The SIMPLE Archive: Building a Collaborative Database Workflow

David R. Rodriguez[1], Kelle Cruz[2,3,4], W. J. Cooper[5,6], Niall Whiteford[7], Clemence Fontanive[8], Ella Hort[9], Sherelyn Alejandro[2], Robert Blackwell[4], Daniel Terach[10]

[1]*Space Telescope Science Institute, Baltimore, MD, USA;*
drodriguez@stsci.edu

[2]*Hunter College, City University of New York, New York, NY, USA*

[3]*American Museum of Natural History, New York, NY, USA*

[4]*Flatiron Institute, New York, NY, USA*

[5]*Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK*

[6]*Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, I-10025 Pino Torinese 24118, IT*

[7]*University of Edinburgh, Edinburgh, Scotland, United Kingdom*

[8]*CSH, University of Bern, Bern, Switzerland*

[9]*Pomona College, Claremont, CA, USA*

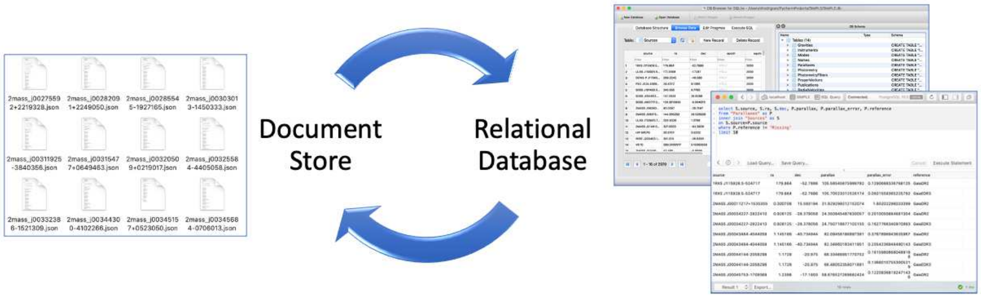
[10]*Pace University, New York, NY, USA*

Abstract.
We present the SIMPLE Archive alongside a Python-based database management package, `AstrodbKit2`. SIMPLE is an archive of low mass stars, brown dwarfs, and exoplanets driven by community curation and review using GitHub. SIMPLE relies on `AstrodbKit2` to convert back and forth from a document-store model of the database, to a more standard relational database that can be used with established packages like SQLAlchemy. In this poster, we present the architecture of the SIMPLE database and how using `AstrodbKit2` facilitates a git workflow for reviewing and approving database modifications.

SIMPLE is available at https://github.com/SIMPLE-AstroDB/SIMPLE-db `AstrodbKit2` is available at https://github.com/dr-rodriguez/AstrodbKit2

1. Motivation and Goals

The goal of SIMPLE is to develop a community-driven archive where individuals can contribute to the knowledge base of these objects. SIMPLE contains a variety of data and supporting metadata for low mass stars, brown dwarfs, and exoplanets, but serves as an example that can be adapted to other fields of study, such as extragalactic or su-


2mass_j00275592+2219328.json
2mass_j00282091+2249050.json
2mass_j00285545-1927165.json
2mass_j00303013-1450333.json
2mass_j00311925-3840356.json
2mass_j00315477+0649463.json
2mass_j00320509+0219017.json
2mass_j00325584-4405058.json
2mass_j00332386-1521309.json
2mass_j00344300-4102266.json
2mass_j00345157+0523050.json
2mass_j00345684-0708013.json
Document Store
Relational Database
Browse Data
Edit Pragmas
Execute SQL
New Record
Delete Record
DB Schema
Tables (14)
Load Query...
Save Query...
Execute Statement
Result 1
Export...

```
select S.source, S.ra, S.dec, P.parallax, P.parallax_error, P.reference
from "Parallaxes" as P
inner join "Sources" as S
on S.source=P.source
where P.reference != "Missing"
limit 10
```

source	ra	dec	parallax	parallax_error	reference
1RXS J115928.5-524717	179.864	-52.7886	[illegible]	[illegible]	GaiaDR2
1RXS J115928.5-524717	179.864	-52.7886	[illegible]	[illegible]	GaiaEDR3
2MASS J00011217+1535355	0.300708	15.588194	[illegible]	[illegible]	GaiaDR2
2MASS J00034227-2822410	0.926125	-28.378058	[illegible]	[illegible]	GaiaDR2
2MASS J00034227-2822410	0.926125	-28.378058	[illegible]	[illegible]	GaiaEDR3
2MASS J00043484-4044058	1.145166	-40.734944	[illegible]	[illegible]	GaiaDR2
2MASS J00043484-4044058	1.145166	-40.734944	[illegible]	[illegible]	GaiaEDR3
2MASS J00044144-2058298	1.1728	-20.975	[illegible]	[illegible]	GaiaDR2
2MASS J00044144-2058298	1.1728	-20.975	[illegible]	[illegible]	GaiaEDR3
2MASS J00045753-1709369	1.2398	-17.1603	[illegible]	[illegible]	GaiaDR2

Figure 1. AstrodbKit2 can convert a database from a document store representation (e.g. a list of JSON files) to a relational database (e.g. SQLite, Postgres, etc).

pernova archives (e.g., The Open Supernova Catalogue). SIMPLE strives to represent each individual astronomical source as a consistent object that can be understood and referenced by others during the community review process using the standard git workflow. Because data in relational databases is organized in a manner that is not always supported by version control, we wrote a python package (`AstrodbKit2`) to facilitate our desired collaborative workflow.

2. AstrodbKit2

`AstrodbKit2` is a Python package that uses `SQLAlchemy` to create and connect to a variety of relational databases (e.g. SQLite, Postgres, MSSQL, etc). `AstrodbKit2` adds a few additional enhancements beyond SQLAlchemy access, including cone searches and Astropy Table output. Tables in `AstrodbKit2` are organized into two types: *Object* tables, which have one-to-many relationships to a single primary object table (i.e. Sources in the SIMPLE database); and *Reference* tables, which have many-to-many relationships against the object tables and are used to store lookup information like publications, telescopes, or instruments. The SIMPLE Archive gathers measurements for low mass objects into a variety of tables all associated to the primary Sources table by their source name. Object tables in SIMPLE include Photometry, Proper Motions, Radial Velocities, and Spectra; Reference tables include Publications, Telescopes, and Instruments. We've designed this schema to be adaptable to other fields of astronomical study.

Using `AstrodbKit2`, we can convert the SIMPLE database from a document store mode, where individual Sources are stored as separate JSON files (which can in turn be loaded into NoSQL databases like MongoDB), to a relational database such as SQLite or Postgres that can be accessed with standard tools (see Figure 1). JSON files, such as the example in Figure 2, serve as the definitive copy of the database for purposes of version control.

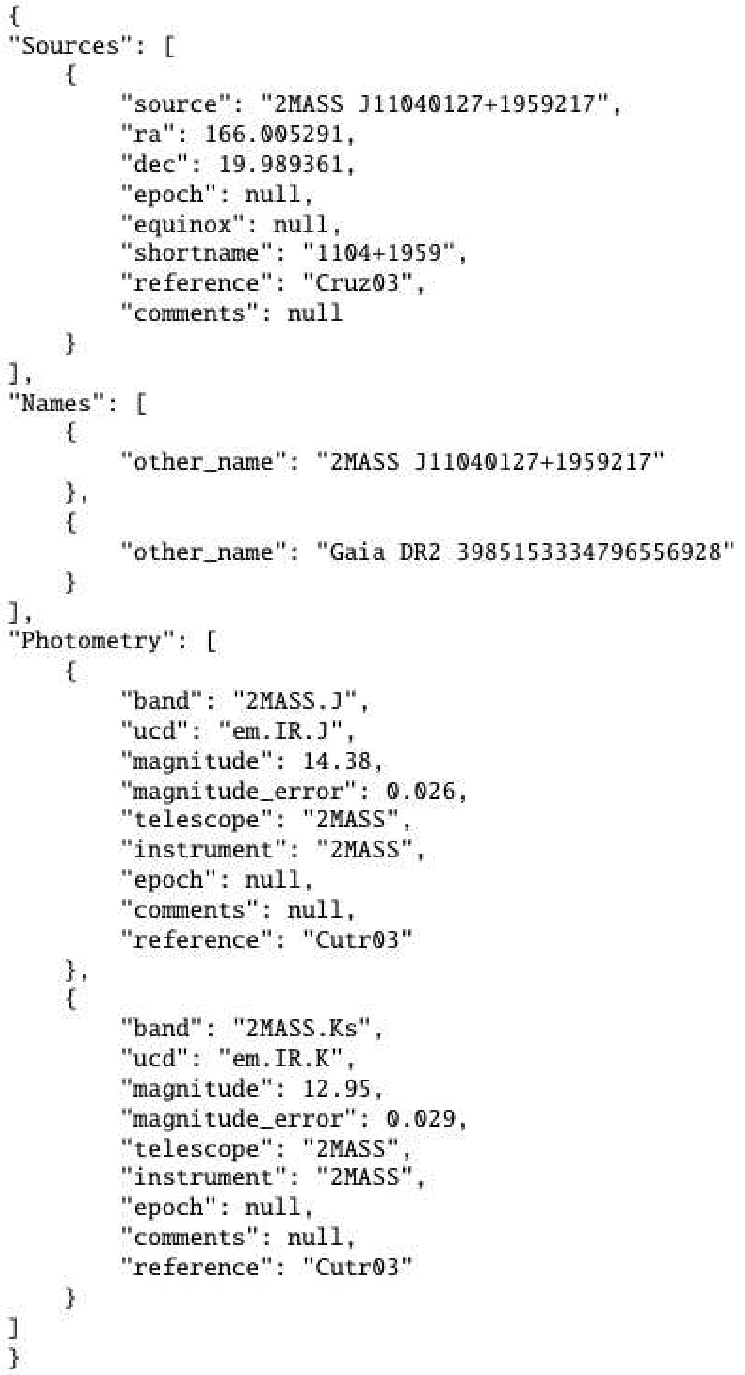
3. Collaborative Workflow

By exporting a database to a JSON document store, we can use git and GitHub to handle version control for our database as well as curate commits via pull requests. An individual user may contain their own copy of SIMPLE, or any other database supported by `AstrodbKit2`. They may make changes in their local branch and push to their copy on GitHub. By issuing a pull request they request their changes be adopted into the main branch of the database. Because the database is stored as individual JSON documents, reviewers can see exactly which objects have been updated and can comment on the changes if needed. By using a plain-text format, we can take advantage of the git differences to provide a clear picture of the changes proposed in a pull request.

As part of the pull request process, automatic tests implemented via GitHub-Actions are run to verify the integrity of the database. This ensures no changes took place that break the functionality of the database and also include some level of verification for the data that has been loaded. Finally, when the pull request is accepted, additional automated tasks can be performed to regenerate the database and push it to external users of the database, such as a graphical user interface.

4. Future Work

We are continuing to develop `AstrodbKit2` and SIMPLE to create a truly collaborative archive. For SIMPLE, we are writing ingests scripts to facility user contributions. These wrap around the underlying `AstrodbKit2` calls to simplify the process. We plan to provide a template GitHub repo that can get users started with a fresh database configured via `AstrodbKit2` and ready for use. We have been building out a `Flask`-powered interface for our SIMPLE database and also plan to convert this to a template to serve other `AstrodbKit2`-enabled databases.



```
{
"Sources": [
    {
        "source": "2MASS J11040127+1959217",
        "ra": 166.005291,
        "dec": 19.989361,
        "epoch": null,
        "equinox": null,
        "shortname": "1104+1959",
        "reference": "Cruz03",
        "comments": null
    }
],
"Names": [
    {
        "other_name": "2MASS J11040127+1959217"
    },
    {
        "other_name": "Gaia DR2 3985153334796556928"
    }
],
"Photometry": [
    {
        "band": "2MASS.J",
        "ucd": "em.IR.J",
        "magnitude": 14.38,
        "magnitude_error": 0.026,
        "telescope": "2MASS",
        "instrument": "2MASS",
        "epoch": null,
        "comments": null,
        "reference": "Cutr03"
    },
    {
        "band": "2MASS.Ks",
        "ucd": "em.IR.K",
        "magnitude": 12.95,
        "magnitude_error": 0.029,
        "telescope": "2MASS",
        "instrument": "2MASS",
        "epoch": null,
        "comments": null,
        "reference": "Cutr03"
    }
]
}
```

Figure 2. Trimmed JSON document for a source in the SIMPLE database