All of the Sky:

HEALPix density maps of Gaia-scale datasets from the database to the desktop

Mark Taylor[1], Grégory Mantelet[2], Markus Demleitner[2]

Introduction

The Gaia catalogue provides access to observations of around a billion sky sources. The primary access to this archive is via TAP services such as GACS and ARI-Gaia, which allow execution of SQL-like queries against a large remote database returning a relatively small result set for client-side use. Such services are generally used for extracting relatively small source lists according to potentially complex selection criteria. But they can also be used to obtain statistical information about all, or a large fraction of, the observed sources by building histogram-like results.

We examine here the practicalities of producing and consuming all-sky HEALPix weighted or unweighted density maps in this way for Gaia and other large datasets. We present some modest requirements on TAP/RDBMS services to enable such queries, and discuss with examples visualisation and serialization options for the results, including some new capabilities in recent versions of TOPCAT.

Tiling scheme

Constructing a sky density map requires a decomposition of the sky into tessellating tiles. Various tiling schemes are available, including HTM, Q3C and HEALPix (in RING and NESTED flavours). We use here NESTED HEALPix, because unlike those other schemes:

- pixels are equal area, making density map analysis easier
- simple SQL-friendly arithmetic degrades pixel index to lower resolution

The HEALPix grid at order N defines tiles with indices in the range $[0, 12 \times 4^N)$. A sky position within tile i at order N falls within tile $i/4^{N-M}$ at a lower order (coarser resolution) M.

Is it useful?

An all-sky or wide-field view of quantities aggregated from a large catalogue can sometimes reveal large scale features or trends in astronomical or instrumental behaviour that would be difficult to discern from other data products. Source density maps are the most obvious application, but there are numerous other possibilities.

Some data centers (including the ESA and ARI Gaia archives) offer for download various all-sky maps in graphical or tabular form that have been pre-calculated on the server, perhaps by different or more efficient methods than that described here.

But sometimes it is useful for end users to be able to assemble density maps themselves, for instance applying custom source selections or weighting functions not foreseen by data centers. Examples of some such custom queries are given in the figures.

TOPCAT HEALPix functions

Recent releases of the TOPCAT/STILTS table analysis suite include new features for working with HEALPix maps.

Tables with an implicit or explicit HEALPix index column can be visualised interactively or exported to bitmapped or vector graphics files. They can be displayed within TOPCAT's Sky Plot window which offers interactive adjustment of colour maps and grid resolution, pan/zoom navigation, a choice of sky projections and coordinate systems, and the option to overlay multiple plots of different types. The figures on this poster were produced using STILTS.

There are also new capabilities to generate HEALPix maps on the client side from local source catalogues and a number of HEALPix-related functions added to the expression language.

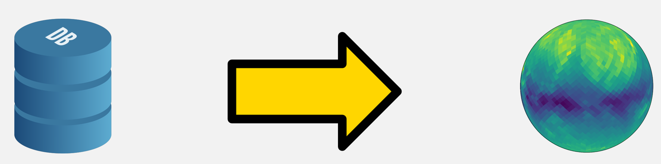
Since HEALPix maps are tables, TOPCAT and STILTS can be used to analyse and manipulate them in general, non-visual ways too, for instance calculating statistics and performing joins.

Conclusions

It is sometimes useful for astronomers to make and use all-sky density maps from large datasets. This is feasible given TAP access, and we present here some discussion and examples. The technique is not novel, but the lack of the features it requires in most existing TAP services suggests that it is not widely practised.

To enable more widespread use of this technique, we recommend that TAP services should make available the User-Defined Function `ivo_healpix_index(order, ra, dec)`, and should also consider the case of sky map creation when setting query timeout and row output limits. We also encourage the IVOA to standardise the representation of HEALPix tile indices in VOTables.

Database → Sky Map Requirements



This panel explains what has to be in place for end-users to be able to make customised weighted or unweighted all-sky density maps for catalogues that would be impractical to download.

1. SQL-like access to source catalogue

SQL-like queries to a relational database are required.

⇒ Public datasets are increasingly exposed via TAP (Table Access Protocol), allowing remote execution of ADQL (SQL-like) queries.

2. HEALPix column or function

Either the table must have a column giving the index of the HEALPix tile in which the source position falls, *or* a User-Defined Function must exist that can calculate tile index for each row (e.g. from RA, Dec columns).

⇒ Most existing TAP services don't currently provide this, but the ARI-Gaia and DaCHS TAP services have introduced such a UDF *(this work)*:

```
ivo_healpix_index(order, ra, dec)
```

An order-12 HEALPix index is also buried in bits 36–63 of the Gaia `source_id` column.

3. GROUP BY query

An SQL query of the form

```
SELECT (agg-func) FROM (table) GROUP BY (healpix-index)
```

calculates the sky map, returning one row per populated sky pixel. The aggregate function defines the weighting (`COUNT(*)` yields unweighted source density) and a `WHERE` clause can optionally be added to restrict the selection of sources.

⇒ The aggregate functions `COUNT`, `SUM`, `AVG`, `MIN`, `MAX` are standard in ADQL. Unfortunately median and other robust estimators are not, since they can be much more expensive to calculate.

4. Query limits

Limits on query execution time and output size must accommodate execution of these aggregating queries.

⇒ These queries usually involve a full-table scan, and runtimes are typically on the order of, very roughly, an hour per billion rows. This is fairly, though not unfeasibly, long, but some TAP services impose execution time limits that may preclude their use (e.g. GACS queries are limited to 30 minutes, so these aggregations usually time out).

Output row limits can also be an issue; density maps suitable for visualisation may be of the order of a million rows (HEALPix $N = 8$ has 786 432 tiles), though finer or coarser resolutions can also be useful. Million-row results are permitted by many, but not all, TAP services.

5. Semantic markup of HEALPix output

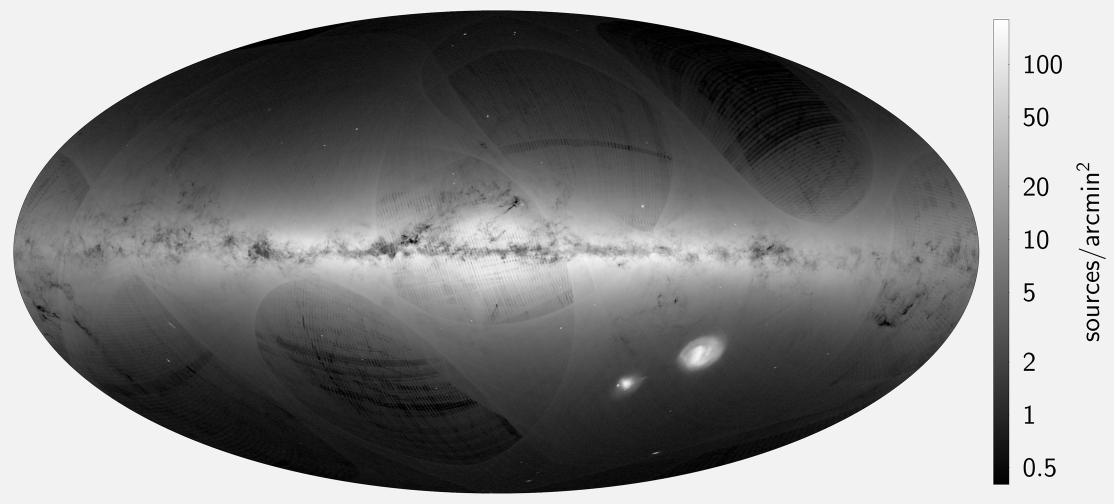
Ideally, the TAP service should mark up output columns containing HEALPix indices describing their content, to assist downstream visualisation/analysis software.

⇒ An undocumented convention exists for encoding HEALPix maps in FITS files. The best way to mark up HEALPix columns in VOTable (the standard TAP output format) is currently under discussion in the IVOA.

6. Analysis tools for HEALPix maps

Having calculated a HEALPix sky map you may want to visualise it, compare it with other maps etc.

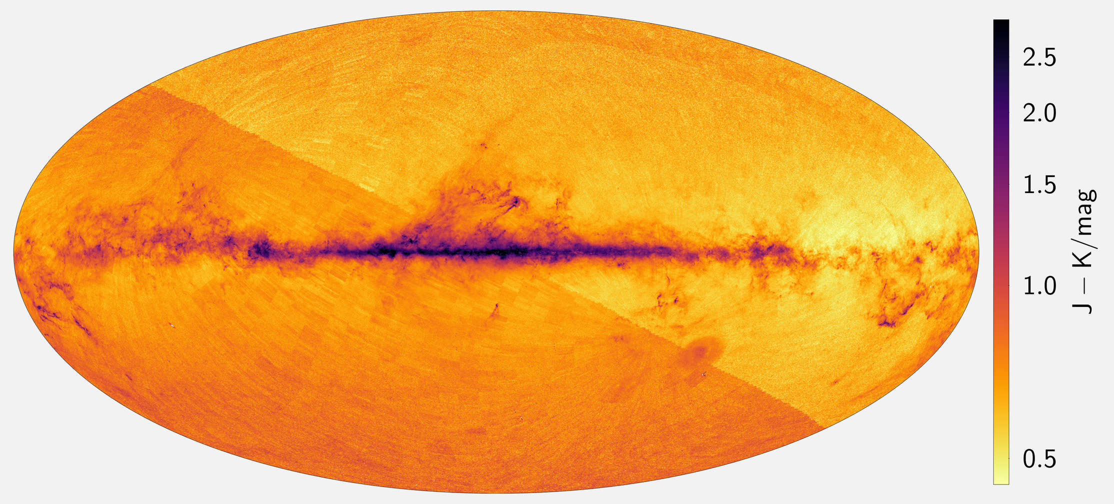
⇒ TOPCAT and STILTS can manipulate, and in recent releases *(this work)* visualise, HEALPix maps. Other options for HEALPix visualisation are also available (Aladin, Healpy, ...)




```
SELECT gaia_healpix_index(8, source_id) AS hpx8,
       COUNT(*) * (POWER(4,8)*12.0/(41253.*3600)) AS srcdens
FROM gaiadr1.gaia_source
GROUP BY hpx8
```

Source density for the full Gaia DR1 source catalogue. A Gaia-specific UDF is used to extract HEALPix index from Gaia `source_id` column. The multiplication factor converts from count per order-7 HEALPix tile to density per square arcminute.

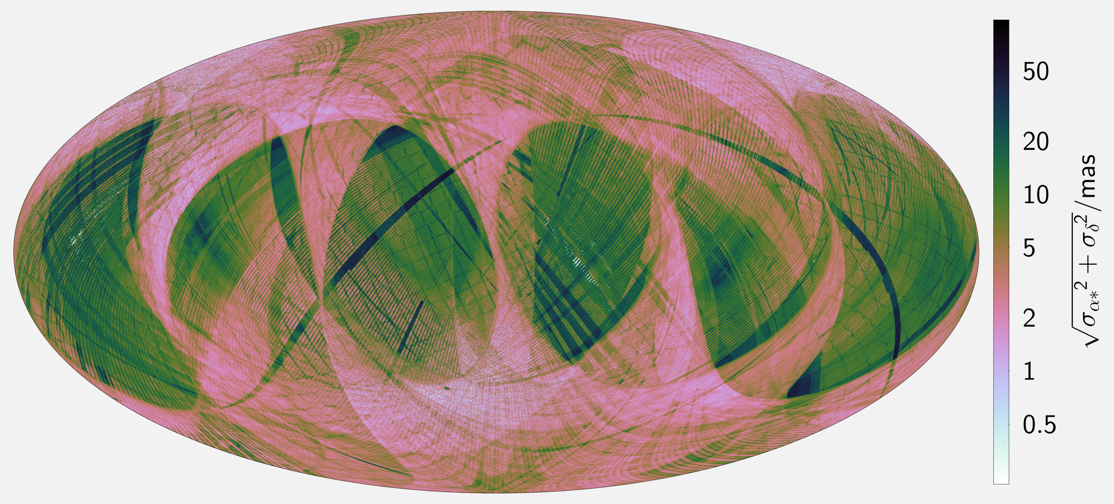
1.1 billion rows, 42 mins runtime on ARI-Gaia TAP service (YMMV).




```
SELECT ivo_healpix_index(9, raj2000, dej2000) AS hpx9,
       COUNT(*) AS nsrc,
       AVG(jmag - kmag) AS j_k
FROM twomass.data
WHERE qflg LIKE 'A_A' AND cflg LIKE '0_0' and xflg = '0'
GROUP BY hpx9
```

$J - K$ colour for 2MASS point sources. The proposed standard UDF is used to calculate HEALPix index from sky position. The upper right half of the image used the `WHERE` clause above, which selects only sources with good J/K photometry, while the lower left includes all sources (no `WHERE` clause). With the custom selection the image is cleaner and the values are lower on average, though not uniformly over the sky.

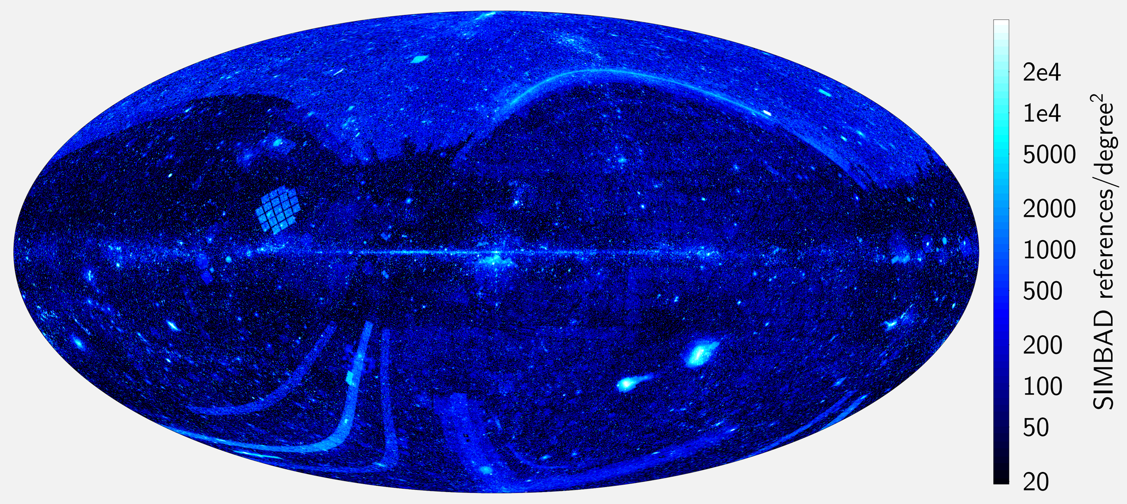
163/471 million rows, 16/39 mins runtime on GAVO DC TAP service.




```
SELECT source_id/2199023255552 AS hpx9,
       AVG(SQRT(ra_error*ra_error +
                dec_error*dec_error)) AS pos_error
FROM gaia.dr1
GROUP BY hpx9
```

Mean isotropic positional error of Gaia DR1 sky positions. The HEALPix index is recovered from the Gaia `source_id` column using integer division. The projection is ecliptic, unlike the other plots.

1.1 billion rows, 70 minutes runtime on GAVO DC TAP service.




```
SELECT hpx/16 AS hpx8, SUM(nbref) AS nbref, COUNT(*) AS nobj
FROM public.basic
WHERE hpx IS NOT NULL
GROUP BY hpx8
```

Sky density of SIMBAD object literature references. The SIMBAD basic table has a column giving order-10 HEALPix index. Conversion to density per degree^2 was done by plotting `nbref/healpixSqdeg(8)`.

8 million rows, 30 seconds runtime on SIMBAD TAP service.

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme (FP7-SPACE-2013-1) for research, technological development and demonstration under grant agreement no. 606740 (GENIUS), the UK's Science and Technology Facilities Council grant ST/M000907/1, and the German Astrophysical Virtual Observatory GAVO, BMBF grant 05A11VH3. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`http://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `http://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. It also makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and IPAC/CalTech, funded by NASA and the NSF, and of the SIMBAD database, operated at CDS, Strasbourg, France.

[1] H. H. Wills Physics Laboratory, University of Bristol, U.K.
[2] Astronomisches Rechen-Intitut, Zentrum für Astronomie der Universität Heidelberg, Germany