SDP Memo 084: Observatory Support Tools: Use Cases

Document number……SKA-TEL-SDP-0000181
Document Type……..……MEMO
Revision……..……..01
Authors……Ferdl Graser, Chris Laker, Rosie Bolton, Lorita Christelis, Antonio Chrysostomou, Rob Simmonds, Robert Laing, Peter Wortmann, Verity Allan
Release Date……..….2018-10-31
Document Classification…….…… Unrestricted

Lead Author	Designation	Affiliation
Ferdl Graser	SDP Systems Engineer	SKA SA
Signature & Date:		

SDP Memo Disclaimer

The SDP memos are designed to allow the quick recording of investigations and research done by members of the SDP. They are also designed to raise questions about parts of the SDP design or SDP process. The contents of a memo may be the opinion of the author, not the whole of the SDP.

Introduction

This is an extract from the SDP Confluence of some use cases, workshopped between SKAO and SDP staff.

Observatory Support Tools Use Cases

Note: There is a list of actors on this SKA confluence page: https://confluence.skatelescope.org/display/RF/Context+diagrams. The context diagrams on this page also appear in the L1 System Requirements Rev11. The list of actors is a useful reference for defining the correct actor in these use cases.

Use Case 1

Title: SKA scientist wants to check the QA logs to search for bad scans in an observing project:

Brief Description: As part of routine follow up of observing, the 'friend of project' notices bad QA flag(s) in the QA report and wants to more closely inspect the QA log generated by the relevant SDP pipeline. They will do this to evaluate the severity of the problem, perhaps where in the processing this problem became apparent, and ascertain any follow up action.

Proposed solution: The "FoP?" will request the full Science Data model for that observation and extract the QA report from that.

Functionality/requirements already supported by SDP: Science Data Model is available as an Observatory Science Data Product.

Missing requirements/functionality: Not yet analysed.

Comments or outstanding issues: RWJS: I don't know what a "friend of project" is. Is there a definition of that somewhere? I assume this is someone who has access to the Project information.
ACC: See link above for list of actors (item A8 in the context diagram). Admittedly, however, this term has not entered the project glossary/dictionary but it does appear in the OCD. The FoP is a staff astronomer who has been assigned to a project following the allocation

process and follows it through from obs design through to data analysis support. They will liaise with the PI/CoIs to help with anything to do with their data QA and project in general.

Use Case 2

Title: SKA scientist wants to inspect raw/calibrated visibilities:

Brief Description: As part of routine follow up of observing, or following a query on the data quality, the 'friend of project' has ascertained from inspecting the QA logs generated by the relevant SDP pipeline, that a close inspection of the data is required. From these logs, they notice that a particular scan exhibits higher noise than other scans. They proceed to inspect the associated visibilities in that scan in an effort to understand whether the problem lies with a particular baseline(s), receptor or group of receptors. This information is used to decide on the appropriate follow-up action.

The relevant workflow will likely start from inspecting the image cube for the observation, then the QA logs (choosing between user-defined filters for the numerical and/or graphical representations of the information), down to the visual inspection of scans and associated visibilities.

Proposed solution: Not yet analysed.

Functionality/requirements already supported by SDP: Not yet analysed.

Missing requirements/functionality: Not yet analysed.

Comments or outstanding issues: (Comments from Rosie) Once processed in the SDPs off-line system there is no intention to retain the raw visibilities. If the QA logs show that a particular scan was bad, this quantitative measurement could have been used to flag that scan before the final SDP products were completed. If they weren't flagged, so if bad data remained in the images, I think the correct thing to do would implement such flagging automatically in future, but I don't see that this requires retention of the visibilities just in case something went wrong. If the bad data has corrupted the final images enough to cause that run to fail, that block would have to be repeated. Since a "friend of project" could only ever be playing with statistics to determine whether or not some data is bad I think that any scenario could be covered by tuning auto-flagging algorithms that can be developed and refined as the telescopes are commissioned and as we continue to learn about it.

More comments from Rosie: It seems that functionality to re-process the visibilities really is needed, even if not used often. Therefore, we need the ability to retain raw visibilities in the cold buffer after they have already been sent to the hot buffer for off-line processing. This might mean that we need some additional overhead in the cold buffer.
RWJS: I think this could only be done at the QA stage, with information that is passed to TM, due to the visibilities not being retained normally. However, of visabilities have been retained (see Use Case 6) then it could be inspected by a project scientist.

Use Case 3

Title: Metadata do not report the correct status of the operative elements.

Brief Description: The metadata in measurement sets do not correctly reflect active components at each receptor. The SKA Scientist needs to be able to interrogate and repair the metadata. This may result in triggering a re-processing of the data (if they still exist) or re-observing if the data have been rendered useless (unlikely, but you never know!).

Proposed solution: Download the Science Data Model and use this to update metadata.

Functionality/requirements already supported by SDP: Not yet analysed.

Missing requirements/functionality: Not yet analysed.

Comments or outstanding issues:

Use Case 4

Title: Manual triggering of data processing.

Brief Description: Problems are noticed (either manually or via QA flags) which render the processing useless. The tool needs to allow you to fix the dataset (e.g. change reference receptor, or do strategic flagging) and then 'schedule' reprocessing.

Proposed solution: Depending on how the workflow is implemented, there are a couple of ways this could be achieved:

1. Schedule two new processing blocks, the first "fixing" the dataset while the second repeats processing. This solution would not require customising pipelines, but might double storage requirements and be fairly slow if the fix is small.
2. Introduce the fix into the workflow by customisation. This would require the workflow template to provide suitable modification points.

Functionality/requirements already supported by SDP: The extent and mechanism of workflow customisation has not been worked out yet, but the architecture allows for this.

Missing requirements/functionality: Not yet analysed.

Comments or outstanding issues:RWJS: Need to check this is held in the Science Data Model.

Use Case 5

Title: Additional RFI masking required

Brief Description: Full processing of a scheduling block instance (of e.g. 1-3 hrs) reveals low-level RFI that was not masked. Change the RFI mask to include these channels and then request re-processing. Aside: if these RFI signals are found to be consistent, then they should be added to the correlator mask.

Proposed solution: Not yet analysed.

Functionality/requirements already supported by SDP: Not yet analysed.

Missing requirements/functionality: Not yet analysed.

Comments or outstanding issues: Comments from Rosie: Does the SDP support reprocessing? I don't think so - so my comment here is exactly the same as for "Scientist wants to inspect raw visibilities": "Once processed in the SDPs off-line system there is no intention to retain the raw visibilities. If the QA logs show that a particular scan was bad, this quantitative measurement could have been used to flag that scan before the final SDP products were completed. If they weren't flagged, so if bad data remained in the images, I think the correct thing to do would implement such flagging automatically in future, but I don't see that this requires retention of the visibilities just in case something went wrong. If the bad data has corrupted the final images enough to cause that run to fail, that block would have to be repeated. Since a "friend of project" could only ever be playing with statistics to determine whether or not some data is bad I think that any scenario could be covered by tuning auto-flagging algorithms that can be developed and refined as the telescopes are commissioned and as we continue to learn about it.

Use Case 6

Title: Ability to write and execute scripts on raw data

Brief Description: SKA scientists should have the freedom and flexibility to write scripts that analyse the raw data to allow different monitoring or preliminary science analysis to be performed. For example, compute the phase structure function, running CASA-style scripts (or equivalent for SDP).

Proposed solution:The scientist will have to request for a raw data product to be created during the processing pipeline. This can then be requested through the delivery system.

Functionality/requirements already supported by SDP: Should have way to insert request for a product to be made at different points within the workflow / pipeline.

Missing requirements/functionality: Need to check this can be done currently.

Comments or outstanding issues:

Use Case 7

Title: Respond to changes in ionospheric behaviour:

Brief Description: SKA scientist has executed a script to determine the quality of the ionosphere during an observation and determines that the data should be processed with a different pipeline recipe to that originally requested. This is changed in the processing block (TBC) and batched processed with the new recipe as normal.

Proposed solution: Determining the quality of ionospheric calibration can be done either in real-time via QA (which could be queried either via TM or directly by the observatory, TBD), after the fact by analysing data products such as the preserved calibration solutions, QA data or the images themselves (via the Delivery observatory interface)

Influence the processing requires re-scheduling and possibly re-formulating of the science data model. Therefore this should be registered with SDP as an entirely new processing block. The old processing block should be cancelled as discussed on SKAO Confluence. All of this should be done by/via TM."

Functionality/requirements already supported by SDP: Providing real-time QA metrics to TM via Tango (that TM stores in the Engineering Data Archive). Cancelling and submitting a new Processing Block.

Missing requirements/functionality: Access to the QA metrics which are stored in the Engineering Data Archive in TM (by the SKA Scientist) is not supported by SDP, but may be supported by TM.

Access to the QA metrics in real-time (via Tango) by the SKA Scientist is supported by SDP, but the functionality to receive (at the observatory) and execute a script on the data is not supported.

Analysis of data products such as the preserved calibration solutions, QA data or the images themselves by the SKA Scientist after obtaining the data via the Delivery observatory interface.

Comments or outstanding issues:

Use Case 8

Title: Analyse effects of strong off-axis source

Brief Description: The processing of a scheduling block instance reveals that an off-axis source has not been properly peeled during the data reduction. It should be possible for an SKA scientist to analyse the cause for this. For this example, it is discovered that the issue is connected to the quality of the input sky model. It should then be possible for the scientist to provide their own sky model for the source and request re-processing (if the raw data are still available).

(Perhaps as a nice-to-have, it would be good to have a tool that can diagnose the impact of nearby sources if observations are carried out around a target position. Such a tool can then inform the observing strategy, as well as help diagnose issue post-observing).

Proposed solution: Cause analysis should be supported by logs and QA data, which should be available to the scientist. In case the available data does not suffice, re-processing with a workflow configuration offering more verbose QA information could be the next logical step. Theoretically the scientist could request the raw visibility and calibration data to work out the problem manually, but in practice this will often be unrealistic due to the amount of data involved.

If the root cause turns out to be the sky model, the scientist should be able to create a new version of the sky model with the modifications applied. They should then be able to request re-processing for the new sky model version.

Functionality/requirements already supported by SDP: Analysis would have to go via QA, as usual. Sky model modifications and processing block rescheduling should be possible using the appropriate interfaces.

Missing requirements/functionality: It is not clear whether the sky model versioning scheme is going to be flexible enough here. Paul seems to worry that promising too much here might cause us technical difficulties.

Comments or outstanding issues:

GSM Operations Use Cases

Also see ticket TSK-1866 - GSM update functionality undefined RESOLVED

Also see ticket SDPCDR-344: Why would automatic updates of GSM be useful? Below is the comment from that OAR from Andre Offringa:

I'm a bit doubtful whether "Automatic updating" of GSM is useful to plan for. I think it would be more useful to plan for these three stages of the GSM:

i) Bright sources only, from existing catalogues – only a few tens of them. Typically manually selected.

ii) Updates to these bright sources, possibly with results from a shallow survey, matched with existing catalogues. This is unavoidably manual labour.

iii) Result from a deep survey. Automated extraction when needed. This also can't be done automatically, as it requires careful analysis of things like beam errors and flux scales.

I think this is more realistic and matching with the building of sky models for e.g. MWA, LOFAR and Apertif.

There are also intermediate model products, which of course should be automatically derived, but can't be automatically propagated back into a GSM as that would give too much uncertainty to the flux scales etc.

A fully automatic-updating database-like GSM has been planned for with almost every telescope (including LOFAR) but has never been really used. Hence, I think that planning for it now is unrealistic as well. I think a bit closer look is required here how astronomers really want to have their data calibrated.

GSM Use Case 1	Query FoV / footprint
Actors	Science operations users
Description	Science operations users want to query a certain FoV or footprint of interest from the GSM. Users would need read-only access to the GSM to query the FoV / footprint. Users would select the format for download (see proposed solutions). One of the uses would be to transform the query results with a

	response. This may either be the SKA Mid or Low beam responses or a user provided synthetic response.
Proposed solution	SDP provides read-only access via VO protocols, either TAP for accessing tables, or a VOView with SAMP support, so that TOPCAT, Aladin, or DS9 can be used, without having to explicitly support them. SDP provides read-only access via TBD database protocol/service. Existing or 3rd party tools can be used to perform the rest of the required functionality: ● Download results in TBD format ● Transform query results with reponse
Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-790 Read access to GSM
Comments / Issues	

GSM Use Case 2	All sky visualisation
Actors	Science Operations Users
Description	Users want to visualise the whole sky as presented in the GSM using 3rd party tools. The users want to visualise the following: 1. All sky through a filtered visualisations. Filter options shall inlcude brigthness, kind, provenance and validity/applicability range. 2. Visualise the transformed data (convolved by the beam response). Beam responses may be a synthetic or user specific response or the SKA Low and Mid Beam responses. 3. After the visualisation is performed, the Users may want to download the visualisation itself and/or the GSM data in the visualisation.
Proposed solution	SDP provides read-only access via VO protocols: either TAP for accessing tables, or a VOView with SAMP support, so that TOPCAT, Aladin, or DS9 can be used, without having to explicitly support them. SDP provides read-only access via TBD database protocol/service. Visualisation is performed using existing or 3rd party tools.

Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-790 Read access to GSM
Comments / Issues	

GSM Use Case 3	Area Visualisation
Actors	Science Operations Users Observatory Staff
Description	Users want to visualise an area of the sky as presented in the GSM using existing or 3rd party tools. The users want to visualise the following: 1. An area of the sky viewed as a sky map 2. An area of the sky as a tabular list.
Proposed solution	SDP provides read-only access via VO protocols: either TAP for accessing tables, or a VOView with SAMP support, so that TOPCAT, Aladin, or DS9 can be used, without having to explicitly support them. SDP provides read-only access via TBD database protocol/service. Visualisation is performed using existing or 3rd party tools.
Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-790 Read access to GSM

GSM Use Case 4	LFAA Retrieve all data
Actors	LFAA MCCS system
Description	The LFAA MCCS system retrieves a copy of all the GSM data on a periodic basis or as triggered by TM for use in station calibration.
Proposed solution	SDP provides read-only access via TBD database protocol/service.
Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-790 Read access to GSM

GSM Use Case 5	Operator updates GSM
Actors	Science operations users
Description	The GSM is updated by a Science operations user following a well defined process (process is TBD). The following update functionality is required: ● modify a individual component. It might be a single position and multiple frequencies or just a single frequency (or new frequency) ● modify multiple (all, one or a filtered selection) entries with an expression (i.e. recalculation of flows) ● revert multiple (all, one or a filtered selection) entries to a given validity date.
Proposed solution	SDP provides read/write access via TBD database protocol/service TBD tools will be used to perform update functionality.
Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-720 Modification to GSM SDP_REQ-850 Changes to GSM

GSM Use Case 6	GSM update from LSM data post processing
GSM Use Case 6	GSM update from LSM data post processing
Actors	Science Operations Users
Description	When SDP processing finishes, the final LSM is available as part of the Science Data Product (TBC). Science operations users can update the GSM with this LSM data following a well defined process (process is TBD). Functionality required is TBD.
Proposed solution	SDP provides read/write access via TBD database protocol/service. SDP provides access to LSM via TBD mechanism. TBD tools will be used to perform update functionality.

Functionality/ requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-720 Modification to GSM SDP_REQ-850 Changes to GSM

GSM Use Case 7	GSM updates from CD/EoR processing at Regional Centres
Actors	Science Operations Users
Description	Deep CD/EoR processing taking place at Regional Centre(s) is expected to produce very accurate sky models. These improved models should be used by SDP during the calibration and source subtraction that occurs before further visibility averaging and transfer to the Regional Centre(s). In general it will also be beneficial if these models are fed back into the GSM for standard SDP processing.
Proposed solution	Global Sky Model updates for EoR SDP shall be able to receive GSM updates from the KSP and update the GSM. (TBC - this is wording of a proposed requirement written by Daniel Mitchell. To be discussed. Updates through standard mechanisms by Science operations users as Uses Cases above, or another process?
Functionality/requirements already supported by SDP	SDP_REQ-866 Hierarchical Components SDP_REQ-720 Modification to GSM SDP_REQ-850 Changes to GSM SDP_REQ-860 (requirement on hold until further discussion, not part of the March L2 submission). Background: Requirement SDP_REQ-860 orinigated during the review (24 11 2017) of the requirement for SDP EoR handover (now SDP_REQ-869). It was decided that a requirement for GSM updates is also necessary.