CU8 Input Data Format and Content. A discussion document

prepared by: Y.Frémat, C.A.L. Bailer-Jones, R. Drimmel, A.-M. Janotto, F. Thévenin, A. Vallenari, R. Blomme, Y. Damerdji, M. Kontizas, O. Kochukhov, A. Korn, M. Outeiro, A. Recio-Blanco, K. Smith, M. Torres

approved by: Approved by the CU8 Management Team
reference: GAIA-C8-PL-ROB-YF-002-01
issue: 1
revision: 0
date: 2009-01-07
status: Issued

Abstract

We summarize the data inputs needed by CU8 from other CUs in order to classify the detected sources and to determine the astrophysical parameters. The aim is to specify in more detail what is required, as a reference basis for further discussions with other CUs.

Document History

Issue	Revision	Date	Author	Comment
1.0	0	07-jan-2008	CBJ	Renamed, changed SP to PL, modified abstract, altered section organization. Issued
1.0	D	20-dec-2008	YF	Comments made by CBJ
1.0	D	25-nov-2008	YF	Status at CU8 Meeting
1.0	D	7-nov-2008	YF	Received comments from A.-M. Janotto and from F.Thévenin. MDB and processing requirements are now in 2 different sections.
1.0	D	26-oct-2008	YF	Received comments from C.A.L. Bailer-Jones.
1.0	D	21-oct-2008	YF	Frédéric Thévenin, Minia Manteiga Outeiro and Miguel Garcia Torres sent me requirements related to the FLAME, OA and OCA packages.
1.0	D	15-oct-2008	YF	First draft. Integrating comments made by C.A.L. Bailer-Jones, F.Thévenin, A.Korn, A.Recio-Blanco, H.Lindstroem, A.Brown, M.Kontizas, K. Smith, A.-M. Janotto, Y.Frémat.

Contents

1 Front matter

1.1 Scope

This document contains the data specifications for CU8. The scope is to provide a description of the data expected from the other CUs by all the workpackages.

1.2 Applicable Documents

[1] CU5 Software Development Plan, July 2008
[2] CU6 Software Development Plan, August 2008
[3] CU8 Software Requirements Specification
[4] Photometry with dispersed images - overview of BP/RP data processing
[5] Data specification discussion on the wiki

1.3 Reference Documents

[1] F. van Leeuwen and P. Richards, CU5 Software Development Plan, July 2008. GAIA-C5-PL-IOA-FVL-001-4

[2] D. Katz, et al., CU6 Software Development Plan, August 2008. GAIA-C6-PL-OPM-DK-003-04

[3] C. Bailer-Jones, et al., CU8 Software Requirements Specification, October 2007. GAIA-C8-SP-MPIA-CBJ-032

[4] A. Brown, Photometry with dispersed images - overview of BP/RP data processing, GAIA-C5-TN-LEI-AB-009

[5] C.Bailer Jones et al., Data specification discussion on the wiki, `http://www.rssd.esa.int/wikiSI/index.php?title=CU8:_Data_specifications`

[6] C. Jordi, et al. Error model for the mean BP/RP spectra for GOG cycle 4 simulations, GAIA-C5-TN-UB-CJ-043-1

1.4 Acronyms

The following is a complete list of acronyms used in this document.

Acronym	Description

AGIS	Astrometric Global Iterative Solution
AP	Astrophysical Parameter
CU	Coordination Unit (in DPAC)
DPAC	Data Processing and Analysis Consortium
DSC	Discrete Stellar Classifier
ESP	Extended Stellar Parametrizer
FLAME	Final Luminosity And Mass Estimator
GOG	Gaia Object Generator
GSP	Generalized Steller Parametrizer
ICD	Interface Control Document
MDB	Main Database
MSC	Multiple Star Classifier
OCA	Object Clustering Analysis
QSO	Quasi Stellar Objects
RV	Radial Velocity
RVS	Radial Velocity Spectrograph
SI	International System of Units
SSU	Specific Source Unit
UGC	Unresolved Galaxy Classifier

2 Processing requests per CU

2.1 Data requested from CU2

2.1.1 GOG

GOG should provide the fluxes and wavelengths in the same units (SI) than those used by CU5 in the MDB.

2.1.2 Auxiliary data

All the calibration data (e.g. AP calibration) and auxiliary data needed for AP determination (e.g. synthetic spectra and observed spectra used by classical minimum distance methods) must be stored in the MDB (DPAC requirement). These auxiliary data should take into account the instrument PSF and instrument response, it is therefore important that CU2 continues to be operational after launch, till the end of the mission.

2.2 Data requested from CU3

2.2.1 Astrometry

- **Proper motion** and corresponding error.
- **Star coordinates** and **algorithms** for catalogue cross-matching. These will be used, for example, by the AP calibration procedure.
- **Parallax and corresponding error**.

2.3 Data requested from CU4

2.3.1 Binaries

- **Total mass** and **mass ratio** for FLAME, as well as their corresponding errors. This requirement only concerns some binaries for which sufficient astrometric, variability (e.g. eclipsing binaries) and spectroscopic data are available. This is compatible with the fact that FLAME is an end-of-mission module.

2.3.2 Detection of unknown objects (UFOs)

- **Label for unknown objects**. Unknown objects will be labelled with an identifier that reflects their outlier nature. A direct input from this set of UFOs is required by

the Outlier Analysis package.

2.4 Overview of what CU5 is planning to provide

FLUX CALIBRATION

CU5 makes the distinction between internal and external calibration.

The **internal calibration** corrects the data for all possible instrumental effects and produces fluxes that are self-consistent and therefore *relatively stable* in time. The "bell shape" of the internally calibrated BP/RP spectra is the result of the internal calibration, it reflects all the instrumental effects affecting the original SED. There is no plan to remove these effects during the internal calibration step, this would be too inaccurate and time consuming.

The **external calibration** will set the zero-point of the flux scale in order to be transformed into physical flux units (W m^{-2} nm^{-1}). This absolute calibration will also get rid of the "bell shape" and reconstruct the true shape of the spectrum, at the low resolution of the Gaia spectrophotometry (dispersion of 4-32 nm in the BP and 7-15 nm in the RP). (Taken from a mail exchange with Carla Cacciari). See also [1].

According to A.Brown (see Data Specification page on the wiki [5]), *the externally calibrated spectra will definitely go into the MDB. It is not decided yet whether the internally calibrated spectra should also go to the MDB (most likely not).*

WAVELENGTH CALIBRATION

The wavelength dispersion curve for BP/RP and the PSF are expected to vary because the prism surfaces are not exactly parallel to the CCD. An accurate wavelength calibration may therefore not be available until late into the mission. The mean spectra (composed of all the individual epoch spectra for each source) will further be oversampled as the information in the data allows this (at least for the brighter sources). The precise sampling and whether or not the wavelength scale will be linear has not been decided yet. CU5 is not planning to correct for RV shifts (Comment made by A.Brown on the wiki [5]).

MDB AND EPOCH DATA

Before producing a combined spectrum the epoch spectra will have to be transformed to a common flux and wavelength scale. So in principle epoch spectra that can be compared to each other will be available. Whether or not they will go into the MDB (and when) is not decided yet. One advantage of CU8 analyzing the epoch spectra would be a consistency check on the APs and also on the disentangling of fluxes for overlapping spectra in crowded regions (Comment made by A.Brown on the wiki [5]).

COMBINED SPECTRA AND PIXEL-TO-PIXEL CORRELATION

Since the epoch spectra will have different dispersion functions due to their different positions on the detector and that many manipulations are required during the extraction and the reduction, a pixel-to-pixel correlation is expected in the mean spectrum. A full covariance matrix of errors could be provided to account for these correlations (Comment made by A.Brown on the wiki [5]).

COSMIC RAY REMOVAL

The aim of CU5 is to provide ”clean” spectra which includes the removal of cosmic rays. Any uncertainties introduced in the removal procedure should be reflected in the error vector (Comment made by A.Brown on the wiki [5]).

2.5 Data requested from CU5

2.5.1 Wavelength calibrated BP/RP spectra and covariances

- **Sky subtracted BP/RP data.**
- **Externally calibrated data (in SI units)** to take advantage of the parallaxes or to adopt more classifical classification criteria.
- If the wavelength scale is expected to be different from one exposure to another, **fluxes AND wavelengths** are mandatory.
- Depending on how much the wavelength dispersion varies, CU8 may be **able to work early** in the processing **without an accurate wavelength calibration** and even data not on an absolute wavelength scale (just a common one).
- CU8 will carry out an EPOCH by EPOCH analysis. **EPOCH spectra should be available during the whole mission**. For calibration and practical reasons, we cannot wait 2 years before getting CU5 data. In particular, the ecliptic pole data should be available sooner. Predictions of the **effective Gaussian 1-sigma noise** spectrum is mandatory.
- CU8 would like to have access to the **epoch spectra before the final calibration**, i.e. we would like to have access to the internally calibrated (flux and wavelength) spectra early in the mission. We additionally assume that the **finally calibrated epoch spectra will be placed in the MDB** (as we presumably want to release these as a Gaia data product).
- CU8 will require BP/RP data early in the mission processing cycles. If **externally calibrated data** are only available relatively late in the mission processing, then **we request** that **internally calibrated** data be made available to CU8.

- A covariance matrix that reflects the pixel-to-pixel flux correlations (i.e. a measure of the flux variances and covariances).
- **Oversampling of the multiple epoch BP/RP** by a factor that still needs to be determined. The **same oversampling** is required for all the spectra, **regardless of magnitude**.
- For **larger RV shifts**, we need to correct the BP/RP wavelength scale.
- **Cosmic ray hits** should be **removed** from the spectrum.

2.5.2 Gaia magnitude and error

- **G, G_BP and G_RP** magnitudes and errors.

2.5.3 Science alerts

- OA deals with objects that were not recognized as known astronomical source types. Information on science alerts from other CUs is thus mandatory.

2.6 Data requested from CU6

2.6.1 Wavelength calibrated RVS spectrum in rest frame and covariances.

- **Sky subtracted RVS spectrum** to check for possible pollution from neighbouring objects.
- Non-(pseudo)continuum-normalized **flux calibrated spectra**. The slope of the RVS spectral range contains indeed information about the effective temperature and interstellar extinction. We also have to take into account the fact that the spectrum normalization of emission line stars may fail.
- **Continuum normalized spectrum**. By *normalized spectrum*, we mean the spectrum divided by the local (pseudo) continuum.
- Same wavelength calibration as the one used for the RV determination. The **wavelength scale** should further be **corrected for RV shift** (we need the spectrum in the stars' own rest frame). To estimate the accuracy of this correction, we further need an estimate of the RV accuracy.
- **EPOCH spectra** should be available during the whole mission, and **effective Gaussian 1-sigma noise** spectrum.

- A **covariance matrix** to take into account the pixel-to-pixel flux correlations (i.e. a measure of the flux variances and covariances).
- **Cosmic ray hits** should be **removed** from the normalized and non-normalized spectra. Pixels affected by the removal should be flagged.
- **G_RVS magnitude and error**.
- If the wavelength scale is expected to vary, **fluxes AND wavelengths** are required for each source.

2.6.2 Projected rotation velocity and error

- $\mathrm{V}\sin i$ **and corresponding error values expressed in km/s**. The auxiliary data (Synthetic spectra) and/or the B-factors (GSP-Spec) need indeed to be broadened to account for rotational broadening.

2.6.3 Emission flag and any available related information

- The **emission flag** is required. It tells if emission was detected in the RVS and provides also its level of significance.
- Any information on emission detected in the RVS wavelength range is welcome.

2.6.4 Duplicity flag

- **Duplicity flag**. It tells if the star is multiple and provides the level of significance of the detected duplicity.

2.6.5 Science alerts

- OA deals with objects that were not recognized as known astronomical source types. Information on science alerts from other CUs is thus mandatory.

2.7 Data requested from CU7

2.7.1 Variability

- The **variability class identifier** is required as well as a membership probability.
- The **G magnitude** time series
- **Outliers of the variability classification** are needed by the OA package.