A census of potential GBOG auxiliary data products for GAP/CU9

prepared by: E. Pancino, A. H. Andrei, G. Bourda, P. Bendjoya, G. Clementini, Y. Frémat, U. Heiter, S. Hodgkin, P. Jofre, A. Lobel, G. M. Seabroke, C. Soubiran, W. Thuillot, L. Wyrzykowski.

reference: GAIA-CD-TN-OABO-EP-009-1
issue: 1
revision: 1
date: 2013-02-01
status: Issued

Abstract

This technical note contains a census of all DPAC auxiliary data and observing programmes that are coordinated by the GBOG WG.

Document History

Issue	Revision	Date	Author	Comment
1	1	01-02-2013	EP	Issued
D	7	28-01-2013	EP	Final revision with autors comments
D	6	6-10-2012	EP	New material on various CUs and major revision
D	5	29-05-2012	PJ	Part of CU3 concerning SEP spectra
D	4	24-05-2012	MA	Part of CU3 concerning GBOT and the EPC
D	3	14-05-2012	GMS	Part on CU6 and document restructuring
D	2	13-05-2012	UH	Part of CU8 section drafted
D	1	02-05-2012	EP	Major revision
D	0	16-03-2012	EP	Creation

Acronym List

Acronym	**Description**
2MASS	2-micron All Sky Survey
A&A	Astronomy and Astrophyscs
AIP	Astrophysikalisches Institut Potsdam
AL	ALong scan (direction)
AP	Astrophysical Parameters
ARI	Astronomisches Rechen-Institut
ASCII	American Standard Code for Information Interchange
BP	Blue Photometer
BVID	Bordeaux VLBI Image Database
CCD	Charge-Coupled Device
CCF	Cross-Correlation Function
CDS	Centre de Données astronomiques de Strasbourg
CU	Coordination Unit (in DPAC)
DPAC	Data Processing and Analysis Consortium
DPACE	Data Processing and Analysis Consortium Executive
DU	Development Unit
ELS	Emission-Line Stars
EP	Ecliptic Pole
EPC	Ecliptic Poles Catalogue
ESA	European Space Agency
ESAC	European Space Astronomy Center
ESO	European Southern Observatory
ESOC	European Space Operations Centre
ESP	Extended Stellar Parametrizer
EVN	European VLBI Network

FLAMES	Fibre Large Array Multi Element Spectrograph
FoV	Field of View (also denoted as FOV)
GAP	Gaia Archive Preparation
GBOG	Ground-Based Observations for Gaia (DPAC)
GBOT	Ground-Based Optical Tracking
GCRF	Gaia Celestial Reference Frame
GIQC	Gaia Initial QSO Catalogue
GSA	Gaia Science Alerts
HARPS	High Accuracy Radial velocity Planet Searcher
HR	Hertzsprung-Russell (diagram)
HS	Hot Stars
IAU	Internation Astronomical Union
ICD	Interface Control Document
IDS	Intermediate Disperstion Spectrograph
ICRF	International Celestial Reference Frame
IGSL	Initial Gaia Source List
IMCCE	Institut de Mécanique Céleste et de Calcul des Éphémérides
INT	Isaac Newton Group of Telescopes
IR	Infra Red
IVS	International VLBI Service for geodesy and astrometry
LCOGT	Las Cumbres Observatory Global Telescope
LoI	Letter of Intent
LQAC	Large QSO Astrometric Catalog
LQRF	Large QSO Reference Frame
LSF	Line Spread Function
LSST	Large Synoptic Survey Telescope
MDB	Main DataBase
MNRAS	Monthly Notices of the Royal Astronomical Society
MOC	Missions Operations Centre
MOU	Memorandum of Understanding
MPC	Minor Planet Centre
NEO	Near-Earth Object
NEP	North Ecliptic Pole
QSO	Quasi-Stellar Object
PHA	Potentially Hazardous Asteroids
R	Spectral resolution (R=$\lambda/\delta\lambda$)
ROB	Royal Observatory of Belgium
RP	Red Photometer
RV	Radial Velocity
RVS	Radial Velocity Spectrometer
SED	Spectral Energy Distribution

SEP	South Ecliptic Pole
SME	Spectroscopy Made EAsy
SNR	Signal-to-noise ratio
SOC	Science Operations Centre
SPSS	SpectroPhotometric Standard Star
SSA	Space Situational Awareness
SVN	SubVersioN
SYRTE	Systèmes de Référence Temps-Espace
TAC	Time allocation committee
ToR	Terms of Reference
UVES	ESO Ultraviolet and Visual Echelle Spectrograph
VLBA	Very Long Baseline Array
VLBI	Very Long Baseline Interferometry
WFI	ESO Wide Filed Imager
WG	Working Group

Contents

1 Introduction

The GBOG WG (CS-004) was created with the main goal of coordinating ground-based observations for Gaia (CS-009) and to avoid duplication of effort. Some support to the observing teams was also given in terms of DPACE-level interactions with the TACs.

The GBOG WG submitted a LoI to the GAP WG, to express the need for archiving and preserving all DPAC auxiliary data, and the letter was accepted. E. Pancino and G. Seabroke are presently representing GBOG in GAP. In particular, *all* auxiliary data used to calibrate Gaia data (and to train Gaia algorithms) should be preserved, so that all Gaia data releases will be accompanied – when appropriate and necessary – by a corresponding set of auxiliary data. The LoI proposal includes not only the GBOG coordinated data, but also auxiliary data that are outside the scope of GBOG (see below), and that will be detailed in a subsequent document.

Moreover, some groups involved in ground-based observations for Gaia feel the need of saving not only the information strictly necessary to make the pipelines run (which will in any case be saved in the MDB), but also *(i)* additional metadata (critically selected literature information), which could be useful for future scientific uses of GBOG data by the community; *(ii)* raw data that come from small telescopes which do not maintain public archives; and *(iii)* intermediate data products, which will allow for the possibility of (partial) reprocessing of GBOG data if necessary. Finally, GBOG data in many cases are also potentially useful to the scientific community at large and thus they need to be published in the most appropriate way.

Most of the observing programmes are still ongoing, and the status of GBOG data is still evolving. The following sections describe the expected data products at the time of writing, including the expected scientific potential, and will be updated regularly.

1.1 Other DPAC auxiliary data outside GBOG

The ground-based observations coordinated by GBOG are not the only auxiliary data for the calibration of Gaia and the training of its algorithms: several groups in various CUs are assembling auxiliary data that are not coordinated by GBOG. There are other data sets, in the form of thoretical spectral libraries or critical compilations of literature and archival data. Examples can be found in DK-015, concerning CU6, and we hope that more CUs will follow the example of CU6, by producing similar documents on their auxiliary data.

1.2 Document organization

Each of the following sections deals with a separate CU and each CU section lists the corresponding projects coordinated by GBOG. Each project is organized as follows.

1. **Use for Gaia.** Brief description of the project and of its use for Gaia. Any relevant ICD (Interface Control Document) already available is mentioned here, and a summary of data products that are expected to go into the MDB is also written here. If any requirements on the CU9 archive derive from the use for Gaia of the data set, they are mentioned here.

2. **Scientific potential.** Description of the scientific potential of the data set. The metadata that is expected to be associated to the data set, such as measurements deriving from the data set, critical compilations of literature data, cross-matches with existing catalogues, and so on, are described. Planned or expected scientific publications, on refereed journals or external archives and databases are mentioned here. Also, the expected usefulness of the data set to the scientific community can be briefly described here, including any requirements on the CU9 archive that may derive from the expected scientific exploitation of the data.

3. **Expected data products.** An itemized list of data products and of any auxiliary information that would be beneficial to publish in the CU9 database, including – when appropriate or relevant – raw data, intermediate or final data-products, measurements, literature information, and so on.

4. **Local or temporary storage.** A description of what data are available at the moment, of how much space it occupies, and an estimate of how much space it will occupy once the project is complete. A description of the facility, archive, database that is hosting the data at the moment, and of any additional data centre or website that is expected to host the data apart from CU9 and the Gaia archives.

2 CU3 data products

One auxiliary data project in CU3 which is vital for Gaia, and is not coordinated by GBOG is IGSL (RLS-001; ASP-001); a census of CU3 auxiliary data programmes similar to what has been done in CU6 (DK-015) would be extremely useful. The programmes presently coordinated by GBOG are listed in the following sections.

2.1 Ground based optical tracking of Gaia (GBOT)

2.1.1 Use for Gaia

Ground **B**ased **O**ptical **T**racking (GBOT) is an observing campaign of the Gaia spacecraft itself conducted by a small network of groundbased telescopes (1-2 m class). The aim is to obtain daily positions of Gaia with an accuracy/precision of 10-20 mas. These will be used by ESOC (MOC) to be included in their orbit determination. Given the ambitious aims of Gaia, its trajectory must be known to a high degree, in order not to suffer tradeoffs in the precision of its most precisely measured stars, the main problem being an adequate compensation of light aberration. Furthermore Gaia intends to measure precise parallaxes of near solar system bodies, e.g. NEO's. For these the baseline, and thus Gaia's position needs to be very well known. Therefore, while for the majority of stars, GBOT will not make much of a difference, it will aid exploitation of the full potential of Gaia. One also needs to consider that residual aberration is a systematic effect, so even with those stars (i.e., the vast majority), for which the improvement by adding GBOT is negligible on an individual scale, this will no longer be the case, when large samples of such objects are being studied. For a more detailed description of GBOT, see (MA-003).

2.1.2 Scientific potential

The scientific potential of the GBOT data as such will be very limited, since it will consist of short exposures and relatively small fields centred on wherever the Gaia spacecraft happens to be at the time of exposure. Depending on the field, there will be apart from Gaia itself, some 10-200 stars, in some cases even less. The probability that there will be something interesting in these frames is very small, but not zero, therefore there is the possibility that one might want to release the data to some public archive, e.g. VO, etc. Looking at the complete data set , the GBOT data will be an unique set of satellite orbital data, which in principle could be used for studies of celestial mechanics (Mignard, priv. comm.). Summarising, given the large volume and little scientific value of GBOT data to the general public, it is probably not wise to have it publicly available in the Gaia archive.

2.1.3 Expected data products

GBOT data will comprise the actual observations, and the necessary calibration data (i.e., flats, bias frames, etc.), as well as orbits and ephemerides, as fed back to GBOT by ESOC. Given a mission length of $\simeq$6 years, i.e. 2200 days and GBOT observations every night, except the three nights around full moon, we will obtain about 2000 nights worth of data. Assuming 4K$\times$4K data (most telescopes will have 2K$\times$2K CCDs, but to be on the safe side, we choose the larger size), 10 science exposures and 20 calibration exposures, this will result in $\simeq$ 2 TB of raw data. Reduced data will add another 4 TB. The actual results, i.e., the files that GBOT delivers and all other items will be negligible in size compared to the imaging data. In order to accommodate for all eventualities we will presume that the total storage space needed for one copy of GBOT data will be 10 TB. A new twist in the methods of GBOT is the newly established possibility of using VLBI to supplement the optical measurements. Since this is still in its very early stages, it is mentioned here, but at current not included in the assessment. In future editions of this document, this may well change.

2.1.4 Local or temporary storage

Within GBOT, it is currently foreseen to have a main database, from which operations are performed at SYRTE in Paris, and a full fledged mirror in Heidelberg. These systems should be as close to identical as possible, so that one can take over from the other, should the need arise. Moreover having the GBOT computing infrastructure in two disjunct geographic locations, ensures redundancy, thus minimising the risk of data getting permanently lost, due to catastrophic event (e.g., fire, flooding) at one site. Since the whole enterprise is somewhat remote from the rest of Gaia, the interaction of GBOT with the main Gaia data system will on a day-to-day basis be limited, and usually restricted to the material exchanged with ESOC (via SOC, see above). Another reason to have the data stored within the system of GBOT is that after $\simeq$2 years, once Gaia data is available to be used as a reference catalogue for GBOT's astrometry, all data collected until then needs to be reprocessed.

However, during the duration of the mission, it may be a good idea to also store the GBOT data somewhere outside of the GBOT system, to prevent data loss due to criminal activity (hackers), which could in principle infect both mirrors. In this case, a rough estimate of the amout of data is between 10 and 20 TB (at most) in total.

2.2 Ecliptic poles catalogue (EPC) and spectra

2.2.1 Use for Gaia

The Ecliptic poles regions will be repeatedly observed by Gaia during commissioning, after the satellite reaches its final position in L2. Thus the South and North Ecliptic pole regions were surveyed with wide field imagers in different photometric bands, i.e. $uBVRi$ in the south (using WFI and MOSAIC2) and $ubri(z)$ in the north (z in the NEP field comes from Hwang et al. (2007), which has also been used to calibrate our photometry in the other bands), to characterise all the potentially observable objects. The optical data goes far deeper than Gaia, this way one can ensure completeness at Gaia's cutoff level. Apart from positions and photometry, the EPC also features proper motions of a subset of stars. The data will not be used in the pipeline, except for the spectroscopic data (see paragraph below). The project and the first version of the photometric catalogue are described in MA-002.

A spectroscopic campaign is also ongoing in the Southern field, observing almost a 1000 stars with FLAMES-GIRAFFE and UVES. Bright stars around the NEP are observed as well with instruments covering the RVS spectral range (see Section 5.2). Data taking for the spectroscopy in both the SEP and the NEP field has been completed. The resulting spectra and radial velocities will be used by CU6 to commission both the RVS instrument and the CU6 pipeline.

2.2.2 Scientific potential

While Gaia will surpass the quality of the astrometric data in the EPC, there is some benefit to publishing the EPC with the Gaia data:

- the spectra will provide radial velocities and astrophysical parameters with higher precision than Gaia;

- the EPC goes far deeper than Gaia, and in several photometric bands;
- while the Gaia astrometry will be better under all circumstances than that of the EPC (which is not only affected by the atmosphere, but also hampered by the quality of today's reference catalogue material, which feature zonal errors, comparable with the field size of the EPC - thus limiting its total accuracy to 50 mas or more), repeating the astrometric reduction using Gaia itself as reference material can be of considerable scientific value, since this would allow the access to kinematics of stars fainter than Gaia's limit at an unprecedented precision. An interesting science case could be brown and white dwarfs, and other intrinsically faint stars, to which Gaia will only grant limited access. In this context it may also be of interest to consider obtaining an additional epoch of data, after Gaia's operational phase has ended.

The EPC will be published in a refereed Journal, and GBOG is collaborating to analyse the spectroscopic data as well, which will yield additional publications.

2.2.3 Expected data products

- Reduced images: fully reduced (photometric), projected, stacked, fully assembled.
- Photometric catalogue for both the North and South EP, with a typical uncertainty of 0.01 mag on the calibrated magnitudes in the well exposed range, and calibrated astrometrically on the PPMXL catalogue with a typical uncertainty of 10 mas (r.m.s.) and 100 mas absolute (see above). Stars brighter than the CCD data's limit are taken from the literature, as well as meridional circle measurements in the northern field.
- Reduced UVES spectra, with a resolution of $R \simeq$47000, a typical SNR of 30 and covering the spectral range of $\simeq$ 50 stars in the range 6600 – 10600 Å.
- Reduced GIRAFFE spectra of almost a thousand stars, typical SNR ranging from 5 to 100, with two grisms, LR2 in the range 3964 – 4567 Å, and a resolution of $R \simeq$6000, and HR21, in the range 8484 – 9001 Å, with a resolution of $R \simeq$16200 (note: not all stars have spectra in both grisms).
- Measured radial velocities with different internal methods, among them cross correlation with masks and templates, fitting to synthetic spectra, and using telluric lines to check for wavelength calibration errors. The results will be compared with those obtained using the CU6 methods to assure robustness.
- Fundamental atmospheric parameters obtained with two independent methods, one based on equivalent widths using DAOSPEC (Stetson & Pancino, 2008) and GALA

(Mucciarelli et al., in prep.) and the other one based on spectral synthesis using the code (SME, Valenti & Piskunov, 1996). In addition, individual abundances, such as the α-elements, will be determined.

- Referencing catalogues crossmatched between various versions of the EPC, as well as between the final EPC and other literature data, e.g. 2MASS
- Possibly compilation of literature data of all the objects contained in the EPC and spectroscopically observed objects (exact necessity and scope needs to be determined)

The full data set (mostly imaging data) will be in the order of 2 TB (somewhat depending on how many different processing stages will be stored).

2.2.4 Local or temporary storage

At the moment the data are stored locally at the ARI in Heidelberg, with redundancy. The released catalogues are either in DPAC's SVN repository, or the later versions, which do not fit into the SVN are stored in the GBOG space at ESA in Madrid/Villafranca. Upon completion (foreseen for the end of 2012), all material (imaging and spectroscopy) will be compiled and deposited in the GBOG repository as well. Given the large volume of the data during processing, it probably does not make sense to store everything (i.e., fits files) somewhere else than locally. Upon publication on international refereed Journals, relevant data will also enter the CDS data system. It might become useful to cross-correlate these with Gaia data at some stage: the whole set of catalogues will occupy no more than 5 GB approximately.

2.3 Photometric observations of QSO

2.3.1 Use for Gaia

The Gaia mission inherited from the Hipparcos mission in the aspect of being free from atmospheric, thermal, and gravity constraints and with full sky visibility, and in the aspect of superposition of widely separated fields to build sky great circles. But Gaia supersedes Hipparcos in the aspect of bearing the manifold results upon a quasi-inertial reference frame that provides a micro-arcsecond consistent materialization of the ICRS, formed by QSOs at unprecedented density, and enabling tens of micro-arcsecond absolute positions, motions, and parallaxes. The determination, definition, and reliability of the QSOs is therefore all important for the Gaia mission. Gaia will observe, presumably, some 500,000 QSOs, from which at least some 10,000 of the most astrometrically stable will form the core GCRF.

It is worth mentioning that two other quite robust extensions of the GCRF and anchored to it will also be produced, to brighter regimens. The one formed by the unresolved galaxies (some 10 million of objects) and the QSOs that did not make it to be in the GCRF (and those would presumably include several radio-loud QSO). And the one formed approximately by half a billion of stars with highly accurate position and proper motions.

QSOs are thus crucial targets and accordingly pipelines are capable of classifying them, by using a library of SED templates. At the same time, a previous QSO catalogue is being built by CU3-GWP-S-335-13000, to ensure an a priori sufficient number of objects to form the GCRF. This catalogue is termed the Gaia Initial QSO Catalogue (GIQC) and belongs in the MDB, as well as it is embedded in the IGSL.

2.3.2 Scientific potential

No ground-based observation comes close to matching Gaia's astrometric precision, and thus the positions in the GIQC are in this sense merely indicative - although the GIQC spun off to the LQRF (Andrei et al., 2009) and to the LQAC (Souchay et al., 2009, 2012). New versions of the latter will be useful before, as well as incorporating, Gaia data. This is because magnitudes, redshifts, morphology, and variability indices are important Gaia inputs, and will not necessarily be delivered in the mission catalogue. The first two parameters enable to adjust the SED template library to the actual Gaia observation output. The other two enable to warn about objects for which no ideally point-like and stable position can be derived. Because the ground-based observations aim to benchmark objects and radio loud quasars, long-term photometry is instructive per se. It remains to be seen if at Gaia sub-milli-arcsecond unprecedented optical precision and at the space sensitivity regimen most QSOs would indeed present either baricentric jitter or an isophotal signature of the host galaxy, or both. Even if only particular classes of QSOs e.g., the brightest, are such plagued, this might already harm the best definition of the GCRF. So, variability and morphology initial flags and along mission study and monitoring may pay off highly to clean up the error budget of important targets, by modeling their photocentre place and behavior. At the same time if the CU9 or other instances of the Gaia consortium regularly deliver the photometric data observed by the satellite, this will provide an important synergy between ground-based and on-board observations, paramount for astrophysical interpretation and to define what is a Gaia QSO in the sense of fulfilling the role of fiducial point in the GCRF.

2.3.3 Expected data products

The main programmes led by GWP-S-335-13000 researchers have as data products medium to high SNR blue and red photometry and multi-colour variability series. The targets are benchmark and radio-loud quasars. The particular output from each programme are listed below.

1. High SNR astrometry and red/blue photometry of 8 QSO crowded regions observed with CFHT/MEGACAM in 2008/09. The aim was to investigate the degradation of

astrometric angular distances on the basis of different morphology types (Taris et al., 2011).

2. High SNR relative astrometry and red/blue photometric variability monitoring of 20 QSOs, using the ESO WFI in 2008/10, with bi-monthly periodicity. The aim is to study the correlation between astrometric and photometric variability.

3. Medium SNR astrometry, red photometry, and morphology observations of the optical counterpart of the QSO planned to tie the GCRF to the ICRFn, using the ESO WFI in 2009.

4. Medium SNR photometry to establish the optical light curve of the QSO planned to tie the GCRF to the ICRFn, using the TAROT telescopes in Calern and in Chile, the Zadko telescope in Australia, and the Haute-Provence telescope in France. The programme started in 2011 and continues to date of this report, and provides frequent observation of the visible targets (Taris et al., 2013, A&A, in press).

5. High SNR red/blue morphology of selected quasars planned to tie the GCRF to the ICRFn, using the Nordic Optical Telescope. The programme will be conducted in 2012/13.

Besides these programmes, a much larger and far reaching amount of QSO data products will be gathered by Gaia itself, namely the LSF, multi-band magnitudes, and great-circle relative astrometry corresponding to each passage of each QSO. If the CU9 or other instance of the Gaia consortium regularly delivers the photometric data observed by the satellite this will provide selected targets for ground observations and will be of enormous value for the understanding of QSOs. Notice that the recovery of such particular daily satellite products as a post-mission effort, if possible at all, will demand a much larger effort than their storage on a post-observation basis. And monitoring, analyzing, and combining those outputs provide elements to remove or unweight individual quasars from the role of building the GCRF. Or, if the photocentre motion could be modelled, to compensate it and consequently relieving individual error budgets.

2.3.4 Local or temporary storage

The present storage of the observation outputs are currently stored at the institutes of the GWPS-335-13000 Researchers PI-ing and CoI-ing the observational programmes. The astrometric data appear in the IGSL, whereas the photometric data is planned to appear as a separate catalog in the MDB. The amount of data should be minimal. No provision is currently made to save raw data or data products other than locally.

2.4 Alignment to ICRF source list

2.4.1 Use for Gaia

The current IAU fundamental celestial reference frame is the ICRF2, composed of VLBI positions of extragalactic radio sources. In order to prepare the alignment between the radio frame and the future Gaia frame in the optical domain, an observing programme has been developed and initiated in 2007 to observe new VLBI sources. The sample consists of 447 optically-bright extragalactic radio sources, selected from a dense catalogue of weak radio sources. A multi-step VLBI observing strategy has been specifically developed to detect, image, and measure accurate VLBI positions for these sources (Bourda et al., 2010, 2011, and Bourda et al. 2013, in preparation), by using the EVN and the VLBA. This is an on-going programme and the final step just started in June 2012.

The radio sources observed during these dedicated VLBI experiments, as well as those from the ICRF2 catalogue, are already part of the GICQ. The important point for the MDB is that the sources identified as suitable for the alignment will be quoted specifically within the GICQ.

2.4.2 Scientific potential

The alignment between the VLBI and Gaia frames will be important not only for guaranteeing a proper transition if the fundamental reference frame is moved from the radio to the optical domain, but also for registering the radio and optical images of any celestial target with the highest accuracy. Such a registration will allow one, for example, to pinpoint the relative location of the optical and radio emission in active galactic nuclei (AGN) to a few tens of μas, placing constraints on the overall AGN geometry. Estimates of this optical-radio core shift indicate that it may amount to $\sim$100 μas on average at 8 GHz, significantly larger than Gaia and VLBI position accuracies. It should thus be directly measurable.

In addition to the observing programme presented in section 2.4.1, we also work at identifying those ICRF2 sources that are the most suitable for the alignment with the Gaia frame. Accordingly, plans are being devised, within the framework of the IVS, to observe all such "link sources" during the Gaia mission in order to control source position stability and accuracy, as well as potential variations in VLBI structures. To this end, the Gaia scanning law should allow us to carry out quasi-simultaneous VLBI and Gaia observations. This will be of high interest for astrophysical purposes (e.g. near real time optical-radio comparisons for studying core-shifts and constraining AGN jets properties).

We plan to go on publishing our results in A&A, linked to the CDS database. We also plan to deliver the results of our VLBI observations through the Bordeaux BVID (`http://www.obs.u-bordeaux1.fr/BVID`), developed locally for VLBI images. And finally, VLBI observations carried out through the IVS will be publicly available for future analyses.

2.4.3 Expected data products

In the framework of the VLBI observing programme detailed in section 2.4.1, data products for each extragalactic radio source observed are:

- VLBI position (right ascension and declination) and related accuracy;
- VLBI maps at S- and X-bands (and related accuracy parameters);
- VLBI total flux density at S- and X-bands;
- S/X spectral index;
- VLBI continuous structure indices at S- and X-bands.

2.4.4 Local or temporary storage

At the moment, the results of the detection step have been published in A&A (Bourda et al., 2010) and the related tables are available through the CDS portal. The results of the pilot imaging experiment have also been published in A&A (Bourda et al., 2011) and the related table is available through the CDS portal, while the VLBI images are stored in BVID (see `http://www.obs.u-bordeaux1.fr/BVID/GC030`). There are plans to make BVID VO-compliant in the future.

Future results will be published in the same way, i.e., through referred journals and related, local or international specific databases. There is actually no plan to publish the data in more ways than BVID, CDS, and refereed journals.

3 CU4 data products

3.1 GAIA-FUN-SSO

The main ground-based data products will come from the Follow-Up Network for Solar System Objects, Gaia-FUN-SSO. This network will provide auxiliary data on Solar System Objects, and in particular NEO objects, which can be of great interest for the scientific community at large.

3.1.1 Use for Gaia

Gaia alerts will be issued from the short-term processing related to Solar System Objects. These alerts will concern some critical objects that the probe will discover but will not be able to monitor for providing robust orbital elements. These alerts will be sent to DU459 and processed in order to provide ephemerides useful for ground based observations (CCD astrometric measurements) by a dedicated network, the Gaia-FUN-SSO network. The ground-based observations (successful observations issued from the 37 observing sites of Gaia-FUN-SSO) will not be directly sent back to the Gaia MDB but they will be sent to the IAU Minor Planet Center in order to complement the international database of asteroids and comets. They will be subsequently used by the Gaia data processing through the MPC database which allow periodical updates of the CU4 DU451 auxiliary database for the identification of objects detected during the next scans. Without such ground-based follow-up, many of the SSO, in particular Near-Earth Objects, would be lost after their detection by the probe.

3.1.2 Scientific potential

The scientific potential of the ground-based follow-up observations of SSO is mainly the improvement of the orbits of newly detected objects, their identification, the guarantee that they will not be lost and the possibility to improve subsequently our knowledge of their physical properties. We must also note that Gaia will have the specificity to detect objects at small Solar elongation. This is particularly important for the objects of our terrestrial space environment (the NEO and especially the PHA objects) and, for example, these data will be able to feed specific databases of the SSA program of ESA.

3.1.3 Expected data products

We foresee to get several data:

- the Gaia data selected for alerts to DU459 (ASCII positioning measurements) which will be directly sent to MPC;
- the Gaia-FUN-SSO astrometric measurements (ASCII positioning measurements too) from the network stations which will be also sent to MPC;
- CCD images from the stations which will agree to send them (we intend to have a data policy defined in a MOU with the observing sites).

3.1.4 Local or temporary storage

IMCCE-Paris Observatory is the central node of the Gaia-FUN-SSO network. All the ASCII data will be stored locally. The CCD images will be collected also locally in order to check the astrometry, or improving the homogeneity of the data, but we have not yet decided about their archiving. We could choose to have a time limited archiving. At the moment, no solution to save the data — other than locally — and/or with VO-compliant formats is foreseen.

Although it is very difficult to estimate the final amount of data, some attempt can be made. We assume we have N alerts by week; one alert would require almost 5 sessions, where 1 session amounts to, say, 20 CCD frames of approximately 5 MB each. If 33% of the 39 stations can observe, then we have $13\times5\times20\times5\times$N MB/week=$6500\times$N. If we have one alert per day, which can be quite realistic because some simulations gave us 1 alert every 4 days, we get less than 2 TB in the 5 years of Gaia operations. Of course, we probably do not need to keep all the images for a long time, although the disk space required is modest. The correcponding ascii data are $13\times5\times0.004\times$N, which amounts to less than 0.5 GB in the 5 years of Gaia operations.

3.2 Spectroscopy of asteroids

3.2.1 Use for Gaia

The task M-458-00900 Asteroid taxonomy is an end mission task dedicated to provide a mineralogical classification of asteroids. This will be the largest asteroid taxonomic catalogue ever. Moreover data in the blue part of the spectrum will be reachable for the first time thanks to the Gaia spectral facilities. The classification will be performed by non-supervised algorithms developed by our group. In order to test these algorithms on the huge data set provided by Gaia we need to create artificial data from typical master spectra in the same range of wavelengths.

We have conducted two campaigns of observation performed at TNGalileo in the Canary Islands, which offers the same spectral range as Gaia. We have been able to sample the previously defined taxonomic classes in order to get master spectra for each of the 25 identified classes. We have got 65 spectra from which we are now able to simulate Gaia spectra. It is then possible to get as many spectra as we want in each taxonomic class in order to test our unsupervised algorithms. The simulated data sets will be of comparable number of spectra to what we will have to deal with at the end of Gaia mission. This will allow to adapt and test our strategies of clustering.

3.2.2 Scientific potential

These data will be useful to feed the algorithm of taxonomic classification which has been adopted to treat Gaia asteroid observations. Gaia-based taxonomy will be based on spectroscopic data covering a wide range of visible wavelengths, including the blue part of the reflectance spectrum, which has been substantially lost in modern asteroid spectroscopy after the development of CCD detectors. The blue part is particularly important to discriminate among different sub-classes of the most primitive objects. This will also permit to better understand the distribution of different classes throughout the asteroid main belt, with important consequences for our understanding of the original gradient of composition of solid matter in the proto-planetary disk, and of the mechanisms of collisional and dynamical diffusion of asteroids in different size ranges. Of course, asteroid taxonomy is important to derive general information about the surface composition. In particular, Gaia will make it possible for the first time to have at disposal masses and bulk densities for about 100 asteroids. Adding taxonomic classification information for these objects will be very important to understand for the first time whether the internal structure of the bodies is in direct relation with their surface composition or not.

3.2.3 Expected data products

Low resolution spectra, covering the whole BP/RP wavelength range, with SNR>100 for each taxonomic class with a characteristic standard deviation for each class. Spectra are calibrated by solar analog stars since we deal with asteroid reflection spectra. Hence solar analog spectra are also required.

3.2.4 Local or temporary storage

The data are currently stored on local disks and observatory archives (TNG) and will not be stored in the MDB. They will be transferred in a structured way to the ESAC disk space provided to GBOG and are not expected to occupy much space: 5 MB in total for the reduced spectra so they pose no storage concern at all.

4 CU5 data products

4.1 Flux standards (SPSS)

4.1.1 Use for Gaia

An observing campaign (Pancino et al., 2012) is ongoing to obtain spectra, absolute magnitudes, and to monitor the constancy of approximately 200 SPSS (EP-007; EP-010). The ultimate goal is to obtain $\simeq$1% precise flux tables of all SPSS that are constant within $\pm$5 mmag, and are calibrated to an accuracy of $\simeq$3% with respect to Vega (FVL-072). These will be used in the CU5 photometric pipeline to calibrate all integrated magnitudes (G, G_{BP}, G_{RP})[1] and BP and RP spectra. A first set of $\sim$100 flux tables will be ready to test the pipelines before launch — albeit at a lower quality than the final release — the rest after launch. The ICD for MDB can be found in MR-002, basically flux tables (wavelength and flux), for each SPSS with their uncertainties are needed for the external calibration part of the pipeline.

4.1.2 Scientific potential

The final flux tables will constitute one of the largest homogeneous database of SPSS in the literature, and the only one with SPSS monitored for constancy on short (1–2 h) and long (3 yrs) timescales within $\pm$5 mmag. The absolute magnitudes will be an added value. A critical compilation of literature data will be helpful for the scientific use of the flux table, which could

[1]The same SPSS will help in the flux calibration of G_{RVS} and RVS spectra (ST-002; GMS-004).

be used also to derive the fundamental parameters of the SPSS, together with the obtained spectra. The light curves of the detected variable stars will have their own scientific value. We intend to publish a series of MNRAS refereed papers presenting various technical and scientific results, so a reference website gathering all the published data and measurements would be extremely useful.

4.1.3 Expected data products

- Flux calibrated spectra of $\simeq$200 SPSS, covering the whole BP/RP wavelength range, with R=1000–2000 and S/N>100 except on the spectral border, calibrated in flux with a precision of $\simeq$1% and an accuracy of $\simeq$3% with respect to Vega, with their uncertainties;
- absolute magnitudes for each SPSS, obtained by comparing and combining at least three independent observations;
- light curves on short-term (1–2 h for all WD and hot subdwarf candidates) and long-term (3 yrs, for all SPSS candidates);
- critical compilation of literature data of all candidate SPSS;
- raw data and intermediate reduced data which originated the above products.

4.1.4 Local or temporary storage

SPSS data are at the moment stored on three different locations:

- our local archives in Bologna for raw[2] and reduced[3] data, with a simple web interface and password protected (EP-008);
- an incremental backup of the local archives is scheduled every week on the ESAC disk space provided to GBOG;
- an incremental backup of the local archives is scheduled every week on the ASDC disk space provided to DU13; this backup is used to design a more professional database and archive hosted by ASDC, with a web interface for browsing and querying the data; this database and archive will become public and could be a pilot project for Gaia auxiliary data publication.

The expected size of the final database and archive at ASDC will be smaller than 1 TB approximately, without considering versioning of the SPSS data relevant to each Gaia data release. The

[2] `http://spss.bo.astro.it/raw.cgi`
[3] `http://spss.bo.astro.it/red.cgi`

total disk space depends on how many Gaia data releases there will be, but in no case we expect to need more than 2–3 TB.

4.2 Science alerts verification campaign

4.2.1 Use for Gaia

We plan to undertake ScienceAlerts verification in two distinct phases. Phase-I will be during the ecliptic-poles scanning campaign at the beginning of Gaia operations. Phase-II will be started 3 months after Gaia begins observing, when a significant fraction of the sky has been repeatedly observed, (which will enable us to better distinguish between variables and true transient events). This phase could last 3 months.

Phase-I We expect to detect limited numbers of transients in the first part of the mission. We will apply for a small amount of telescope time to begin the building of our own Gaia-based classification training set, and to test the end-to-end system and suitability of the various instruments.

Phase-II Will comprise a major ground-based observing campaign, ideally from Northern and Southern hemispheres, probably via the INT (with IDS), ESO NTT and suitable imaging telescopes (Liverpool John Moore's, LCOGT for example). The goals here are to:

1. Verification of the alerts.
2. Test and fine-tune the AlertPipe detection threshold.
3. Test and quantify the AlertPipe classification strategies.
4. Significantly improve our training sets by observing significant numbers of transient phenomena from various classes.

Once the ScienceAlerts data products are demonstrated to be scientifically useful and well described, then we will start releasing them to the community at large, hence this verification is a requirement.

4.2.2 Scientific potential

The methods of classification and the full data set s and results will be published to the community at large in a timely (fast) fashion, to enable feedback and comments early in the lifetime of Gaia. If we do our job well, then the scientific community will understand the full potential of Gaia as a transient discovery and classification machine and will respond accordingly. We expect our results during verification to impact significantly on the remaining 4 years of the

Gaia mission, and the resulting data set to have legacy value for future transient time-domain surveys such as LSST.

4.2.3 Expected data products

We aim to build a large library of well classified templates, with:

- high-precision Gaia G-band photometry (multi epoch), NB with preliminary calibration.
- Gaia BP/RP photometry (ideally multi-epoch), although with limited calibrations.
- Gaia cutouts.
- Ground-based multi-colour multi-epoch imaging/photometry.
- Ground-based low-/intermediate-dispersion spectroscopy (multi-epoch).

This will be accompanied by a publication, and a release of the data set to the community at large.

4.2.4 Local or temporary storage

The amount of expected data is rather modest. For imaging, let us assume that we will obtain of the order of 1000 alerts with approximately 10 data points each, in 5 bands. Each data-point is produced based on follow-up observation each containing, say, $\simeq$500 stars for calibration. This makes 25×10^6 measurements, which is probably less than 1 GB in total to store. For spectroscopy, if a large amount of time is obtained to follow-up the same $\sim$1000 objects, say 30 nights on two telescopes (for example IDS@INT or EFOSC2@NTT), amounting to 2 GB/night in total, including raw and processed data, the grand total would not exceed 100 GB. Let us consider less than 0.5 TB if we later decide to store the raw images as well.

Locally, we currently envisage storing the data products on the IoA ScienceAlerts cluster. We have redundancy and a backup strategy in place for the cluster. We will need full browsing capabilities (i.e., most probably a database structure and a web interface) to enable all collaborators — and later, the scientific community — to access the data. It remains to be decided if it is worth publishing the data through CU9, it also depends on manpower and resources.

In any case, the data will most probably not be public during the verification phase, but the results will be published on an international refereed Journal with electronic material at CDS or in similar form.

5 CU6 data products

As described in (DK-015), the auxiliary data used in the CU6 data processing consist of:

- a compilation of atmospheric parameters;
- a catalogue of G_{RVS} magnitudes;
- libraries of spectra in the RVS range (synthetic and observed[4]);
- compilations of radial velocities, including the main catalogue of RV standard stars plus asteroids velocities;
- a compilation of rotational velocities.

Probably, not all these auxiliary data are worth to be delivered by CU9. Not considered here — because not coordinated by GBOG — are: *(i)* atmospheric parameters compiled from various sources in the literature which are only used at the beginning of the processing until CU8 provides the APs of all the observed stars, *(ii)* catalogues of stellar radial velocities compiled for verification purposes at the beginning of the mission, *(iii)* G_{RVS} magnitudes estimated from IGSL, *(iv)* the library of synthetic spectra provided by CU8. The compilation of rotational velocities might be worth to consider but there is not yet enough information on its status to do it now. The auxiliary data coordinated by GBOG for CU6 are the catalogue of radial velocity standard stars and the library of observed spectra. They have been built specially to fill the CU6 requirements and they required new ground-based observations.

5.1 The catalog of Radial Velocity Standards Stars (RV-STD) for Gaia

5.1.1 Use for Gaia

If based only on Gaia data, the RVS wavelength calibration is mathematically degenerate: i.e. a shift of all the RVs can be compensated by a ‘shift’ of the dispersion law. The degeneracy can be removed by the use of ground-based RV-STDs which provide the zero-point of the radial velocity scale, which in turn provides the zero-point of the dispersion law. The CU6 DU630 Calibration module Reference Selection selects F, G and K stars as reference stars for wavelength calibration because they display deep and narrow lines for which centroids can be accurately determined. The specifications for RV-STDs and their ground-based observed RV

[4]The observed spectra are coming from CU3, CU6, CU7 and CU8 GBOG observing programmes: CU3 SEP spectra (Section 2.2), CU6 NEP spectra for commissioning (Section 5.2), CU7 spectra of LPVs and Be stars (Section 6.1) and CU8 OBAFGK benchmark and AP reference stars and emission-line stars and other stars with extended parameters (Section 7).

accuracy are described in DK-015. When RV-STDs are observed by RVS and selected as wavelength calibration reference stars for a particular CCD and FoV by Reference Selection, their ground-based heliocentric RV will be converted to a Gaia-centric RV given the time of the observation. This Gaia-centric RV is then injected into the A and b matrices of the $A \cdot x = b$ linear system in the CU6 DU630 Calibration module Wavelength Calibration, where x is the vector that contains the coefficients that relate wavelength to field angles for the CCD and FoV being calibrated (see HEH-005 for more details).

Since the zero-point might differ according to the spectral type of the stars, it is important to have also O, B, and A RV-STDs with the same constraints as for F, G, and K stars. Such a set has not yet been assembled.

5.1.2 Scientific potential

RV-STD are important to *(i)* establish a velocity scale to which observations coming from various spectrometers can be linked, *(ii)* estimate the accuracy of such observations. At IAU, there is a dedicated WG on Radial-Velocity Standard Stars under Division IX Commission 30 in charge of following the activity on that subject. A new catalog as the one we built for Gaia will be useful for other projects. It is the largest catalog of RVs with uniform distribution on the sky, covering well the HR diagram, and with a good magnitude interval ($6 \leq V \leq 10.5$ mag). A series of A&A papers is in preparation following the first one describing the selection of the 1420 candidates (Crifo et al., 2010): one to present the pre-launch version of the catalog, one to set the RV scale with asteroids, one to take into account the convective blue shift and gravitational red shift. It is planned to determine the APs of the stars for which we have spectra which have not been taken with the Thorium-Argon simultaneous exposure, which is the case for our GBOG observations on ELODIE, SOPHIE and NARVAL.

5.1.3 Expected data products

The catalogue consists of two tables. One table gives the 10220 individual measurements (Julian date, barycentric RV, error) taken with the 5 instruments (new observations and archived data from SOPHIE, CORALIE, NARVAL, HARPS, ELODIE). The second table gives for each of the 1420 stars the weighted average of the RV measurements expressed in the SOPHIE scale, with error, standard deviation and time baseline.

5.1.4 Local or temporary storage

All the observational material related to the catalog of RV-STD is stored in a dedicated database maintained at AIP. This includes the reduced 1D spectra and measured CCF coming from our GBOG observations on ELODIE, SOPHIE and NARVAL and from the archives of these instruments. For CORALIE, we have only the measured CCFs. We have also retrieved relevant spectra and CCF in the ESO-HARPS archive.

5.2 Library of observed spectra in the RVS range

5.2.1 Use for Gaia

CU6 DU630 Calibration modules Wavelength Calibration and AL LSF Calibration currently rely on template spectra. However, given that uncertainties in spectra modelling in the templates will generate errors on the calibration that propagates as systematic calibration errors to all the observations, the best templates that could be used in CU6 processing are the spectra of the stars themselves, observed from the ground. Using observed templates removes the dependency on CU8 stellar parameters to choose synthetic templates, which will not be available until cycle 03 (CBJ-054). The observed templates can also be used a-posteriori to check the quality of the processing: e.g., background modelling and subtraction, de-blending and RVS calibration.

The observed templates will be especially useful during commissioning. This will take place during the Ecliptic Pole Scanning Law. Therefore it is optimal to select the ground-based spectra as observed templates in the vicinity of the ecliptic poles, as these will be the stars observed the most frequently by the RVS during the first two months of the mission. These observed templates may include GBOG South Ecliptic Pole spectra (see Sect. 2.2).

5.2.2 Scientific potential

RVS empirical templates will be extracted from NARVAL spectra which cover the full optical range from 370 to 1 050 nm at a resolving power of $\simeq$65000 or 80000 depending on the observing mode. Together with the RVS empirical library, a more general library covering a wider spectral range could be delivered. Such empirical libraries have many application as reference spectral datasets, e.g., testing methods of AP determination, improving atmosphere models, calibrating spectrocopic indices, or feeding population synthesis codes (see also Section 7).

5.2.3 Expected data products

We have a dedicated GBOG programme on NARVAL to obtain high resolution, high SNR spectra for 63 stars selected at the NEP for the initialization phase. We will also compile all our NARVAL observations of RV-STD plus CU8 reference stars, plus all the relevant data to be gathered for the NARVAL archive. This compilation will result in an homogeneous empirical library of stellar spectra.

5.2.4 Local or temporary storage

The RAdial Velocity Experiment database, which is not yet public, plots each observed spectrum and its corresponding template. If CU9 also do this and CU6 use the GBOG library of observed spectra instead of templates for these stars, then the GBOG library of observed spectra will automatically be published in the CU9 Gaia Archive.

6 CU7 data products

The CU7 activities are mainly based on the use of existing databases for variable stars, such as the OGLE, EROS, ASAS, and Hipparcos catalogues; however, ancillary data are also being collected using the CU7 network of telescopes (this includes 9 telescopes from 50 cm to 2 m in size, distributed in a number of different countries in Europe and including also facilities in Chile and New Mexico) for a number of specific programs (see Section 6.1.1), and also as contributions to the validation of the GSA system (see also Section 4.2) in collaboration with CU5, and to the GBOT program for monitoring the Gaia satellite (see also Section 2.1). Initial tests of the GSA pipeline were carried out by CU7 telescopes using as triggers alerts issued by the Catalina transient sky survey and for the GBOT program by obtaining multi-epoch photometric observations of the Planck satellite (see the respective sections for more details).

6.1 Supplementary observations of variable sources

6.1.1 Use for Gaia

A number of monitoring programs are being carried out on specific types of pulsating variables (namely, Cepheids, RR Lyrae stars, Long Period Variables – LPVs, Be stars, short period variables) to test the algorithms and procedures being developed to classify and characterize the variable sources that will observed by Gaia. They include:

- multi-band CCD photometry with the 1 m telescope of the Konkoly Observatory (Hungary) of 25 Galactic Cepheids of various subtypes to follow up period changes; and "lucky imaging" with the EMCCD camera of the 50 cm telescope of the Konkoly Observatory of 76 Galactic Northern Cepheids, to search for faint close (optical) companions with the goal of possibly defining a general procedure for the identification of binary/multiple Cepheids;
- photometry of ultra long period Cepheids (periods exceeding 100 days) with the DOLORES Camera of the Telescopio Nazionale Galileo and the GMOSS of the Gemini South telescope, to characterize these objects that are often saturated in existing databases such as OGLE or EROS;
- time-series photometry of the Ecliptic Poles regions. The program on the NEP is based on V, I_C photometry obtained with a 30 cm robotic telescope in New Mexico, USA, plus the twin telescope of the Tartu Observatory in Estonia. About 700 stars were monitored down to V$\simeq$15.6 mag. Forty-seven candidate variables were identified, of which 24 deserve further investigation. The program on the SEP is

based on K_s time-series photometry being obtained with the VISTA telescope (in the framework of the VISTA Magellanic Cloud ESO Public Survey, VMC) along with EROS 2 optical data and the recently published OGLE IV V, I_C time-series photometry of the Gaia SEP field (Soszyński et al., 2012). Time series photometry in B, V, R, I of unclassified variable stars in the LMC 562 and 563 fields (Soszyński et al., 2012) with the DFOSC/DK-154 photometer/telescope instrumentation at La Silla, Chile. The aim is the full characterization of the variable stars present in these regions (with particular interest for Cepheids and RR Lyrae stars) in order to test the performance of the classification and characterization algorithms, as well as the capability to recover multiple Cepheids even in case of sparse observations;

- spectra of 15 LPVs (mainly Miras and irregular variables) in the spectral range of RVS at a resolution around 10 000 obtained with the Perek 2m telescope (P2mT) of the Ondrejov Observatory (Czech Republic);
- photometry of short period variables that may be detected on a per-CCD photometry or from the knot points of the Gaia scanning law, obtained with Eddington fast-read CCDs mounted on the 1.2 m Euler telescope in La Silla, Chile, and the 1.2 m Mercator telescope in La Palma, Canary Islands;
- medium resolution spectroscopy of 180 Be stars in the H_α and RVS spectral regions obtained with the P2mT telescope of the Ondrejov Observatory (Czech Republic).

6.1.2 Scientific potential

The ancillary data being collected for the aforementioned programs will often form the first catalogues of specific classes of variable stars in specific photometric bands, or the first spectroscopic information ever obtained. Furthermore, by expanding the time baseline of existing databases they will allow a better definition of the stars' periods, as well as the study of period variations and the detection of amplitude changes due to specific phenomena, such as the presence of companions in the case of Cepheids, or the Blazhko effect for RR Lyrae stars. Results from these studies are being published on refereed journals (e.g. Ripepi et al., 2012; Szabados & Nehéz, 2012). At least one paper is planned based on the data of the variability search near the NEP.

6.1.3 Expected data products

- Reduced multiband time-series photometry of different types of variable stars. Specifically, multicolor CCD photometry of about 300 Galactic Cepheids; multisite observations of subdwarf variable sources; K_s-band, B_{EROS}, R_{EROS}, V, I_C light curves for 130 RR Lyrae stars, 26 Cepheids, 30 δ Scuti, 530 LPVs and about 260 binaries in the SEP field; V, I_C light curves for the NEP variables;

- light curves, periods/frequencies, amplitudes of the light variations and average magnitudes;
- reduced spectra;
- radial velocity curves, amplitude of the radial velocity curve and average radial velocity;
- a catalogue of Be stars;
- cross-identification with existing photometric/spectroscopic catalogues of variable stars.

6.1.4 Local or temporary storage

A large fraction of the observations mentioned above are being obtained with facilities that are part of the CU7 network of telescopes. They are stored locally and made publicly available (after verification) either through dedicated local web pages (e.g., for the NEP data the dedicated local web-page of the Tartu Observatory at `http://www.aai.ee`), or upon request to the colleagues responsible for the specific program/data set . Data from existing surveys (such as OGLE, EROS, VMC, Hipparcos, ASAS, etc.) are stored in the repositories of each specific survey and generally have open access. The OGLE, ASAS, EROS and Hipparcos data are also stored in the CU7 database in Geneva. At the present time we do not foresee to store these data in the MDB or to publish them through CU9.

7 CU8 data products

The General and Extended Stellar Parametriser algorithms – the core of the CU8 software – are based on synthetic spectra. The purpose of the ground-based observations for CU8 is to assess the realism of the underlying stellar atmospheric models and spectral line formation codes, and to provide an external empirical calibration of astrophysical parameters (APs). The calibration stars are divided into two categories, which we call “benchmark stars” and “AP reference stars”. The first category comprises a number of bright well-known stars covering the HR diagram (on the order of 50 – 100 stars). Observing campaigns are ongoing to obtain high-resolution, high SNR spectra in the optical and near-IR wavelength regions for all of them. The AP reference stars are an extended sample of fainter stars, which will be observed by Gaia, but are also accessible for high-resolution spectrographs. These improve the AP coverage compared to benchmark stars. The spectral data are complemented by independent compilations of APs for all stars. The ICD for the observed spectra and associated parameters can be found in UH-002. In the following, we describe data products for two spectral-type categories, FGK- and OBA-type stars.

7.1 Benchmark stars (FGK-type stars)

7.1.1 Use for Gaia

The current sample of FGK-type candidate benchmark stars comprises about 40 stars. The sample and their APs are being updated on a regular basis. The spectra and APs of these stars are being used for critical tests of the models used by the CU8 algorithms, and these stars serve as reference objects for the analysis of the FGK-type AP reference stars.

7.1.2 Scientific potential

The final set of benchmark stars will for the first time provide the possibility to calibrate spectroscopic astrophysical parameters for large and diverse samples of stars. Unlike in the field of photometry or radial velocities, up to now there are no standard stars defined for stellar spectrum analyses. The Sun has always been the single common reference point for studies of FGK-type stars. A first application will be seen in the use of benchmark stars within the Gaia-ESO public spectroscopic survey.

7.1.3 Expected data products

- Spectra with high resolution ($R \geq 80000$) and high SNR (≥ 300), between 300 and 1000 nm, wavelength calibrated and flux-normalized to the local continuum.
- Information about the observations (telescope, instrument, observer, ...) and the targets (coordinates, radial velocity, ...). See ICD for possible parameters.
- Broad-band photometry compiled from the literature.
- Critical compilation of effective temperatures, based on direct observables such as angular diameter and bolometric flux.
- Critical compilation of surface gravities, based on direct observables such as angular diameter and parallax, on dynamical determinations in the case of binaries, or on asteroseismology.
- Homogeneous determinations of metallicities and element abundances from the ground-based observations.

7.1.4 Local or temporary storage

The data are currently stored on local disks and observatory archives (ESO, TNG, TBL). They will be transferred in a structured way to the ESAC disk space provided to GBOG. One complete set of data (one extracted spectrum per instrument for each star) occupies about 1 GB of disk space. If we aim to store two sets of data (intermediate and final processing products), we will thus need 2 GB of disk space for this part of the GBOG data. With minor revisions, the information in the MDB will be sufficient to publish these data in the CU9 archive (see UH-002 and related Mantis issues). The library of spectra for FGK benchmark stars will be published in a scientific journal, including description, example application and link to data in a format useful for the community.

7.2 AP reference stars (FGK-type stars)

7.2.1 Use for Gaia

FGK-type AP reference stars are about 500 field stars, open cluster stars, and stars in the SEP field, which will be observed by Gaia. For these stars, parameters determined from ground-based observations will be compared directly to those determined by the CU8 algorithms from the Gaia data. This provides an additional external AP calibration. The stellar samples are described in UH-001, and the ground-based data obtained and gathered from archives so far are described in CS-008.

7.2.2 Scientific potential

Here, we give some examples for potential scientific applications. The abundance analysis of open cluster stars will contribute several data points for the study of the chemical structure of the Galaxy. The SEP sample will allow to characterize different stellar populations in the Milky Way and the LMC, and the relation between these two neighboring galaxies. Stars in LMC clusters of different age will help to constrain the timescale for the dissolution of clusters into the field stellar population.

The AP reference stars can also serve as a pool from which to draw common calibration star samples for ongoing and planned stellar spectroscopic surveys.

7.2.3 Expected data products

- Spectra with medium to high resolution ($R \geq 20000$) and high SNR (≥ 100), between 300 and 1000 nm, wavelength calibrated and flux-normalized to the local continuum.
- Information about the observations (telescope, instrument, observer, ...) and the targets (coordinates, radial velocity, ...). See ICD for possible parameters.
- Broad-band photometry compiled from the literature.
- Critical compilation of literature data for effective temperature, surface gravity, and metallicity.

- Homogeneous determinations of effective temperature, surface gravity, and metallicity from the ground-based observations, using benchmark stars as reference objects.

7.2.4 Local or temporary storage

The data are currently stored on local disks and observatory archives. They will be transferred in a structured way to the ESAC disk space provided to GBOG. We do not have a complete overview of the data volume yet. Extrapolating the estimate for the benchmark stars library (a factor of 10 for the number of stars, and a factor of 2 for the number of useful spectra per star), we arrive at a disk space of 40 GB for this part of the GBOG data. With minor revisions, the information in the MDB will be sufficient to publish these data in the CU9 archive (see UH-002 and related Mantis issues). It is not yet decided if and how the library of spectra for FGK reference stars will be published to the community.

7.3 Benchmark stars (OBA-type stars)

7.3.1 Use for Gaia

An observing campaign is ongoing at the ROB for collecting high-resolution spectra ($R \simeq$80000) of SNR$\simeq$1000 in the Gaia-RVS wavelength region with the HERMES spectrograph on the Mercator Telescope at La Palma. Four spectra of A-type stars, 5 of B-type stars have been observed between 2009 and 2012, while another 5 A- and B-type spectra will be obtained between 2012 and 2015. Six more benchmark spectra of selected O-type stars will be proposed for observation commencing around fall of 2012.

The benchmark OBA spectra are required for the development and training of algorithms for the calibration of APs of hot stars in ESP-HS (DPAC-CU8: WP Extended Stellar Parametrizer - Hot Stars). These data will complement ongoing work in the GBOG WP for FGK- (cool) stars (CS-008). The 20 benchmark stars are bright hot stars (V>7 mag) with well-known (literature) APs for quality testing and/or to improve spectrum synthesis codes in the Gaia-RVS wavelength domain. They are carefully selected narrow-lined stars having optical spectra with vsini <50 km/s. These stars will cover the O, B, and A spectral subtypes as well as possible within the total observing time available on HERMES.

7.3.2 Scientific potential

No description of the scientific potential of these data was provided by the CU8 authors.

7.3.3 Expected data products

1. Continuum flux normalized and wavelength calibrated spectra covering the RVS wavelength range with $R \simeq$80000 and SNR$\simeq$1000. The spectra will be provided in the heliocentric wavelength frame.

2. Information about the HERMES spectra observations (telescope, instrument, observing conditions) and pipeline calibration frames (flat, bias, ThAr frame dates, observation sequence numbers, and calibration pipeline version), from the FITS file headers.

3. A compilation of astrophysical parameters provided in the literature for all observed benchmark targets. In case the literature values can be improved they are complemented with APs derived from detailed synthetic spectrum fits to the RVS region or to other parts of these benchmark star spectra.

4. A list of absorption lines identified in the RVS spectra based on the synthetic spectrum calculations using literature APs or improved AP values. The line ids. (species, observed and rest wavelength) can be supplemented with a list of equivalent line width values in case of significant line detection (e.g. exceeding $\sim$10 mÅ, with respect to the H Paschen lines), together with the atomic line data utilized for the detailed spectrum synthesis calculations.

7.3.4 Local or temporary storage

Mercator-HERMES spectra of OBA Benchmark stars are currently stored at the ROB. The 20 RVS spectra with metadata will occupy 1 GB at most. The data will also be offered through the online SpectroWeb database (`http://spectra.freeshell.org`), which provides a user-friendly web-interface for interactively displaying both observed and synthetic spectra of each star, combined with spectral line identifications and the corresponding atomic line data.

7.4 AP reference stars (OBA-type stars)

7.4.1 Use for Gaia

A program is carried out at the ROB for collecting high-resolution optical and near-IR spectra ($R >$50000) with SNR$\simeq$50 of AP reference hot stars that also include the Gaia-RVS wavelength region. The spectra will complement benchmark OBA star spectra required to improve the astrophysical parameters ($T_{\rm eff}$, log(g), [Fe/H], $V_{\rm mic}$, vsini) coverage for the calibration of hot star spectral grids employed in Gaia data analysis. We currently expect to observe 50 to 100 AP reference OBA spectra in the Mercator-HERMES 'filler' program of P. Royer (Univ. Leuven) & A. Lobel (ROB) (2009 – 2014). Note that the selected hot stars may have shallow absorption lines (e.g., vsini>50 km/s) compared to the benchmark hot stars (see above) because target selection in the HERMES filler program chiefly depends on lesser seeing conditions or the visibility of bright hot stars for non-standard observations at dusk and dawn.

7.4.2 Scientific potential

The AP reference OBA stars spectral catalog will have great scientific potential for Gaia data analysis, but also outside the Gaia context. The data will be very valuable for training the Gaia ESP-HS algorithms, currently under development in CU8, but they will also provide important spectroscopic reference information available outside the Gaia-RVS wavelength region. The detailed profiles of important AP calibration lines, such as the extended wings of the Balmer H_γ and H_δ lines (outside RVS), will be useful to test the accuracy of surface gravity (log g) values derived from H Paschen lines in the RVS domain. The high-resolution catalog will be very useful to, for example, estimate possible line opacity effects in the RVS region for anomalous abundance stars such as in Am (metallic-line)-type stars observed with RVS spectral resolution.

7.4.3 Expected data products

1. Continuum flux normalized and wavelength calibrated merged spectra sufficiently covering the optical and near-IR wavelength ranges, including the RVS wavelength domain with R>50000 and SNR>50. The spectra will be provided in the heliocentric wavelength frame.

2. Information about the high-resolution spectroscopic observations (telescope, instrument, observing conditions) and pipeline calibration frames (flat, bias, ThAr frame dates, observation sequence numbers, and calibration pipeline version), typically also provided in the FITS file headers.

3. A compilation of astrophysical parameters provided in the literature for all observed AP reference stars. In case available literature values can be improved they are complemented with APs derived from detailed synthetic spectrum fits to the RVS region or to other parts of the AP reference spectra.

4. A list of absorption lines identified in the RVS wavelength region, based on the synthetic spectrum calculations using literature APs or improved AP values. The line ids. (species, observed and rest wavelength) can be supplemented with a list of equivalent line width values in case of significant line detection (e.g. exceeding $\sim$10 mÅ, with respect to the H Paschen lines), together with the atomic line data utilized in the detailed spectrum synthesis calculations.

7.4.4 Local or temporary storage

The high-resolution spectra of AP reference OBA stars are currently stored at the ROB. We estimate an inventory of 50 to 100 AP reference hot star spectra that, together with the metadata, will occupy 100 GB at most. The Mercator-HERMES spectra will also be offered in the HHRespect database, currently under development at the Univ. of Leuven. The latter database will offer advanced data access tools for user-friendly (raw and pipeline calibrated) retrieval of the echelle data frames, together with graphical tools for database users for displaying the offered spectra in various wavelength regions.

7.5 Emission-line stars and other stars with extended parameters

7.5.1 Use for Gaia

New reference spectra for various types of emission line stars are/will be collected at high-resolution and good SNR. Additionally, data is also collected from currently available databases. These datasets are used to train, develop and test the ESP-ELS algorithm aiming to classify the emission line stars.

Being able to identify emission line stars from the huge amount of Gaia data will help to conduct statistical studies on the appearance of certain peculiar ELS and to better understand the origin and formation of the spectroscopic emission features. However, to correctly and unambiguously interpret the observations, one needs also to have access to the reference data used to classify the ELS.

7.5.2 Scientific potential

No description of the scientific potential of these data was provided by the CU8 authors.

7.5.3 Expected data products

We propose to provide the continuum normalized spectra, a full description of the observations as well as the classification we considered for the reference stars.

7.5.4 Local or temporary storage

The new data are currently stored on a local fileserver at the ROB. At the end of the mission, we hope to gather data for 500 to 1000 reference emission line stars, which should take from 3 to 5 GB of disk space. No other plan for publication has been finalized yet.

8 References

[MA-003], Altmann, A., 2011, *Ground Based Optical Tracking of Gaia*,
GAIA-C3-TN-ARI-MA-003,
URL http://www.rssd.esa.int/llink/livelink/open/3094163

[MA-002], Altmann, M., Bastian, U., 2009, *Ecliptic Poles Catalogue Version 1.1*,
GAIA-C3-TN-ARI-MA-002,
URL http://www.rssd.esa.int/llink/livelink/open/2885828

Andrei, A.H., Souchay, J., Zacharias, N., et al., 2009, A&A, 505, 385, ADS Link

[CBJ-054], Bailer-Jones, C., 2011, *CU8 Operations Manual. Running Apsis*,
GAIA-C8-PL-MPIA-CBJ-054,
URL http://www.rssd.esa.int/llink/livelink/open/3079670

Bourda, G., Charlot, P., Porcas, R.W., Garrington, S.T., 2010, A&A, 520, A113, ADS Link

Bourda, G., Collioud, A., Charlot, P., Porcas, R., Garrington, S., 2011, A&A, 526, A102, ADS Link

Crifo, F., Jasniewicz, G., Soubiran, C., et al., 2010, A&A, 524, A10, ADS Link

[UH-001], Heiter, U., Soubiran, C., Korn, A., 2008, *Calibration of the General stellar parameterizer algorithms using ground-based observations*,
GAIA-C8-TN-UAO-UH-001,
URL http://www.rssd.esa.int/llink/livelink/open/2864468

[UH-002], Heiter, U., Thevenin, F., Fremat, Y., et al., 2009, *Interface control document for Gaia observed spectral libraries*,
GAIA-C8-SP-UAO-UH-002,
URL http://www.rssd.esa.int/llink/livelink/open/2906426

[HEH-005], and others Huckle, H., 2011, *WP630 Software Design Document - RVS Calibration*,
GAIA-C6-SP-MSSL-HEH-005,
URL http://www.rssd.esa.int/llink/livelink/open/2737088

Hwang, N., Lee, M.G., Lee, H.M., et al., 2007, APJS, 172, 583, ADS Link

[DK-015], Katz, D., et al., 2012, *Auxiliary data in CU6 processing*,
GAIA-C6-TN-OPM-DK-015,
DRAFT

[EP-007], Pancino, E., Altavilla, G., Bellazzini, et al., 2011, *Ground-based observations of the Gaia Spectrophotometric Standard Stars. Present status and future steps.*,
GAIA-C5-TN-OABO-EP-007,
URL http://www.rssd.esa.int/llink/livelink/open/3081178

[EP-008], Pancino, E., Altavilla, G., Rossetti, et al., 2011, *The local Bologna archive of SpectroPhotometric Standard Stars observations*,
GAIA-C5-TN-OABO-EP-008,
URL http://www.rssd.esa.int/llink/livelink/open/3081255

Pancino, E., Altavilla, G., Marinoni, S., et al., 2012, MNRAS, 426, 1767, ADS Link

[EP-010], Pancino, E., et al., 2012, *Ground-based observations of the Gaia Spectrophotometric Standard Stars. Incremental 2012 status update.*,
GAIA-C5-TN-OABO-EP-010,
DRAFT

[MR-002], Riello, M., De Angeli, F., 2008, *PhotPipe Interface Control Document*,
GAIA-C5-SP-IOA-MR-002,
URL http://www.rssd.esa.int/llink/livelink/open/2863103

Ripepi, V., Moretti, M.I., Marconi, M., et al., 2012, MNRAS, 424, 1807, ADS Link

[GMS-004], Seabroke, G., Soubiran, C., Pancino, E., Thuillot, W., 2013, *Minutes of the 12th GBOG meeting*,
GAIA-CD-MN-MSSL-GMS-004,
URL http://www.rssd.esa.int/llink/livelink/open/3161294

[RLS-001], Smart, R., Morbidelli, R., Nicastro, L., et al., 2008, *The Initial Gaia Source List 1.0; Production and Plans*,
GAIA-C3-TN-OATO-RLS-001,
URL http://www.rssd.esa.int/llink/livelink/open/2819603

Soszyński, I., Udalski, A., Poleski, R., et al., 2012, Acta Astronomica, 62, 219, ADS Link

[CS-009], Soubiran, C., 2010, *Ground-Based Observations for Gaia. 2010 status*,
GAIA-CD-PR-LAB-CS-009,
URL http://www.rssd.esa.int/llink/livelink/open/3056184

[CS-008], Soubiran, C., Heiter, U., 2010, *Primary AP reference stars for the calibration of the General Stellar Parametrizer algorithms*,
GAIA-C8-TN-LAB-CS-008,
URL http://www.rssd.esa.int/llink/livelink/open/3044507

[CS-004], Soubiran, C., Mignard, F., Allende-Prieto, C., et al., 2008, *GBOG Terms of Reference*,

GAIA-CD-PL-LAB-CS-004,
URL http://www.rssd.esa.int/llink/livelink/open/2856717

Souchay, J., Andrei, A.H., Barache, C., et al., 2009, A&A, 494, 799, ADS Link

Souchay, J., Andrei, A.H., Barache, C., et al., 2012, A&A, 537, A99, ADS Link

[ASP-001], Spagna, A., Castaneda, J., 2009, *IGSL XM Algorithm and Software Description*,
GAIA-C3-TN-OATO-ASP-001,
URL http://www.rssd.esa.int/llink/livelink/open/2848864

Stetson, P.B., Pancino, E., 2008, PASP, 120, 1332, ADS Link

Szabados, L., Nehéz, D., 2012, MNRAS, 426, 3148, ADS Link

Taris, F., Souchay, J., Andrei, A.H., et al., 2011, A&A, 526, A25, ADS Link

[ST-002], Trager, S., 2010, *Spectrophotometric calibration of RVS using CU5-DU13 flux calibration tables*,
GAIA-C5-TN-UG-ST-002,
URL http://www.rssd.esa.int/llink/livelink/open/3027880

Valenti, J.A., Piskunov, N., 1996, A&AS, 118, 595, ADS Link

[FVL-072], van Leeuwen, F., Pancino, E., Altavilla, G., 2011, *GB-Obs Software Requirement Specifications*,
GAIA-C5-SP-IOA-FVL-072,
URL http://www.rssd.esa.int/llink/livelink/open/3086593