A PROPOSAL FOR THE BULK-DOWNLOAD REPOSITORY FOR GAIA DR4

Prepared by	Jos de Bruijne
	Directorate of Science

Document Type	TN - Technical Note
Reference	GAIA-CZ-TN-ESA-JDB-128
Issue/Revision	1 . 1
Date of Issue	21/10/2024
Status	Approved

APPROVAL

Title	A proposal for the bulk-download repository for Gaia DR4		
Issue Number	1	Revision Number	1
Author	Jos de Bruijne with key support from individuals as defined in the abstract	Date	21/10/2024
Approved By	Gaia Mission Manager and Gaia Project Scientist	Date of Approval	21/10/2024

CHANGE LOG

Reason for change	**Issue Nr**	**Revision Number**	**Date**
Initial draft release	D	1	23/09/2024
Incorporated feedback from Enrique Utrilla Molina	D	2	26/09/2024
Incorporated feedback from Héctor Cánovas	D	3	27/09/2024
Incorporated feedback from David Teyssier, Gonzalo Gracia, Mark Taylor, Nigel Hambly, Jordi Portell, Xavier Luri, and Jorgo Bakker	1	0	07/10/2024
Incorporated feedback from Uwe Lammers and Johannes Sahlmann	1	1	21/10/2024

CHANGE RECORD

Issue Number 1	Revision Number 1		
Reason for change	Date	Pages	Paragraph(s)

DISTRIBUTION

Name/Organisational Unit
ESA Gaia mission, Gaia Science Team (GST), ESA Euclid mission, ESAC Science Data Centre (ESDC), ESA SCI-S programmatic IT, ESA Datalabs, Gaia DPAC (including partner and affiliated data centres), SPACIOUS, Vera Rubin HiPSCat team, IVOA

Table of Contents

ABSTRACT

We present a science-driven proposal, for wide stakeholder feedback, for the contents, data format, metadata, partitioning scheme, and delivery scheme of the Gaia DR4 bulk-download file repository.

The proposal is to change the Gaia SOC/DPAC DPCE-internal data format for Gaia DR4 from Gaia binary (gbin) to Parquet, and also to deliver files with this format to the Gaia partner and affiliated data centres. Furthermore, we propose to change the Gaia DR4 bulk-download data format from enhanced CSV (ECSV) to Parquet. The bulk-download repository is proposed to be partitioned using (a to-be proposed evolution of) the HiPSCat scheme, which is an emergent IVOA standard developed by the Vera C. Rubin Observatory team. The bulk-download file repository is also proposed to be encapsulated into an Apache Iceberg table (one table per product) so that industry-standard analytics tools can benefit from a standardised set of big-data "performance metadata", with added VOTable-style "scientific metadata" in XML format for Virtual Observatory (VO) interoperability at individual Parquet-file level. The Gaia DR4 file repository (Apache Iceberg table) that will be NFS-mounted in ESA Datalabs is proposed to contain all Gaia DR4 data products (566 TB[1]) to allow scientists conducting large-scale data discovery and exploitation use cases. The above proposals combined allow to drastically reduce the number of copies of the Gaia DR4 dataset at ESAC, in principle only requiring one master set (and one backup thereof).

As for the delivery scheme of the bulk-download data for Gaia DR4, the proposal is to use the Content Delivery Network (CDN) for the first 6 months after the release, after which a cost-usage-benefit-risk assessment should be made whether the CDN could be switched off such that all bulk-download requests will fetch all data from the "geafiles" master node at ESAC with an associated (moderate to small) risk to occupy a significant fraction of the ESAC internet bandwidth for sustained periods of time. The public bulk-download file repository is proposed to contain all Gaia DR4 data products except for 6 "large" epoch DataLink products, thereby reducing the bulk-download data volume from 566 TB to 66 TB as well as the associated CDN cost. We do, however, identify a desire to add a bright-source subset for these missing epoch products as a nice-to-have which would add <30 TB to the CDN, bringing it to ~100 TB.

The open points are listed in Section 9.

Disclaimer Much of the intellectual ownership of this proposal belongs to Enrique Utrilla Molina, whose knowledge, competence, pro-activeness, and support translated into many Jira tickets describing ideas, tests, tradeoffs, and assessments that made it nearly trivial to compile this document. This proposal has furthermore greatly benefited from input from (in alphabetical order): Jorgo Bakker, Héctor Cánovas, Nigel Hambly, Pedro García-Lario, Xavier Luri, Gonzalo Gracia, Mark Taylor, David Teyssier, and Jordi Portell.

[1] All data volume estimates in this TN are the "current best estimates" from 24 September 2024, with 10% margin added.

1. INTRODUCTION

The Gaia DPAC consortium delivers all data that are to be published in Gaia binary (gbin) format to its ESAC-based DPCE data processing centre. After conversion, mapping, and filtering at DPCE, the resulting gbin files serve various, sometimes interconnected, purposes:

- Distribution to ESDC for ingestion into the Gaia ESA Archive (GEA, or GACS for Gaia Archive Core System);
- Distribution to Gaia partner and affiliated data centres for public release (of all or part of the data, in original or modified format) after the data has become public;
- Use in the DPAC CU9 Data Analysis Framework (DAF) for DPAC-internal, pre-release validation purposes;
- Conversion to Enhanced Character Separated Value (ECSV) format for population of the bulk-download repository that serves the data "in bulk" through a commercial Content Delivery Network (CDN) that synchronises itself with the master node at ESAC (sometimes referred to as "geafiles");
- Exposure of public Gaia data in ESA Datalabs, for exploitation and data mining purposes, through NFS-mounting of the "geafiles" disk of this bulk-download repository (as defined in the previous bullet).

This TN proposes several changes in various elements in the above. These changes are proposed in view of:

- The radically increased data volume of Gaia DR4 (566 TB, >95% of which are DataLink products) compared to Gaia DR3 (~10 TB). This not only requires efficient data formats and careful partitioning to avoid time-consuming data operations at DPCE, underperformance of the DPAC CU9 DAF, underperformance of ESA Datalabs, etc. but also strengthens the need to minimise the number of copies (and backups thereof) of Gaia DR4 that will be floating around at ESAC. As an example, Gaia DR3 data exist in various incarnations and in various places. The final released data are available in both gbin and ECSV format at DPCE for transfer to the partner and affiliated data centres. The CU9 DAF contains a copy of selected tables in Parquet format. The data are available on the "geafiles" machine in ECSV format as master node of the CDN bulk-download repository and for ("in-house") use in ESA Datalabs. Finally, GACS contains a copy of the data inside its PostgreSQL database. That represents four full copies of the final data in three different formats. With the proposed move to Parquet for Gaia DR4, one shared master volume with Parquet files could in principle serve all these purposes simultaneously (at least for the DataLink products);
- The unavoidable introduction – being 100% transparent for users – of a massively parallel processing (MPP) distributed database infrastructure for the Gaia ESA Archive at ESDC ("Greenplum") in which the DataLink data themselves are stored in files in a file repository "behind GACS" instead of being ingested

in the single-instance (SI) GACS database that currently serves Gaia DR3 (see Section 4.1). This design, which is driven by GACS performance and GACS database storage considerations, requires a future-proof and efficient file format, ideally supported by the ESDC common code team for use by other archives;

- The emergence of massive (big-data) scientific use cases (see also GAIA-C9-TN-ESAC-EUM-136, "*Exploitation of Gaia data in commercial public clouds*" – DPAC-internal). These use cases, which can be categorised in three domains, are well-known to be poorly catered for through relational databases such as GACS, regardless of whether they are distributed ("Greenplum") or not:
    - When you don't want to download all the data, but when you want something (non-trivial) done to those data, e.g., group epoch data by SourceId, compare epoch data to a variability or orbit model, find the residuals, etc. (e.g., "*Give me all sources for which the astrometric residuals contain periodic features with periods between 100 and 300 days*").
    - When you want to go through all the data (for instance spectra), compare them to a reference (spectrum), compute some ("similarity") metric, and only return the instances where the metric exceeds a certain threshold (e.g., "*Give me all RVS mean spectra where the flux over the interval 850 – 852 nm exceeds the continuum level by 5% or more*").
    - When you want to do large-scale joins with other massive data sets such as LSST for multi-wavelength and/or time-domain studies (e.g., "*Give me light curves for all Gaia quasars from Gaia, ZTF, and LSST*").

  The state-of-the-art in these cases is using (industry-standard) analytics tools operating on files, allowing distributed (cloud or science-platform) processing through partitioning of the data.

2. GAIA DR3 BULK DOWNLOAD SET-UP

As explained in the ReadMe file, "*the Gaia DR3 bulk download repository allows downloading of the Gaia data for offline analyses in which the full content of one or more tables in the Gaia Archive is needed*". All tables can be found at http://cdn.gea.esac.esa.int/Gaia/ "*as [gzip-compressed] enhanced csv [ECSV] files. These files are regular csv files with an extra header with metadata added on top in which each line is preceded by the hashtag comment character.*" The addition of metadata using enhanced CSV files, importantly making the data self-descriptive and self-contained, was only implemented in Gaia DR3 (the preceding Gaia EDR3 and DR2 releases use regular CSV, while Gaia DR1 offers regular CSV, as well as FITS and VOTable).

Besides the legacy and data-preservation aspects, the scientific use cases of the bulk-download repository are threefold:

(i) Allow download of full tables (or even the full data set);

(ii) Allow download of subsets of a table based on sky coordinates;

(iii) Support efficient (local, science platform, or cloud-based) large-scale analytics (including distributed joins and cross-matching with other large catalogues).

The Gaia DR3 bulk-download repository only supports the first two use cases:

(i) Downloading a full table (e.g., GaiaSource from Gaia DR3) is as easy as typing in a terminal: *wget --recursive --no-parent 'http://cdn.gea.esac.esa.int/Gaia/gdr3/gaia_source/'*

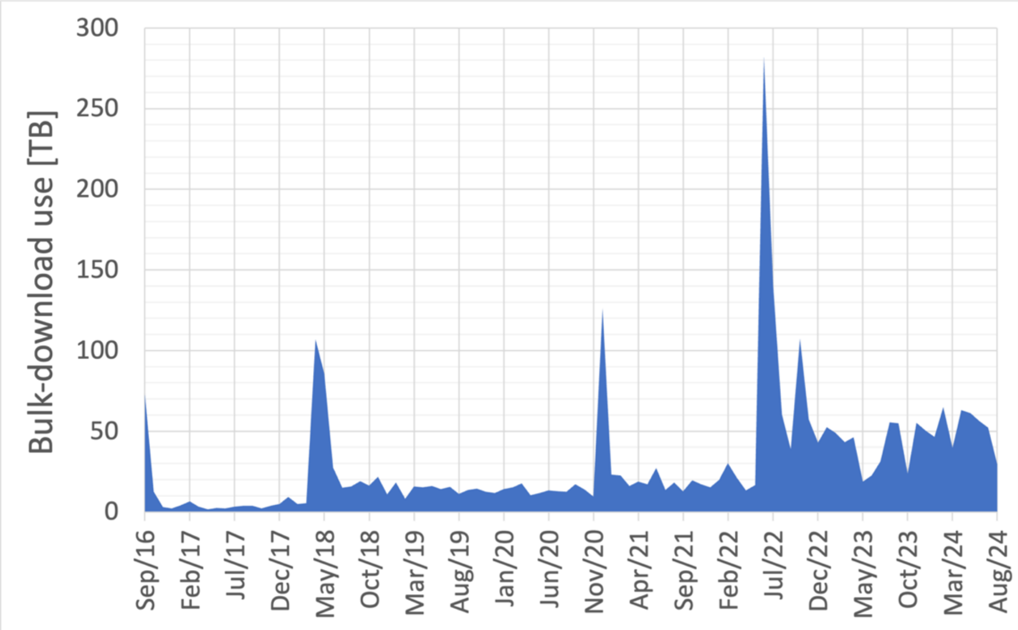
(ii) The content of GaiaSource in Gaia DR3 is partitioned in multiple files based on 3386 contiguous ranges of HEALPix level-8[2] indexes so that, by design, each file weighs ~230 MB and contains ~550,000 sources that are all located in a well-defined area on the sky, the size of which depends on the local object density. Since each file name contains the first and the last HEALPix level-8 indexes that are included in the file, it is straightforward to do efficient "cone searches" in the bulk-download repository using, for instance, the Astropy HEALPix package (see this Gaia Jupyter tutorial notebook for an example). With only a few exceptions, the same tessellation has been used for all other data products in Gaia DR3. The drawback of this approach is that it results both in datasets with many files with few entries per file (e.g., the variability tables) and in datasets with very large files (e.g., the MCMC samples and XP continuous mean spectra). In addition, it results in a skewed file-size distribution within a table, for instance the smallest XpContinuousMeanSpectrum file is 236 MB and the largest one 1.7 GB in size. This is problematic for parallel processing in memory-limited environments such as ESA Datalabs (or a potential future Gaia science platform). The current proposal is meant to remedy these drawbacks (among other things).

(iii) Large-scale use cases are not efficiently supported by the (CSV, row-based and file-size-skewed) Gaia DR3 bulk-download repository, and the current proposal is put forward to remedy this (among other things).

[2] Level 8 was selected for Gaia DR3 as a balance between a manageable number of tiles (786,432) allowing a relatively compact representation of the intervals in the filenames, and a sufficiently high resolution to generate detailed spatial plots. Any other HEALPix level could be selected instead as reference for Gaia DR4, but in order to keep files small enough without subpartitions that would likely imply going down to level 12 (201,326,592 tiles), which in less populated areas would leave a lot of empty or almost empty files.

3. GAIA DR3 BULK DOWNLOAD USAGE

The Gaia bulk-download repository, that has been delivering data to users through a CDN since Gaia DR1 in September 2016, continues to be used intensively with the following monthly download-volume profile:



As of Gaia DR3 (released in June 2022), the monthly volume of data downloaded is of order ~50 TB. The ~3-fold increase of this volume between Gaia DR2 (released in April 2018) and Gaia DR3 is due to users downloading DataLink products in full (RVS mean spectra, BP/RP sampled spectra, BP/RP continuous spectra, etc.). From the above time evolution, there is no evidence for any decreasing trend after a given data release has taken place.

4. PROPOSED DATA FORMAT FOR GAIA DR4

The Gaia DR3 bulk download repository serves data in Enhanced Character Separated Value (ECSV) format. The CSV decision was made before Gaia DR1, around a decade ago, and most likely involved arguments of being human readable, platform independent (Windows, Mac, Linux), and technology-wise "legacy proof" as in Egyptian clay tablets surviving for millennia (albeit the use of gzipped files actually kills this argument). Although the ECSV format is self-descriptive through the addition of metadata in the header, it is not Virtual Observatory (VO) interoperable (while noting that TOPCAT can actually deal with this format). More importantly, for Gaia DR4, the element of human readability is considered to be inferior in weight to the so-far not-considered element of "big-data performance" (in cloud, ESA Datalabs, other science platform, or local applications), which is extremely poor for the row-based ECSV format.

In that light, the proposal presented here, formally in response to C9OPS-837 "*Define format for DR4 bulk download*", is to move from ECSV to Apache Parquet format. Parquet is an open-source, column-oriented data file format designed for efficient (binary) data storage and retrieval of large tables with many columns. It provides high-performance (per-column) compression and encoding schemes to handle complex data in bulk, and is supported in many open-source and/or industry-standard tools allowing parallel processing. Within the Vera C. Rubin Observatory (LSST, based on collaborative work involving STScI and IPAC, also in view of the upcoming Nancy Grace Roman and SPHEREx missions), the baseline is to release data in Parquet format. The Gaia DataMining Platform in Edinburgh is also based on Parquet files, as is the Gaia Data Analytics Framework (GDAF) platform concept at the University of Barcelona and the CU9 DAF at ESAC. In addition, the NASA astrophysics archives have been working on publishing multi-mission catalogue data in Parquet format in the cloud while Parquet is also considered the default format to be used for large catalogues in the NASA Fornax Astrophysics Science Platform Initiative. As a result, the IVOA has recognised that "*Parquet is the new CSV*" (quote from Gregory Dubois-Felsmann) and has included associated developments in its roadmap. Gaia now joining this initiative, as per this proposal, will certainly provide sufficient weight to make this ongoing (and already non-stoppable) development a lasting success.

4.1. ESDC proof of concept

In preparation of a potential move from gbin to Parquet format, ESDC has done an exhaustive trade-off and proof of concept exercise reported here [ESA-internal] (see also C9OPS-1295 "*Feasibility study: storing Gaia DR4 DataLink products in Parquet format*"). After successful conclusion of this proof of concept in May 2024, the associated software developments to ingest Parquet instead of gbin files and to let GACS retrieve DataLink products from a Parquet file repository rather than a database table, have been started following this roadmap [ESA-internal]. Currently (October 2024), an investigation is ongoing to let GACS fetch the DataLink products from a Parquet-based Apache Iceberg table (see Section 5.2) created by DPCE (e.g., using the Iceberg Metadata or Iceberg Scanning look-up APIs), which would remove the ESDC / GACS dependency on Gaia-specific custom (YAML) metadata files as well as eliminate the need for a dedicated ESDC copy of the DataLink products (545 TB for Gaia DR4), although this could also be achieved independently of such an Iceberg / datalake solution.

4.2. Both Parquet & ECSV

One option is to provide a small subset of Gaia DR4 products in the bulk-download repository in Parquet as well as ECSV format. The most innocent hybrid proposal in this direction would be to serve GaiaSource in (gzipped) ECSV as well as Parquet format. This could serve "senior astronomers", also somehow recognising the inertia and "laziness" of the community at large, and would provide a trusted reference version of this key file, avoiding a potential non-controlled proliferation of informal conversions appearing around the globe. However, although this would only add a "negligible" ~1 TB of data to the total repository, there are no legitimate scientific use cases that would benefit from this approach, that would at most provide emotional satisfaction and illusionary peace of mind of being a true legacy-proof solution. It is therefore proposed, pending an assessment of the popularity of the Gaia DR3 GaiaSource table in the bulk-download repository, to **not** follow this path but to provide a tutorial and Jupyter notebook for interested users to do this conversion themselves (C9DRD-411).

5. PROPOSED METADATA FOR GAIA DR4

As of recent, user-defined key-value pairs, i.e., custom metadata fields, can be defined in the Parquet footer. Broadly speaking, the question of the definition of metadata has two elements:

1. Which VO interoperability metadata shall be included? Where shall these "scientific metadata" be stored? Which format shall be used?
2. Which big-data performance metadata shall be included? Where shall these "operational metadata" be stored? Which format shall be used?

The proposal presented here, formally in response to C9OPS-1331 "*Specification of the dataset metadata format*", addresses both questions.

5.1. Scientific metadata

As realised by the IPAC-IRSA/Rubin team (see this presentation), the optimal way to embed scientific, per-column metadata (name, datatype, description, unit, UCD, and UType) into Parquet files is to use a VOTable-style XML format, benefitting from its extensive standardized set of metadata (recently extended with Model Instances, MIVOT) that is defined around VO interoperability. Active developments (driven by Rubin) are ongoing in this direction (see this presentation), supporting both single Parquet files and partitioned datasets, and embedding VOTable metadata (i.e., a DATA-less VOTable) in Parquet files is intended for inclusion in Astropy 7.0 (courtesy Brigitta Sipőcz) while prototypes also exist in Rust (courtesy François-Xavier Pineau) and

STILS/TOPCAT (courtesy Mark Taylor). Not surprisingly, for Gaia, it is proposed to adopt this route (obviously following any forthcoming IVOA standard that is likely to appear).

As for the location of the scientific metadata, the proposal presented here is to ensure that individual Parquet files in a partitioned data set are self-contained and self-descriptive, i.e., carry their metadata around with them. This has the unavoidable consequence that metadata is duplicated within a dataset, (marginally) increasing the file size and adding some risk for potential inconsistencies. These minor disadvantages are considered to be far inferior to the significant advantage that individual (or small subsets of) Parquet files retrieved from the bulk-download repository remain scientifically interpretable and usable (for instance with TOPCAT).

5.2. Performance metadata

Examples of "performance metadata" are:

- The minimum and maximum value of each primary key, which can be useful to identify in which file a given source is located;
- The number of not-null values for each field, plus its minimum and maximum values, which can be used to detect inconsistencies or problems (e.g., values out of the valid range, fields not being populated, etc.) or create statistics;
- Information on the partitioning (which may be encoded in the folder structure and naming convention, in performance metadata, or in both) and/or on the sorting of data inside each partition.

In short, "performance metadata" support validation but are key for big-data tools and frameworks such as Spark and Dask to operate efficiently on large, partitioned data sets with (tens of) thousands of files.

Since there is no VO recommendation (neither existing nor in the making) and no need for Parquet files to be self-contained regarding performance metadata, we propose to use a standardised approach that is recognised by industry-standard big-data solutions. Following this assessment [DPAC-internal] (see also SCO04TSK-118), the most promising (i.e., simple, general, and open-source) candidate is the Apache Iceberg table format. The main benefit of this approach is that the performance metadata follows an industry standard, facilitating and optimising the processing of the data both in distributed processing engines (e.g., Spark, Trino, ...) and in single-process libraries (that can be used in, e.g., ESA Datalabs through PyIceberg), without custom solutions based on the conventions that were used when generating the data. One caveat exists for this approach: the current Apache Iceberg specification only supports absolute (and not relative) URIs to identify the location of the data.

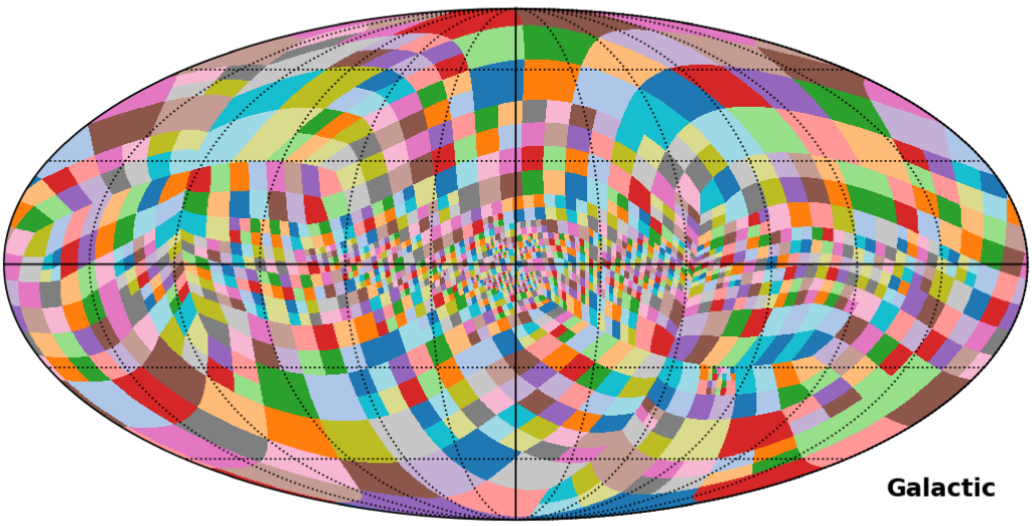
This may be addressed in a future update of the specification while a "home-grown solution" to work around this has already been implemented (see C9UTIL-166) so that this is not considered to be a showstopper.

6. PROPOSED PARTITIONING FOR GAIA DR4

As explained in Section 0, the Gaia DR3 partitioning has several drawbacks. Whereas some of these could be mitigated by refinement of the current scheme (e.g., use of a different level of HEALPix indexes to increase the granularity) or "reformatting" of it (e.g., rather than including the range of HEALPix indexes in the file name, include a simple partition Id, with the range corresponding to each Id documented elsewhere), we here propose a new concept. This proposal, formally in response to C9ACT-268 "*Define partitioning scheme for DR4 bulk download*" and its related ticket C9OPS-1409 "*Changes in partitioning*", is to use a multi-level partitioning, as outlined below.

Rubin is actively working on a similar scheme, differing in detail, aiming to have it become an IVOA standard in a (few) years' time. Their approach uses an adaptive-depth HEALPix tessellation and is referred to as "HiPSCat" since it can be seen as an extension of the widely-used IVOA HiPS standard. HiPSCat is based on a multi-level approach, i.e., level-4 tiles in low-density areas, level-10 tiles in high-density areas, and anything in between for intermediate densities. It also includes the duplication of some entries around each tile in a margin zone (neighbour cache, extending each tile by of order ~10 arcsec in each direction), stored in a separate HiPSCat tree to keep the original catalogue "clean", to support shuffle-less crossmatches between partitions with other HiPSCat catalogues (see also GAIA-C9-TN-ESAC-EUM-100, "*Efficient cross-matching in Spark*" – DPAC-internal). In parallel, the Rubin team is actively developing a Python-based (Pandas-like) analytics framework known as Large Survey DataBase (LSDB) for high-performance (parallelised) cross-matching of HiPSCat catalogues.

With requirements, limitations, constraints, and trade-offs for Gaia DR4 provided in GAIA-C9-TN-ESAC-EUM-127 ("*Gaia DR4 Dataset partitioning*", DPAC-internal), mostly focused on considerations relevant for Gaia-internal processing and validation operations at DPCE and DPCB, the basic idea as outlined in C9OPS-1409 "*Changes in partitioning*" is to also use a multi-level approach, with (currently) 1788 top-level partitions ("regions", chosen to have at most 3.5 million sources each) for which the ranges have been constrained to match a single tile of a lower-resolution level (resulting in 59 partitions at level 2, 368 partitions at level 3, 463 partitions at level 4, 645 partitions at level 5, and 252 partitions at level 6; see the below figure). The ratio between the maximum number of sources and the average number of sources is ~2.2, which is considered to be an acceptable skewness.



The spatial distribution of the 1788 Gaia DR4 multi-level tiles compatible with a design-choice of a maximum of 3.5 million sources per partition. For reference, with ranges of HEALPix-8 tiles as used in Gaia DR3, GaiaSource for Gaia DR4 would have 798 partitions with at most 3.5 million sources each.

The above scheme, that will be adopted for Gaia DR4-internal processing and is also proposed for the Gaia DR4 bulk-download repository, differs from the HiPSCat scheme since HiPSCat uses a directory structure that encodes the order/pixel number for each HEALPix pixel. This use of folders per resolution level and subfolders for each tile at that resolution is unnecessarily complex in operations and, moreover, introduces three useless columns (norder, dir, and npix) that are not uniquely interpretable between different resolution levels. On the contrary, the Gaia DR4 scheme only has one mandatory folder level ("regions") and, more importantly:

- It names the partitions after the tile in a unique way that makes it simpler to use them operationally (through a unique string composed of a hexadecimal character from 0 to B, representing the 12 level-0 tiles of HEALPix, plus one extra character in the range 0 to 3 for each additional resolution level);
- It retains the natural order of the HEALPix tile indexes and, thus, also the order of the Gaia SourceIds, which is useful in join operations using the sort-merge join algorithm.

Additional folder levels (e.g., "sectors") can be introduced if needed for organisational reasons, e.g., to limit the number of individual files in a folder to avoid stressing the file system.

Since it will be in everyone's interest to adopt an IVOA recommendation for hierarchical partitioning of large, all-sky data sets that suits everyone, in particular Gaia and Rubin, it is proposed to discuss this "conflict" with the Rubin team and investigate whether they are sensitive to our constructive input with the ultimate aim to base the forthcoming IVOA HiPSCat standard on common ground, following the approach presented here. Should that not work out (either not in time for the Gaia DPAC Early Release Operations that commence in spring 2025 [C9ACT-377] or not at all), an emergency fallback solution that could be considered could be to artificially introduce the HiPSCat partitioning scheme in the Gaia DR4 bulk-download repository by means of soft (symbolic) links that mimic the HiPSCat folder structure.

7. PROPOSED CONTENTS FOR GAIA DR4

So far, the bulk-download repository has served (and continues to serve[3]) all public data of all Gaia data releases (Gaia DR1, DR2, EDR3, DR3, and the FPR, in total some 15 TB). This includes all TAP and DataLink products (the latter serialised with the DPAC data model, also known as the RAW data structure). With Gaia DR4, we propose to partially break with this "tradition" and distinguish two use cases, the first one of which does not yet exist but will be relevant for Gaia DR4:

1. Data in the ESA-internal ("geafiles") file repository that can be accessed through ESA Datalabs but not (easily) downloaded in bulk;
2. Data that can be accessed and also (easily) downloaded in bulk from an externally accessible bulk-download file repository ("CDN", see Section 8).

These use cases are addressed in the following two sections that make proposals for each, formally in response to C9OPS-1045 "*Define contents for DR4 bulk download*".

Note: Gaia DR4 auxiliary data products, i.e., data sets that reside in "Cosmos" instead of GACS, are not covered in this proposal. Should some of these products (e.g., CU8 files for astrophysical-parameter inference, CU6 models of the along-scan line-spread function of RVS, or CU5 photometric passbands) need to be made available on "geafiles" in view of their relevance to data exploitation use cases in ESA Datalabs, this is possible but would require further discussion.

[3] Although it may look tempting to remove "obsolete" and/or not-often-accessed products from earlier data releases, it is deemed important – in view of legacy value, relevant for reproducibility of science results based on earlier data releases – to not delete any old data products from the bulk-download repository (while admitting that a valid counterargument would be that all data is anyways available from GACS).

7.1. ESA-internal file repository for use in ESA Datalabs ("geafiles")

Since the ESA-internal file repository ("geafiles") is the one and only direct (NFS-mounted) access point of public Gaia data in ESA Datalabs, the proposal presented here is to serve all TAP and all DataLink products from Gaia DR4 in the ESA-internal file repository. Providing all data products is important to allow large-scale data discovery and exploration in ESA Datalabs, even – or especially – in products that may look uninteresting at first sight (e.g., noisy epoch spectroscopy of faint sources).

7.2. Externally accessible bulk-download file repository ("CDN")

In terms of which Gaia DR4 products to serve through the Content Delivery Network (CDN, see Section 8) for easy bulk download, the proposal presented here – and justified below – is to serve all TAP and all DataLink products **except** the following six "large" DataLink epoch data products:

- EpochAstrometry [72 TB] and EpochImage [166 TB],
- EpochPhotometryCcd [32 TB, covering just the CCD-level G-band photometry],
- RvsEpochSpectrum [61 TB], and
- XpEpochSpectrum + XpEpochCrowding [163 + 6 TB],

totaling 500 TB. Recall that these "large" datasets can of course be queried through GACS[4] and can also be used in massive use cases from within ESA Datalabs without downloading any data. In addition, initiatives under development such as the "Science PlAtform Cloud Infrastructure for Outsize Usage Scenarios" – SPACIOUS – may provide cloud-based access to all Gaia DR4 data. The proposal can hence be summarised as follows:

21 TB of TAP products, **included** in the CDN;

500 TB representing six "large" DataLink epoch data products, **excluded** from the CDN (see list above);

45 TB representing all other DataLink products[5], **included** in the CDN, including (transit-level) EpochPhotometry [9 TB], RvsMeanSpectrum [3 TB], and XpContinuousMeanSpectrum [32 TB].

---

[4] As per C9GACS-1002, the GACS DataLink downlink concept will change for Gaia DR4. The current retrieval strategy is based on a compressed bundle of individual files, created on-the-fly "inside" GACS, that is subsequently offered to the user for download. Since this is CPU, memory, and disk-space intensive for the GACS server, users are limited to access DataLink products for at most 5000 sources per request. In the new scheme, if more than one DataLink product is requested, regardless of the number of sources, a simple ASCII file will be returned that allows users to download the products themselves. This can be through curl or wget commands or through the Metalink standard. This new scheme will allow a significant relaxation of the current 5000-sources limit, making it easier (albeit not faster) for users to download DataLink products for many sources.

[5] "Small" DataLink epoch data products are proposed to be **included** in the CDN, including CrowdedFieldEpochAstrometry, CrowdedFieldEpochPhotometry, NssEpochAstrometry, NssEpochPhotometry, NssRvsEpochData, RvsEpochDataDouble, Rvs-

The reasoning behind **not** serving six "large" epoch DataLink products through the CDN is two(-and-a-half)-fold:

1. It provides a significant CDN cost saving by excluding ~90% of the data volume of Gaia DR4 (Section 8);
2. It is hard to imagine a use case in which an individual without very significant local computing resources could make any use of downloaded datasets weighting tens of TBs. Such use cases would mimic affiliated data centres which already have their own, separate access mechanism defined. Hence, should individuals or institutes request bulk download of one of the excluded products, this could be catered for on request using the standard mechanism defined for affiliate data centres[6].
3. An added advantage of this proposal is that the transfer time to and synching time within the CDN network right after the Gaia DR4 release will be reduced drastically by not including these "large" epoch data products.

An extra option for the bulk-download repository would be the addition of high signal-to-noise, bright-source subsets of the six "large" epoch DataLink products that are missing in the above proposal. Such subsets should (by design) result in at least decimation of the DataLink product volumes, i.e., downloadable and workable file sizes for power users with semi-massive use cases at well-supported institutes. Two scenarios can be envisaged:

1. First define a uniform, self-consistent "golden sample" subset among all products, for instance using G_{RVS} = 14 mag as defining criterion, covering the brightest ~40 million sources. This threshold is used by the DPAC spectroscopic processing pipeline in CU6 to split the bright and faint runs and CU6 delivers all epoch products for the bright subset while the epoch results of the faint sources are more limited. Subsequently complement the RVS epoch spectra with the other epoch products for the same source list. For this particular example, the added data volume would be ~20 TB.
2. Define a separate magnitude limit per epoch product at which the quality of the epoch data shows a significant degradation, for instance G = 16 mag for epoch astrometry and XP epoch spectra (~80 million

EpochDataSingle, RvsEpochSpectrum, SsoXpEpochSpectrum, SsoEpochInfo, SeaEpochPhotometry, BrightSourceEpoch-Astrometry, VariEpochPhotometry, VariEpochAstrometry, and VariRvsEpochData.

[6] Such requests should be made by email, accompanied by a short justification, to the Gaia Helpdesk. Upon confirmed availability of storage and computing resources at the receiving end, access will be granted, or users will be recommended to use the Gaia ESA Archive or ESA Datalabs. In practice, the access will be based on scp (Gaia DR3) or sftp (Gaia DR4, TBC) and only involves a few trivial steps (e.g., provide the access credentials, add the public key, etc.) such that access can be provided "on the fly" without delay and/or limitations to several or even dozens of users / institutes, the only caveat being the effective bandwidth during heavy parallel downloads. To preserve the preferential treatment of affiliated data centres, all of which are given access to the full data set on the date of the Gaia DR4 release, it may be considered to only provide access to others starting (say) one week later.

sources) and G_{RVS} = 14 mag for RVS epoch spectra (~40 million sources). For this particular example, the added data volume would be ~30 TB.

Since DataLink products are not arranged by magnitude, neither of the above is trivial to accomplish in practice (albeit the tests with the ESACGrid reported in C9OPS-1414 appear promising), so that we propose this use case as nice-to-have (with a back-up solution to only make these subsets available after the Gaia DR4 release, i.e., buying time to create them in quiet periods). In order to avoid any confusion, it is preferable to release these bright-source bulk-download subsets in a separate catalogue (i.e., a separate folder in the CDN), e.g., "Gaia DR4 bright", but with identical product names (and data model). Such a catalogue could include other popular tables filtered in the same way, e.g., GaiaSource and/or AstrophysicalParameters, not because these tables are hard to query in Gaia DR4, but because joins between the epoch data and these reduced tables would be much simpler.

8. PROPOSED DELIVERY SCHEME FOR GAIA DR4

As for the delivery scheme, the analysis in Section 3 clearly demonstrates that users actively use the bulk-download repository, which has so far been delivered through a commercial Content Delivery Network (CDN). With the very substantial increase in the number (and data volume) of (DataLink) products in Gaia DR4, in addition to the anticipated hunger of the community for (epoch) data products allowing (semi-)massive science use cases, a further "permanent" increase of the data volume downloaded monthly is deemed very likely.

The CDN can be seen as a commercial extension of the ESA-internal file repository offering users around the globe efficient bulk download as well as alleviating pressure on the ESAC internet bandwidth which hosts the CDN master node. The CDN service, however, could be paused or stopped at any moment. In that case, any users accessing the "ESA-owned" bulk-download URL (http://cdn.gea.esac.esa.int/Gaia/) would simply be pointing to and fetching the data from the ESA-internal "geafiles" machine. In that light, should the CDN service be stopped, there is a risk of an increase in outgoing traffic from ESAC on its current 10 Gbit/s connection to the academic internet. To put some numbers to that risk in the future Gaia DR4 era:

- Some 66 + 30 TB of the 566 TB in total for Gaia DR4 is proposed to be served through the CDN, which represents a factor ~10 increase from the 10 TB for Gaia DR3 and which would increase the total CDN volume by a factor (15 + 96) / 15 ~ 7.
- A hypothetical Gaia DR4 bulk-download volume of 330 TB per month (which reflects a factor ~6 increase from the Gaia DR3 usage – see Section 3) corresponds to an average traffic of 1 Gbit/s, i.e., 10% of the current bandwidth.

- Note: as part of the Euclid Data Space project (formerly known as Digital (data) Exploration Platform, DExP), an increase of the ESAC bandwidth by a factor 2.5 from 10 to 25 Gbit/s is targeted to become available in mid-2025, well before Gaia DR4 (which will "not [be published] before mid-2026"), linked to the Euclid internal DR1 data release in June 2025.
- Note: the monthly bulk-download values mentioned above represent one-month averages. Peaks loads, which may sustain over a few days as per the CDN time-evolution plots in the Gaia ESA Archive monthly reports, can have instantaneous download rates that are a factor ~3 higher than the average. This effectively cancels the upcoming bandwidth gain from the previous bullet.

In short: in the Gaia DR4 era, in case the CDN would be switched off, the risk of bulk-download users fetching data in bulk from ESAC and thereby congesting the ESAC bandwidth (i.e., using >10% for sustained periods) is moderate to small. As a backup mitigation, the ESAC bandwidth to "geafiles" could actually be throttled such that it will never exceed a maximum of (say) 1 Gbit/s if other systems are demanding.

The proposal presented here is to provide the limited CDN service as defined above for Gaia DR4 for the first 6 months following Gaia DR4. This will have a significant but still reasonable cost, despite the "risk" that the ESA bulk-download repository may be the one-and-only access point for the full Gaia DR4 dataset since most (if not all) partner data centres plan to **not** serve all Gaia DR4 products.

It is furthermore proposed that, at 6 months after the Gaia DR4 release, a cost-usage-benefit-risk assessment should be made aimed at the question whether the CDN can be switched off in view of cost savings. The implication would be that all bulk-download requests will fetch all data from the "geafiles" master node at ESAC, with an associated (moderate to small) risk to occupy a significant fraction of the ESAC internet bandwith for sustained periods of time.

As part of the proposed trade-off assessment, a further investigation could be made to offer "free" or "requester pays" data repositories in (commercial or open-source) cloud environments, provided there is Directorate-level support for this.

9. OPEN POINTS

This table summarises a number of open points – ordered chronologically following their appearance in this document – that are (sometimes peripherally) linked to the proposal in this document.

Sect.	Issue	Ideal solution	Backup solution	Contact	Comment
All	All data volumes are "current best estimates" from 24 September 2024, with 10% margin added	Use release-candidate data (expected to become available end 2025 at the earliest)	Use 10% margin on "current best estimates"	Gonzalo Gracia (DPAC Project Office)	The major "uncertainty" on the CDN volume is the nice-to-have bright-source subset (Section 7.2)
4.1	Use of Apache Iceberg and standard tools at ESDC for GACS DataLink back-end	Adoption of Iceberg Metadata API to minimise Gaia dependencies (C9OPS-1416)	Continue use of home-grown scheme (already developed and tested)	Pilar De Teodoro	This is a parallel study within ESDC without further impact on Gaia DR4
4.2	Provision of GaiaSource in ECSV format	Provide a conversion tool (tutorial with notebook; C9DRD-411)	Provide the ECSV data (in addition to Parquet)	Jos de Bruijne	This requires scientific input from the Project Scientist
5.1	Standard to add VOTable-style XML scientific metadata to Parquet footer is missing	Rubin-driven prototyping becomes (IVOA) standard	Use home-grown solution to insert these metadata (and adapt to standard later, if possible)	Enrique Utrilla Molina (supported by Mark Taylor)	This is not a showstopper for Parquet or Iceberg or Gaia DR4
5.2	Apache Iceberg does not support absolute URIs	Upcoming specification update (with unknown timescale) caters for this	Use home-grown solution (C9UTIL-166)	Enrique Utrilla Molina	This is not a showstopper for the use of Iceberg or Gaia DR4
6	Proposed Gaia DR4 partitioning scheme is not HiPSCat compatible	In collaboration with Vera C. Rubin, evolve HiPSCat IVOA proto-standard to meet the Rubin and the Gaia needs	Use soft links in file repository to mimic HiPSCat / adhere to IVOA standard	Jos de Bruijne	This is not a showstopper for Gaia DR4

6	Decision is needed on HiPSCat margin zones (neighbour caches)	Use 10 arcseconds as maximum typical crossmatch radius and include margins only for GaiaSource and AllSourceAstrometry in a separate HiPSCat tree	N/A	Jos de Bruijne	Decision is not needed prior to the Early Release Operations since HiPSCat stores the duplicate sources in the margins in a separate tree
7	Some Gaia DR4 auxiliary data on “Cosmos” may need to be made available in ESA Datalabs	Include selected data on “geafiles” and link to them from Cosmos to avoid duplication and inconsistencies	Duplicate some auxiliary data on “geafiles”	Pedro Garcia-Lario	This is not a showstopper for Gaia DR4
7.1	Accessibility of ESA Datalabs for large-scale exploitation and data mining of Gaia DR4	Unmoderated, open access starts before mid-2026	Offer (part of) Gaia DR4 in the cloud (cf. SPACIOUS)	Jan Reerink (and Pedro Garcia-Lario)	ESA Datalabs is considered to be a “must have” for Gaia DR4
7.2	Pre-release distribution scheme to partner and affiliate data centres is undefined	Secure FTP (CrushFTP?)	Cloud (TBC)	Pedro Garcia-Lario (supported by Jorgo Bakker)	Risk has been reduced through Early Release Operations + most data centres do not want all data
7.2	Define bright-source subsets of “large” epoch products for bulk download	Data volume remains manageable in size for individual (power) users	N/A	Jos de Bruijne	This requires scientific input from the Project Scientist (C9ACT-380)
7.2	Add bright-source subsets of “large” epoch products to CDN for bulk download	Publication together with Gaia DR4	Publication after Gaia DR4	Enrique Utrilla Molina	This is not a showstopper for Gaia DR4
8	Conduct CDN price negotations for 2026	Get discount on storage and/or networking cost(s)	Pay full price and/or reduce CDN contents (data volume)	Richard Schimmer and Uwe Lammers	This is not a showstopper for Gaia DR4