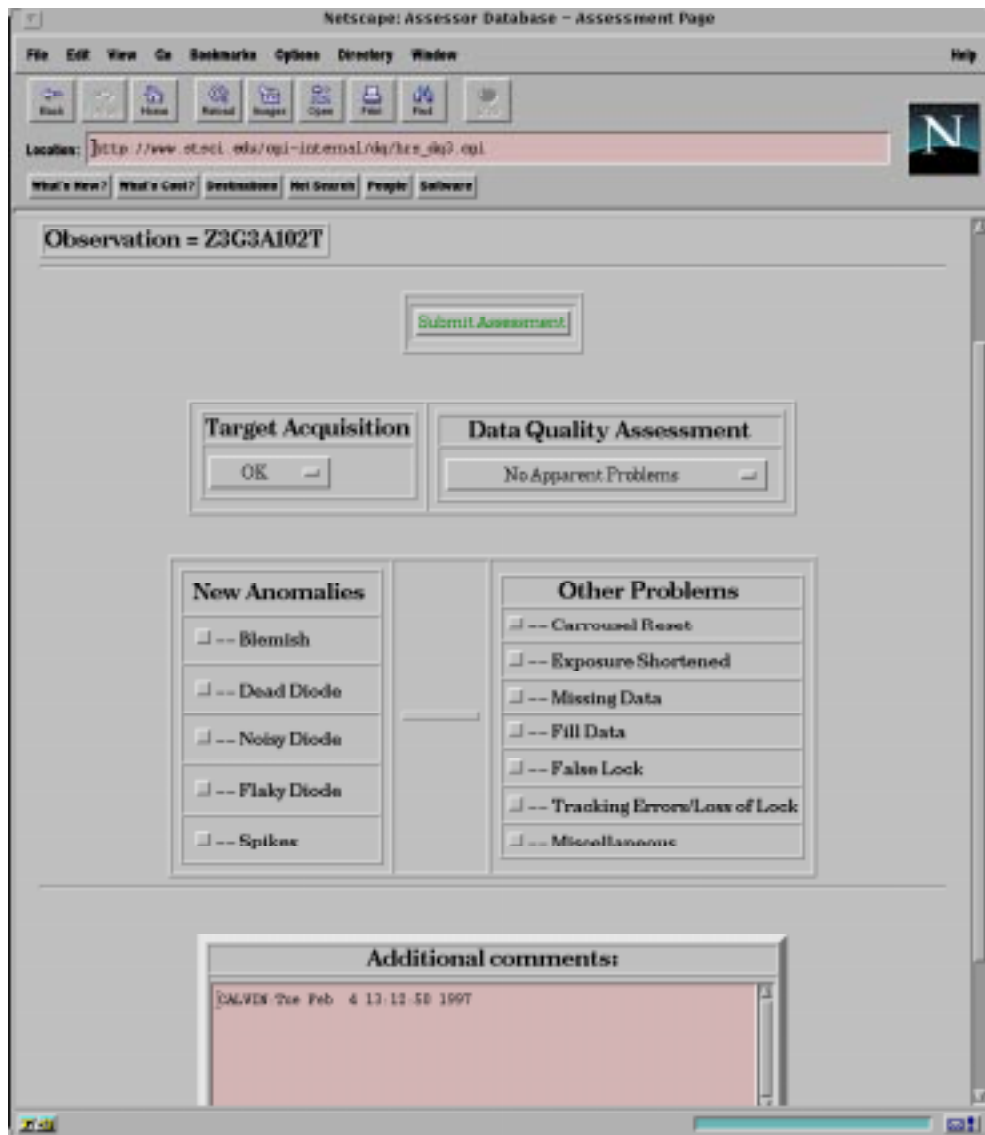
**FOC**:

| Poor Target Position | High Background | Nonlinear/ Saturation |
|---|---|---|
| Eight-Bit Overflow | PSF Degraded | Miscellaneous |

**FOS**:

| Too Few Counts | Crowded Field | Guide Star Aquisition Failure | Possible Coordinate Problems | Other Cause | Noisy Diode |
|---|---|---|---|---|---|
| Uncorrected Dead Diode | Unusual Cosmic Ray Hit | Unusual Background Correction | Flat Field Problem | Unusual Scattered Light Correction | Flux Calibration Not Applied |
| Unusual Calibration Spectrum | | | | | |
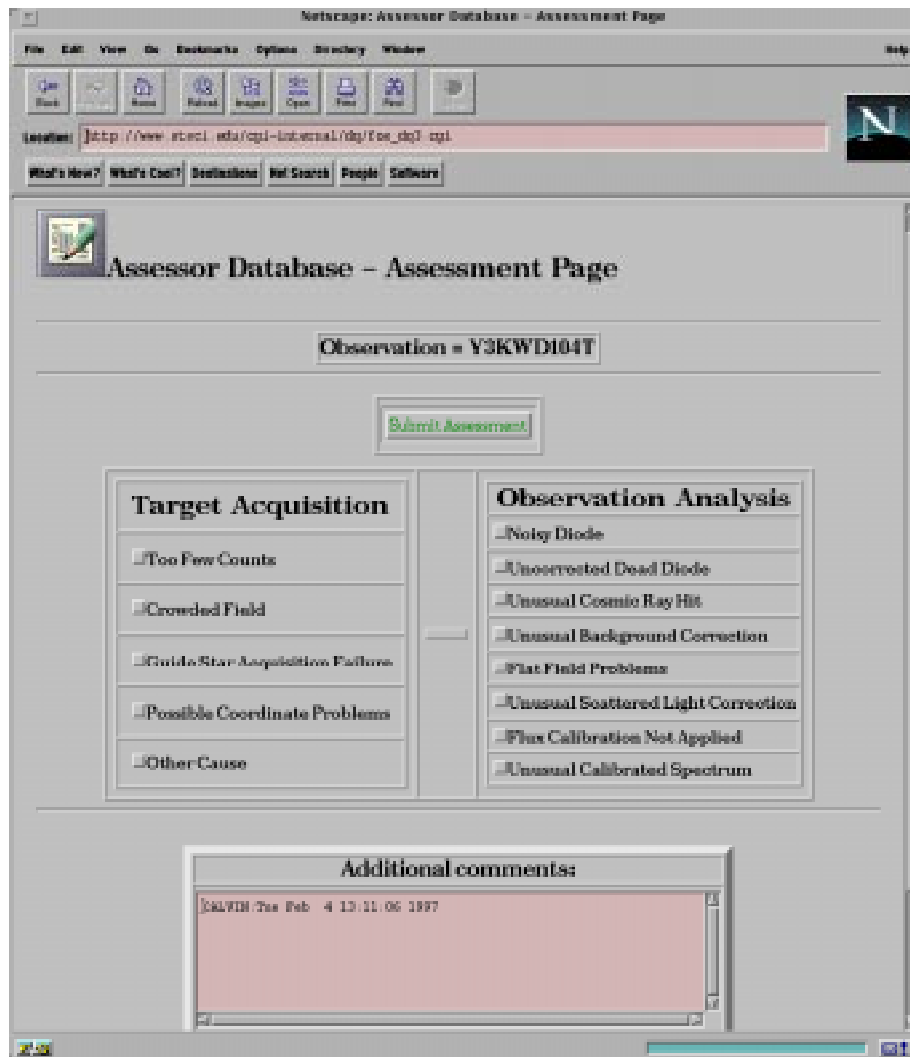
**WFPC2**:

| Bias Ripples | Horizontal Smearing | Residual Images | Contamination Worms | Stray Light |
|---|---|---|---|---|
| Dragon Breath | Ghosts | OTA Reflections | Moving Targets | Missing Data |
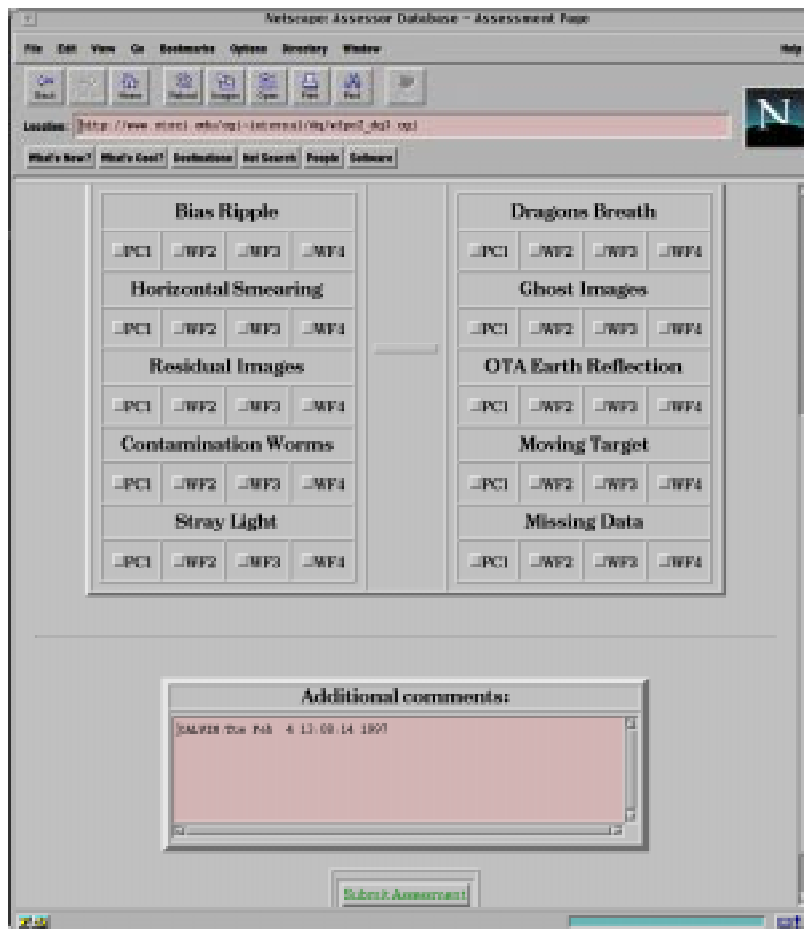


In addition to these, the tool allows for a virtually unlimited space for text comments to be recorded with these selections. This text field can be appended to at any time, without changing the assessed observations.

Once the "*Submit Assessment*" button is clicked, the assessment is written to the database and viewed on the page final Results page, which is as follows:

At this point, the user would want to assess the remaining observations, and would click on the Netscape "**BACK**" button twice, arriving at the "**Matching Observations** Page". The user would need to use the Netscape "**RELOAD**" to update the form posted. Following the suggested scenario, the user would select the "*Assess Remainder*" option and submit the remainder to be assessed in the default mode of "OK". The tool brings the user to the "**Assessor Database - Results**" page and assessment recording for that set of observations is concluded.

## 4. Procedures

*A] It is anticipated that most contact scientists will use the tool in the following way, with the paper products from one visit of a program at hand.*

1.  Select the proposal number and visit number corresponding to the paper product set.

2.  Having looked at your paper products, one or more datasets may have a certain anomaly. Select those datasets from the list, select the "*Assess selected datasets*" box and hit "*Continue"*. Note that all the datasets selected must have the same anomaly (or anomalies). This will bring up the "**Assessment"** page.

3.  Select the anomalies that these datasets share. Put in any comments that are appropriate. Hit the "*Submit Assessment*" button. This will bring up the "**View Assessment**" page.

4.  There the results of these assessments are given. Hit the "**BACK**" button on the browser twice to get back to the "**Matching Observations**" page AND HIT "**RELOAD**" so that the updated form is posted.

5.  Repeat the select/assess/comment process until all the anomalous datasets are assessed.

6.  After the anomalous datasets are assessed, the rest of the datasets can be done by selecting the "*Assess remainder*" box. This will bring up the assessment page with the default (no anomalies) settings. In most cases there is no anomaly to record, but sometimes a comment might be needed.

7. The last option of the list allows the user to add a text comment to ALL the datasets without changing the existing assessments.

8. On completion and having "**RELOAD**ed" after making the last assessment, it is possible to select "*View Existing Assessments*" to go over it one last time. Otherwise, hit "**BACK**" to get to the **Observation Selection Form** and choose another proposal/visit combination to assess.

9. For those who don't choose to do their data quality assessment in the way outlined above, the functionality provided by the tool should give the means to do the assessment in a variety of ways. If there are any extra functions that would make life a whole lot easier that we haven't anticipated, please contact Robert Jedrzejewski (rij@stsci.edu).

### B] What if a mistake is made?

It is always possible to change an assessment by selecting the dataset(s) to change and selecting "*Assess selected datasets*". Hit the "*Continue*" button to bring up a page that displays the existing assessment(s), and either change the assessment, or else just append another comment without changing the assessment.

Note that the assessments stored in the database have one set of values only, which is whatever the last assessor made them. On the other hand, comments always APPEND, so there is always a record of someone making an assessment.

## 5. Additional Features for the Future

1. We will be adding the capability to generate a ASCII report directly from the resulting Web pages to file with paper products.

2. To expand the selection criteria capability.

3. All of the information residing in the instrument results database, containing the assessments will be propagated back into DADS. STARVIEW would then have access to all of the DQ information. Work is currently in progress to define how this will be done.