Experimental Astronomy manuscript No.
(will be inserted by the editor)

A Parameter Data Base for Large Scientific Projects: Application to the Gaia Space Astrometry Mission

M.A.C. Perryman · J.H.J. de Bruijne · U. Lammers

Received: date / Accepted: date

Abstract The parallel development of many aspects of a complex space science mission like Gaia, which includes numerous participants in ESA, industrial companies, and a large and active scientific collaboration throughout Europe, makes keeping track of the many design changes, instrument and operational parameters, and numerical values for the data analysis and simulations, a challenging but crucially important problem. A comprehensive, easily-accessible, up-to-date, and definitive compilation of a large range of numerical quantities is required, and the Gaia parameter database has been established to satisfy these needs.

The database is a centralised repository containing, besides mathematical, physical, and astronomical constants, many satellite and subsystem design parameters. Version control provides both a 'live' version with the most recent parameters, as well as previous 'reference' versions of the full database contents. Query results are formated by default in HTML, while an important feature is that data can also be retrieved as Java, ANSI-C, C++, or XML structures for direct inclusion into software codes, such that all collaborating scientists can use the retrieved database parameters and values directly linked to computational routines.

Keywords Gaia · Reference Values · Database · Software Systems

PACS 95.10.Jk · 95.55.Br · 95.75.Pq · 95.80.+p

Mathematics Subject Classification (2000) 85-04

M.A.C. Perryman & J.H.J. de Bruijne
Research and Scientific Support Department, ESA–ESTEC
Postbus 299, 2200 AG Noordwijk, The Netherlands
E-mail: jdbruijn@rssd.esa.int

U. Lammers
Research and Scientific Support Department, ESA–ESAC
P.O. Box 50727, 28080 Madrid, Spain

1 Introduction

An important issue for large, complex, scientific collaborations is ensuring that all participants use the same relevant set of numerical parameters. In the case of the development of ESA's Gaia space astrometric satellite and associated data analysis, several hundred scientists in dozens of institutes in Europe, and across some 15 scientific 'working groups', make use of many hundreds of numerical parameters which originate from the satellite industrial teams or operations centre, or are based upon numerous other physical or astronomical constants. Instrument parameters evolve as the satellite design develops up until launch, with some (e.g. calibration) parameters being updated thereafter. Relevant quantities, ranging from the prescription of the optics, focal plane parameters, satellite spin rate, etc., may be scattered amongst thousands of industrial documents.

In a complex software system comprising many hundreds of thousands of lines of code, developed and updated across many individuals and institutes over more than a decade, the potential for errors due to conflicts in parameter usage stem from many trivial and non-trivial causes, including: (a) errors in representing or approximating basic mathematical constants; (b) errors, including truncation errors, in computing derived quantities; (c) failing to update the relevant sections of numerical code to reflect the updated hardware design, for various reasons (e.g. being unaware of updates, or overlooking or incorrectly implementing them); (d) using different values of basic physical or astronomical constants whose very definition (e.g. the astronomical unit) or assumed numerical value (e.g. the constant of gravitation) evolves with time; (e) conflicts created by different choices adopted by different expert groups, e.g. a differences between the state-of-the-art determination of the mass of Solar System bodies and those implicit in even the latest self-consistent Solar System ephemerides; (f) other types of error related to how software is developed and updated over intervals of many years. Software versioning *per se* does not solve the issue of what numerical values lie buried within such complex systems.

In practice, these types of error can be almost impossible to detect, and a robust method developed from the start of the software development would go a long way in eliminating them, and their effects, in the final products of the data analysis.

2 The Gaia Parameter Database

The parameter database for Gaia was designed and developed to address the two central issues of this problem: providing a centralised repository of a wide range of Gaia-related parameters, and providing a convenient and controlled access to the up-to-date numerical values. It was released in 2004 after an extensive contents review by the Gaia Science Team and various associated working groups. At the end of 2007, between 1000–2000 parameters are included, the number depending on how certain multi-dimensional parameter sets are defined. It is now actively used in various environments, and mandatory for certain core elements, notably in the telemetry software, in the simulation software, in the global iterative solution software, and in on-board detection software simulator. Major simulation runs, and the associated global iterative solutions, can also be identified with specific reference versions of the parameter database. Its consistent use by the Gaia community is helping to ensure that correct numerical values are being used throughout the complex software system being built up.

We describe the principles of the resulting Gaia parameter database, believing that the underlying procedures are not generally applied for other comparably complex scientific collaborations, and that the rationale and adopted solution should be of interest and value to other such efforts.

Content Part of the database comprises all relevant mathematical, physical, and astronomical constants. Examples include fundamental mathematical constants (π, e); fundamental physical constants from the 2006 revision of CODATA[1] (speed of light, Newton's constant of gravitation, Planck constant); Solar system ephemeris data (astronomical unit, planetary masses); astronomical constants (obliquity of the ecliptic, light year, 4π sr in square degrees, precession constants, etc.); Earth and Moon parameters (equatorial radius, dynamical flattening, mass, angular rotation, etc.); properties of the Sun (spectrum, mass, apparent radius, radius, quadrupole moment, absolute magnitude, bolometric correction, surface gravity, luminosity, etc.); planets and major moons (radii, masses, mean densities, geometric albedos, orbital semi-major axes, periods, eccentricities, and inclinations, post-Newtonian light deflection, etc.); asteroid data; representative star spectra and filter transmission curves; properties of the local standard of rest (Oort constants, Galactocentric distance, etc.).

The database also contains a large number of satellite and subsystem design parameters relevant for the data analysis on ground, and for simulations of the mission, accuracy budgets, and telemetry data stream. Examples include the L2 orbital description, mission lifetime assumptions; system dead times; scanning-law parameters, pointing and scan-rate requirements; size and layout of the various fields of view; focal plane properties (image scale, operating temperature, star transit speeds, transit times, processing rates, etc.); telescope properties (number of mirrors, mirror reflectivity, pupil area, focal length, coefficients of thermal expansion, wave-front error maps); detector properties (pixel size and angular area, dead zones, integration period, binning characteristics, read-out noise characteristics, quantum efficiency and modulation transfer function of the detectors as a function of wavelength).

Naming Scheme All database entries have unique and self-explanatory names, unambiguous descriptions, and literature references. If appropriate, they carry conventional (LaTeX-format) symbol (e.g., 'ω' for the satellite spin rate), the

[1] `http://www.codata.org`

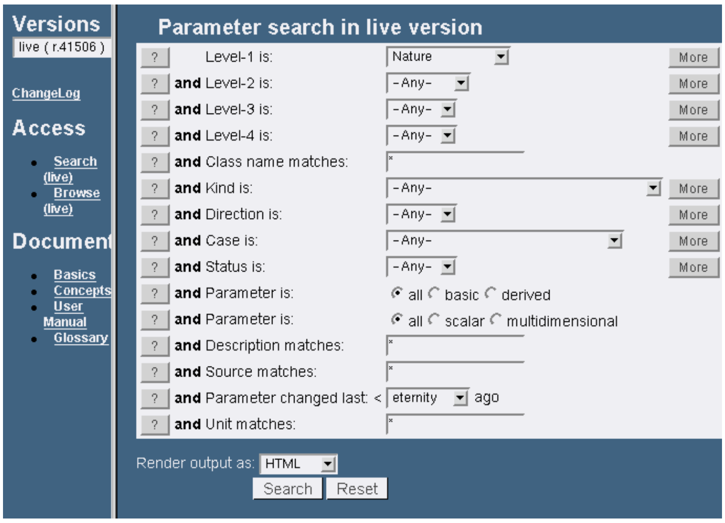

Fig. 1 The default Gaia parameter database search form, as it appears on the main parameter database entry page. Activating the 'Search' button will return a form like the extract displayed in Figure 3.

date of last modification of the parameter, and a status flag indicating whether the numerical value is confirmed or pending confirmation. SI units, with appropriate prefixes, are consistently used throughout. In cases where a parameter is frequently expressed in more than one 'logical unit' (for example, the astronomical unit in units of light-seconds or meters, the Julian year in units of Julian days or SI seconds) then units are also included in the parameter names (thus `AstronomicalUnit_Second` and `AstronomicalUnit_Meter`).

Reflecting the specific design of the satellite and payload, and with a view to minimising parameter redundancy, the Gaia parameter database employs a top-down four-level hierarchical parameter naming scheme. This simply reflects the instrumental system symmetry and commonality for the specific experiment for which it was constructed (for example, the nominal read-out noise for all CCDs in the astrometric field are identical, and entered only once in the database). Parameters related to mathematical or astronomical constants are essentially all at the first such naming level.

Data Categories All parameters in the database are categorised as either 'basic' or 'derived'. Basic parameters are set explicitly in the database as hard-coded [numerical] values. Derived parameter entries are calculated on-the-fly from other parameters, which may themselves be either basic or derived. For all derived parameters, the relevant dependencies are provided as formulae in an 'expression' field. All dependencies are expanded and processed, in double-precision arithmetic, with each new query: all dependencies are thus truly transparent.

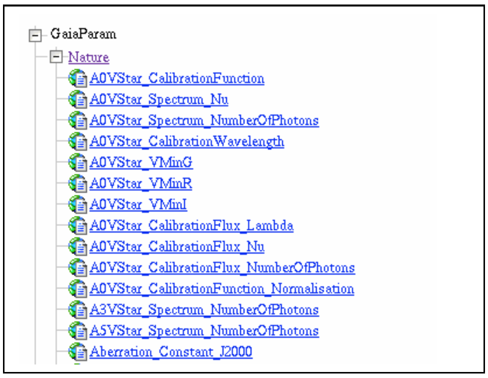

Fig. 2 Example of an intermediate step of a browse operation. The 'Nature' hierarchical node has been expanded, exposing all associated sub-levels (none in this case) and all individual parameters available at this hierarchical level. A mouse click on any of these parameters will return a result window like that displayed in Figure 3.

As an example, the effective temperature of the Sun is not entered as a number, but is calculated from the Sun's radius and luminosity. In turn, the Sun's radius is derived from its measured angular diameter and the value of the astronomical unit; the latter is in this case taken from the numerical planetary ephemerides INPOP06 developed at the IMCCE, Obs. de Paris [1]. The Sun's luminosity is also a derived parameter, derived from the astronomical unit and the solar constant. This example illustrates that a request to change the numerical value of the Sun's effective temperature (for example), cannot be implemented directly.

Parameters are also categorised as either 'scalar' or 'multi-dimensional'. Scalar data comprise integer-type, double precision-type, and string-type entries. Multi-dimensional data (such as quantum efficiency versus wavelength, wavefront error maps versus coordinate location, reference stellar spectra, and 3×3 coordinate transformation matrices) are currently stored in FITS format, although other output formats, such as XDF, could be provided.

Access Modes The database can be searched or browsed on-line using a regular Web browser.

The basic 'search form' (Figure 1) represents the four-level hierarchical structure of the database, and contains several selection options (e.g., free-text search in the 'description' and/or 'source' fields, selection of all 'basic parameters', selection of all 'scalar parameters', etc.). Multi-dimensional data files can either be downloaded or displayed on the screen. By default, queries return up-to-date parameter values from the 'live' version of the database, although provision is made for retrieval of previous reference versions of the database. A change log tracks major modifications, such as the suppression of parameters, name or value changes, etc.

The 'browse' mode has no search options, and instead functions by means of clickable, expandable nodes which represent the database hierarchy (Figure 2). Individual parameters are returned as for the search mode.

An off-line access mode is also available, enabling users to download (selected parts of) the contents of the database either manually from the command line or automatically, possibly at regular intervals, by means of a script.

:Nature:INPOP06:AstronomicalUnit_Second	4.990047838061357e+2	s	false	true	TBC	Astronomical unit (AU) light time (TCB−compatible value in SI units; INPOP06 value)
:Nature:INPOP06:AstronomicalUnit_Meter	1.495978706910000e+11	m	true	true	TBC	Astronomical unit (AU) length (TCB−compatible value in SI units; INPOP06 value)
:Nature:INPOP06:Earth_GM	3.98600439077e+14	m^3 s^−2	false	true	CONF	Geocentric gravitational constant (TCB−compatible value in SI units; INPOP06 value)
:Nature:INPOP06:Earth_EquatorialRadius	6378137	m	true	true	CONF	Equatorial radius of the Earth (INPOP06 value)
:Nature:INPOP06:Earth_JSub2Dot	−3.0e−9	cy^−1	true	true	CONF	Secular (long−term) variation of the dynamical form−factor J_2 of the Earth (also known as oblateness and as Stokes' second−degree zonal harmonic of the geopotential) due to the post−glacial rebound of the mantle (INPOP06 value)

Fig. 3 Example of a query result (obtained after pressing the 'Search' button in Figure 1) containing five parameters (entries). The various columns are: parameter name, value, unit, basic/derived flag, scalar flag, confirmation status, and description/reference. The same form, but containing only one parameter, also results from each browse operation.

Output formats Query outputs are formated by default in HTML, convenient for visual inspection (Figure 3). Other 'readable' output formats, notably LaTeX and PDF, are also supported. A particularly powerful feature of the database design is that data can also be retrieved as Java, ANSI-C, C++, or XML structures for direct inclusion into software codes in these languages (e.g. Figure 4). A Comma-Separated Value (CSV) output format also caters for Microsoft Excel spreadsheets, while options for Fortran-77 and Fortran-90 have been discontinued due to their discouraged use in the Gaia project, in part due to their restriction in handling complex structures.

3 Technical Implementation

The core of the Gaia parameter database is a central, relational, SQL-based database system which holds the parameters and all associated attributes in a collection of logically linked tables. The open-source relational database MySQL[2] has been chosen as baseline. The on-line user is not directly exposed to the SQL-level, but interacts with the system through either a dynamically generated query form or a tree-like browse interface. Both of these interfaces are realised through a central software layer implemented in the PHP[3] scripting language. The software serves three main purposes: (i) dynamically creating the user interfaces (query form and tree-view); (ii) generating SQL-queries to the database system from the contents of the query form; and (iii) receiving matching result sets from the database and rendering the results into the desired output format.

The result rendering is a two-stage process. First, the query results from the database are converted on the fly into an in-memory XML[4] structure by means of PHP's DOM facilities. From there, the rendering into the final output

[2] http://www.mysql.com
[3] http://www.php.net
[4] http://www.w3.org/XML

```
static const char *const DBVersion = "-live-:2008-03-03T14:03:52";
class Nature {
public:
    class INPOP06 {
    public:
        // Geocentric gravitational constant (TCB-compatible value in SI units)
        // Source: A. Fienga, J. Laskar, H. Manche, M. Gastineau, 19 April 2007
                'Solar System Planetary Ephemeris Delivery: INPOP06'
        // Status: CONF
        // Basic : false
        // Scalar: true
        static const double EARTH_GM = 3.98600439077e+14; // [m^3 s^-2]
    };
};
```

Fig. 4 Example of part of the file header, and a single database parameter, given by a C++ search request. The code includes the unique parameter name (here `EARTH_GM`), numerical value and unit, reference source, and associated flags, here indicating that the value has been confirmed, that it is a derived parameter, and that it is a scalar quantity.

formats is driven by a collection of XSLT (eXtensible Stylesheet Language Transformation) stylesheets on the server side. There is exactly one stylesheet per supported output format, one of which is XML itself.

Using XML as intermediate data format is a flexible approach allowing the seamless conversion to almost any other format through XSLT. All XSLT stylesheets that control the conversion to the supported output formats are publicly accessible in a directory under the main URL. They may be used as templates for special client-side XSLT rendering of the XML output.

Version control is implemented through the widely-used Concurrent Versions Systems (CVS). The user is not exposed to the raw CVS level but the central management software provides a simple interface for the selection of all available content versions and their transparent retrieval.

One of the authors (JHJdB) has been responsible for defining the detailed database entries, and is the single-point contact for maintaining and updating the data content. Another (UW) has implemented and maintains the detailed data retrieval and output structures.

Although access to the Gaia parameter database is limited to members of the Gaia scientific and industrial teams, a demonstration version, with full functionality but without the detailed database content (i.e. excluding most of the Gaia instrument-specific parameters, and including only a few of the physical or astronomical entries by way of illustration) is accessible via the ESA Gaia www pages[5]. TBC

Acknowledgements We thank the many members of the Gaia scientific community who provided feedback on the database, and also validated the current entries through the various scientific working groups. We thank Ivo Busko (STScI) for providing the dedicated Java applet for downloading and displaying the multi-dimensional data files.

[5] `http://www.rssd.esa.int/gaia`

References

[1] Fienga A, Manche H, Laskar J, Gastineau M (2008) INPOP06: a new numerical planetary ephemeris. Astron. Astrophys. 477:315–327