BRINGING IMAGE DATA QUALITY TO THE VIRTUAL OBSERVATORY: AN ASTRO-WISE EXAMPLE

Gijs Verdoes Kleijn, Andrey Belikov, John McFarland, and Edwin Valentijn

OmegaCEN / Kapteyn Astronomical Institute, University of Groningen, The Netherlands

ABSTRACT

The Virtual Observatory provides access to a rapidly growing volume of astronomical data. For the uptake of the VO infrastructure by astronomers in their research it is now increasingly important that they can assess in detail the quality of published data. This paper describes the approach to quality control in the Astro-WISE information system for wide-field imaging surveys and how the data quality information together with the data can be accessed via the VO.

Key words: Virtual Observatory, Astro-WISE, wide-field imaging.

1. INTRODUCTION

The Virtual Observatory (VO) provides access to many types of astronomical data, such as images, spectra and catalogs. Access to detailed quality information for that data can be of paramount importance to obtain and trust scientific results newly derived from the VO published data. The required data quality criteria can depend among other things on science goal, type of data, wavelength regime and reduction steps.

At the starting line there is the quality of the raw observations which depends on instrument characteristics and status and atmospheric quality at time of observations. At the finish line is the quality of the final product after data calibration and reduction such as image distorsions and illumination variations across the field of view for a mosaiced image. The desired level of detail in the description and quantification of the data quality can vary widely. The astronomer might need to know in detail every step in the creation and calibration of the final data product to decide if the data meets the quality standards set by the science goal. Alternatively, it is sufficient to have a high level overview of the data quality and general limitations of the data.

High-level data quality information can be made accessible already via the VO. The frame work for detailed data descriptions is provided by VO standards for data models. The Data Model for Astronomical DataSet Characterisation [1] standardizes high level metadata necessary to describe the physical parameter space of observed or simulated astronomical data sets, such as 2D-images. It has the the status of recommendation. Furthermore, a VOTable provides the ability to put links to on-line journals describing the data or to other VO or web services. A VO application such as VOExplorer can provide in this way links to on-line journals.

Access to comprehensive detailed quantified descriptions of astronomical data quality is a major outstanding challenge in the VO. For the uptake of the VO infrastructure by astronomers it is now very important to have easy access to such detailed data quality information, whether it is standardized or not (e.g., Gilmore 2007).

This article focuses on the use case of data quality of optical and near-infrared astronomical imaging surveys which is the expertise of the authors. The approach to data quality control advocated is to provide direct access via a database to extremely detailed metadata in addition to the pixel data as generated by the full chain of image processing (from raw image to final coadded image and catalog). Standard quality control tools and pipelines are provided which extract relevant information from the database. More importantly, the researcher has direct access her/himself to the full data lineage to built a customized quality control pipeline if desired. This paper describes this method and how the resulting detailed information on quality control can be accessed from the VO.

2. QUALITY CONTROL IN ASTRO-WISE

Astro-WISE[2] stands for Astronomical Wide-field Imaging System for Europe. Astro-WISE is a system consisting of hardware and software which is federated over about a dozen astronomical institutes and datacenters in the Netherlands, France, Germany and Italy (Valentijn

[1] http://www.ivoa.net/Documents/latest/CharacterisationDM.html
[2] http://www.astro-wise.org

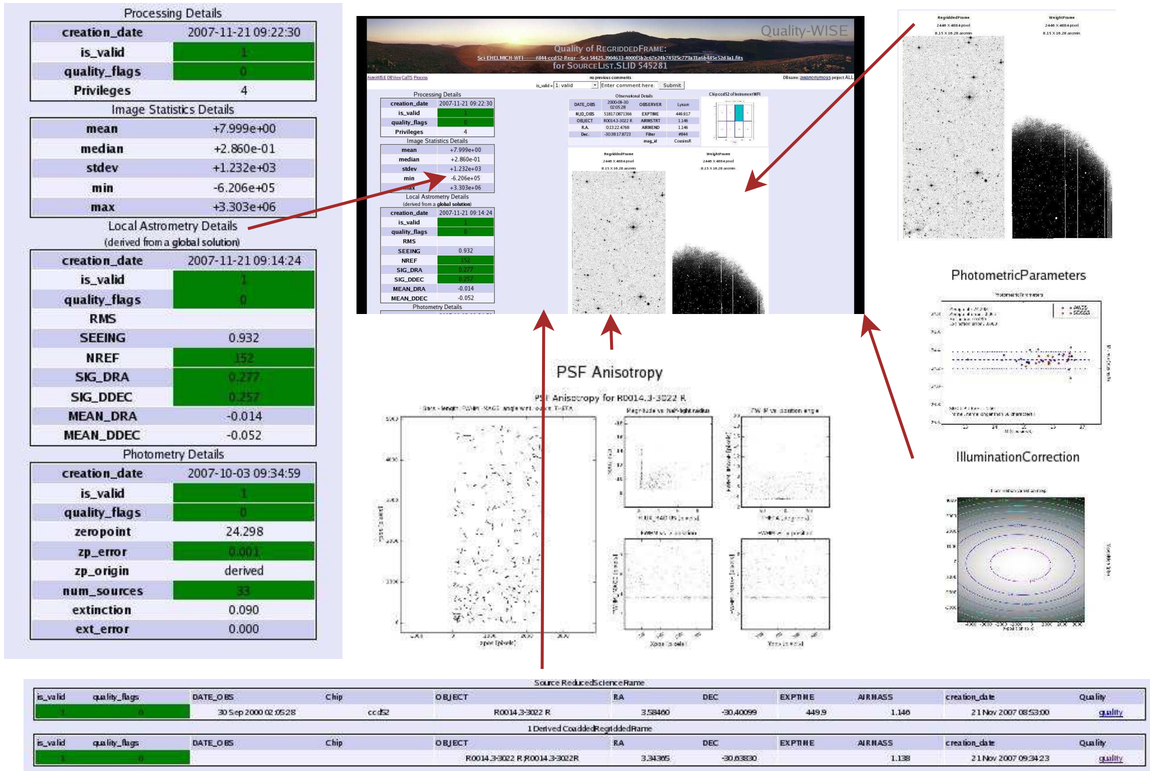

Figure 2. The QualityWISE webservice page is partially shown in the middle of the top row. The arrows link to zoom-ins. See text for a detailed description of the displayed information.

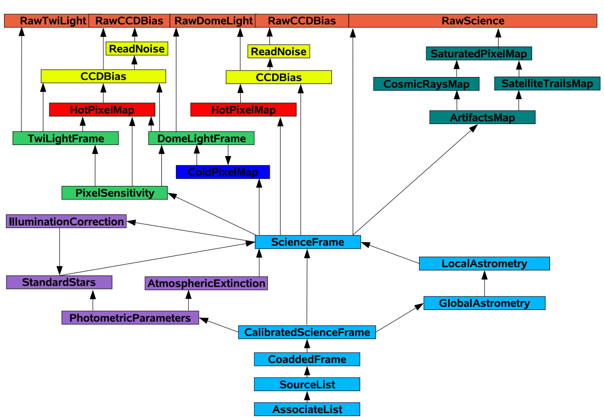

Figure 1. A slightly simplified view of the dependencies of Astro-WISE processing targets to the ocean of raw observational data of astronomical wide field imaging. Arrows indicate the dependencies leading back to the raw data.

et al. 2007; Valentijn & Verdoes Kleijn 2006). It has been developed to scientifically exploit the ever increasing avalanche of data produced by large astronomical imaging surveys. The system now has a capacity of 364 Terabyte storage, access to compute clusters with over 200 nodes and a large database federated over all four countries. The system handles data from more than a dozen different imaging instruments, some outside the optical/near-infrared domain. The Coma Legacy Survey (Carter et al. 2008), various WFI surveys and upcoming ESO VST surveys such as KIDS use Astro-WISE.

Quality control is at the heart of the Astro-WISE architectural design because it is one of the most critical and time-consuming steps in survey processing. The following architectural characteristics are crucial for how quality control is performed in Astro-WISE.

Object Oriented Paradigm The Astro-WISE architecture is based on an object-oriented paradigm (using the programming language Python). This means that raw, intermediate and final science and calibration data, processing parameters, pipeline code version, astrometric and photometric calibrations, source catalogs, galaxy surface brightness models, etcetera, etcetera, are stored in objects. Each class of objects has its methods (functions acting upon the object) and attributes (values associated with the object).

Integrated system with full data lineage Astro-WISE is an all-in-one system: it stores all relevant information of all objects in a single system. Objects have logical links to dependencies, i.e., other objects on which they depend (see Figure 1) giving access to the full data lineage.

To create such an all-in-one fully linked information system which is robust requires combining features of object-oriented and relational data storage: the metadata (i.e., attributes and links) are stored in relational database (Oracle) while at the same time each data object can be accessed as an object. Metadata in this case is defined as all non-pixel data and hence many more than usually ends up in a conventional FITS header. Pixel data is stored on massive dataservers. In conclusion: every bit of information can be traced in the system and is preserved in the system.

Each object for which quality control is relevant (i.e., 'Process Targets' such as bias frame, de-biased flat-fielded science image, photometric and astrometric solution, coadded image) has quality control methods. All quality control in Astro-WISE is built using three basic quality control methods:
`1. Verify:` In verification the quality of the object is assessed using the properties of the object as input. For example is the noise level of a bias too large? Is the signal-to-noise of a masterflat high enough and is the signal in the linear regime? Is the RMS of an astrometric solution too large? Are there too few or too many reference stars for a robust astrometric solution?
`2. Compare:` In comparison the quality of the object is assessed by comparing the properties of the object to previously derived versions of the object. For example, are patterns in a defringing frame significantly different from an earlier version?

Both verify and compare do not require human intervention. When verification or comparison is performed its verdict on data quality is stored in the object attribute `quality_flag`. This is a bitwise quality flag which is non-zero when a quality criterion is below specifications.
`3. Inspect:` Inspection allows quality control via human intervention by visually inspecting pixel data, diagnostic diagrams (e.g., positional offsets of astrometric reference objects, magnitude offsets of photometric reference objects), and values of the attributes of the object. If the inspector concludes the object is not in order, the inspector invalidates the object by setting its attribute `is_valid` to 0.

From these three building blocks implemented per class of objects one can build (customized) quality control within pipeline processing. These can take the form of calls to `verify`, `compare` and `inspect` methods in pipeline recipes in python. The diagnostic diagrams, images and verdicts of quality control can be bundled also and visualized via a webpage. Astro-WISE has such a webservice called *QualityWISE.*

For each science frame, from raw via intermediate products to final coads, the QualityWISE service creates a webpage at an unique url dynamically. Figure 2 shows an example QualityWISE page which can be accessed at this url[3].

On top of the page the filename of the pixeldata of the object that represents the science frame is a hyperlink to the fits data. This hyperlink is generated from information stored in the object. The row below contains in the middle entry fields to (in)validate the object (the `is_valid` attribute discussed above), optionally with a comment. Information entered here is stored in the Astro-WISE database. The boxes in the column on the left contain (from top to bottom): general information on processing, information on image quality, astrometric quality and photometric quality. All this information is accessed via the object that represents the science frame which contains links to dependency objects such as the astrometric and photometric solution. To the right are thumbnail images of the science frame and its weight frame which can be clicked to obtain enlarged previews. Thumbnails and previews are created by the Astro-WISE image server on the fly or retrieved from its cache. Directly above the two thumbnail images observational information is listed together with a graphical display of the location of the frame within the mosaic. Below the thumbnail images are the diagnostic plots for the astrometric solution, the photometric solution and the illumination correction. These diagrams are obtained on the fly or from cache via the quality control `Inspect` method of these objects. At the bottom are diagnostic plots on the PSF anisotropy and properties of extracted sources. Urls containing larger versions of the diagnostic plots are reached by clicking the plots. At the bottom of the page are links to the QualityWISE pages of objects which precede and follow the the current science data in the pipeline processing. These links are constructed from information stored in the object representing the science data and queries to the database.

3. ASTRO-WISE VO SERVICES

The Virtual Observatory interfaces[4] are an integral part of Astro-Wise. Currently it provides a Simple Image Access (SIA), Cone Search and a local publishing registry for Astro-WISE.

Research groups using Astro-Wise share their imaging data in "projects". One project member can access and re-process data of other members. Research groups publish imaging data in the VO by setting a flag on the related object. They Astro-Wise local registry will spot this and automatically publish the imaging data to the VO. The VO user can now fetch the data via Cone Search or SIA. A VOTable is returned which contains a link to the QualityWISE webservice on data quality. Figure 3 illustrates step by step the access to QualityWISE. **Step 1:** a WFI R-band image from project "ALL" is selected in Aladin (Ochsenbein et al. 2000). The VO section of the Astro-Wise webpages

[3] http://quality.astro-wise.org/QualityWISE?object_str=RegriddedFrame&object_id=3F6DE4DB740E9563E0407D81C5064581&project=ALL
[4] http://www.astro-wise.org/portal/aw_vo.shtml

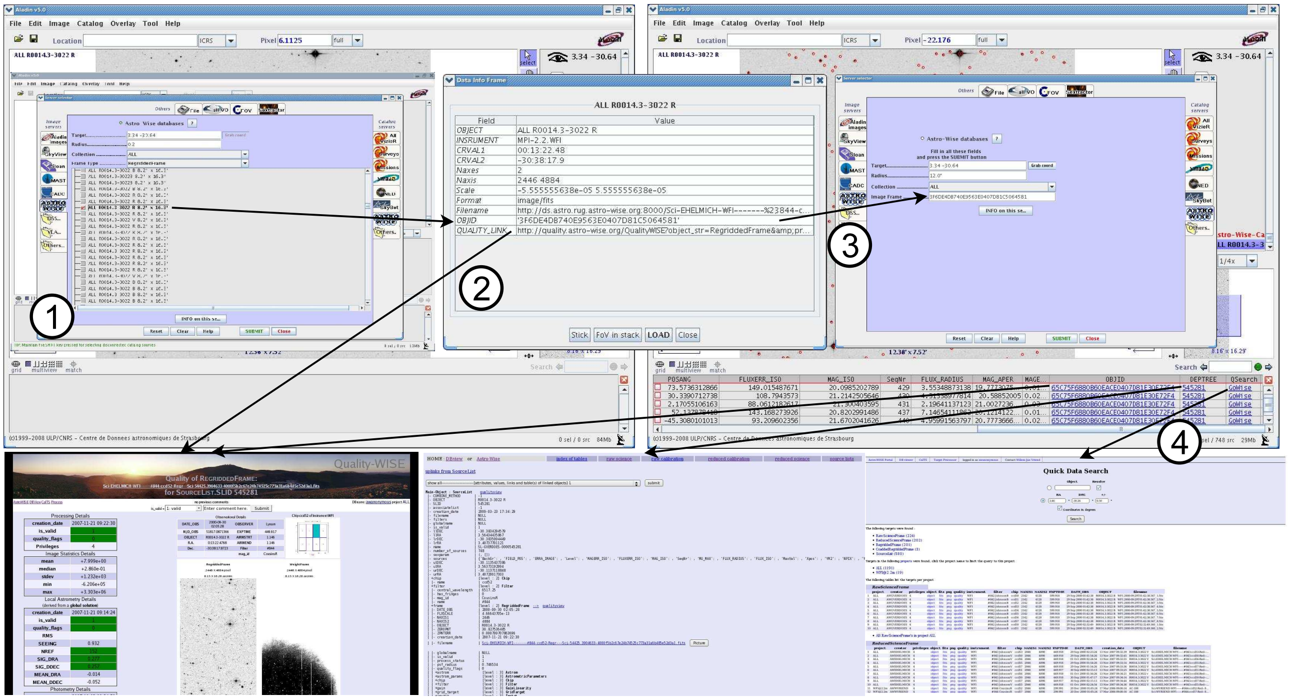

Figure 3. The use of the QualityWISE service of Astro-Wise from VO: the VO user accesses images or catalogs of Astro-Wise with links to Astro-Wise services, see Section 3 for the description.

explains how to update Aladin to make the Astro-Wise SIA and Cone Search available. **Step 2:** the VOTable which is returned contains the url to the QualityWISE server page (`QUALITY_LINK` field) for the image. **Step 3:** the Astro-Wise internal ID for the image (`OBJID`, OBJID='3F6DE4DB740E9563E0407D81C5064581' in step 2) can be entered to load the image in Aladin. **Step 4:** via the Astro-WISE Cone Search in Aladin the user retrieves source catalogs for the image. These VOTables contain links to (i) the QualityWISE server page for the associated image and (ii) the Astro-Wise DBviewer which shows the full lineage for the data object and (iii) QuickDataSearch server which provides links to data entities in Astro-Wise within a 0.5deg circle around the selected source.

4. DISCUSSION AND CONCLUSION

It has been shown that a combination of VO interfaces and native Astro-Wise web interfaces connected by links from a VOTable gives an easy and fast method to obtain quality information on imaging data published in the VO by Astro-Wise survey projects.

Generically at least three ways can be proposed to access quality information from the VO.
Via VOTable. The user stays within the VO and quality data is packed inside the VOTable. The advantage is that a single XML table contains the data and data quality description requiring no extra tools. The disadvantage is the limited capability of a VOTable to handle complex-structured data with a number of links and references.
Via external web services. This is the case realized in Astro-Wise. The complete data quality information provided by the data producer is easily available but is outside the VO.
Via a new XML schema. A data producer can create his own type of XML document to describe the data quality with all necessary parameters and lineage. The user will receive the data and the quality information but in a XML document with an (initially) unfamiliar schema.

Both the VOTable inclusion and the link to an external service are easy to realize for an existing VO service which makes them a natural choice for publishers to link data quality in the VO currently. Third parties can then decide of the offered data meets the quality standards required for their work.

REFERENCES

Carter, D., et al. 2008, ApJS, 176, 424

Gilmore, G. F. 2007, Highlights of Astronomy, 14, 587

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23

Valentijn, E. A., et al. 2007, Astronomical Data Analysis Software and Systems XVI, 376, 491

Valentijn, E. A., Verdoes Kleijn, G. A. 2006, ERCIM News 65, 20, see http://www.astro-wise.org/Public/published_versionEN65_astrowise_only.pdf