Astronomy and Computing 30 (2020) 100356

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

Mega-Archive and the EURONEAR tools for data mining world astronomical images

O. Vaduvescu [a,b,*], L. Curelaru [c], M. Popescu [b,d]

[a] Isaac Newton Group (ING), Apt. de correos 321, E-38700, Santa Cruz de La Palma, Canary Islands, Spain
[b] Instituto de Astrofisica de Canarias (IAC), Via Lactea, 38205 La Laguna, Tenerife, Spain
[c] Amateur astronomer and software engineer, Brasov, Romania
[d] Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain

ARTICLE INFO

Article history:
Received 25 March 2019
Accepted 2 December 2019
Available online 12 December 2019

Keywords:
Data mining
Asteroids
Near Earth asteroids (NEAs)
Image archives
Mega-Archive
Mega-Precovery

ABSTRACT

The world astronomical image archives offer huge opportunities to time-domain astronomy sciences and other hot topics such as space defense, and astronomical observatories should improve this wealth and make it more accessible in the big data era. In 2010 we introduced the *Mega-Archive* database and the *Mega-Precovery* server for data mining images serendipitously containing Solar system bodies, with focus on near Earth asteroids (NEAs). This paper presents the improvements and introduces some new related data mining tools developed during the last years. Currently, *Mega-Archive* indexed 15 million images available from six major collections and other instrument archives and surveys. This meta-data index collection is daily updated by a crawler which performs automated query of five major collections. Since 2016, these data mining tools are installed on the new dedicated EURONEAR server, and the database migrated to SQL which supports robust and fast queries. To constrain the area to search for moving or fixed objects in images taken by large mosaic cameras, we built the graphical tools *FindCCD* and *FindCCD for Fixed Objects* which overlay the targets across one of seven mosaic cameras, plotting the uncertainty ellipse for poorly observed NEAs. In 2017 we improved *Mega-Precovery*, which offers now two options for the ephemerides and three options for the input (objects defined by designation, orbit or observations). Additionally, we developed *Mega-Archive for Fixed Objects* (MASFO) and *Mega-Archive Search for Double Stars* (MASDS). We include a few use case scenarios and we compare our data mining tools with other few similar services. The huge potential of science imaging archives is still insufficiently exploited. Their use could be strongly enhanced by defining a standard format needed to index the image archives. We recommend to the IAU to define such a standard, asking the observatories to index their image archives in a homogeneous manner.

1. Introduction

The world astronomical image archives provide valuable means to improve the physical properties of Solar System bodies, and in particular of near Earth asteroids (NEAs) which typically remain observable only for short period of times. NEAs represent laboratories for studying the formation and evolution of the minor planets and their physical interactions with the Sun and major planets. Part of NEAs, potentially hazardous asteroids (PHAs) and virtual impactors (VIs) could pose some risk due to their possibility of impact, but they also represent an opportunity for more accessible space missions and eventually future mining industries.

Upon discovery, the recovery and follow-up of NEAs are essential for providing the initial orbital solution and for searching of possible linkage with previously known objects. In most cases, smaller NEAs fade rapidly and become invisible even for largest telescopes which are expensive to access and usually lack time for urgent reaction. However, the existing image archives represent a free opportunity to improve the orbital knowledge based on serendipitous encounters of targets searched using dedicated data mining tools.

Searching for fixed objects (like stars or galaxies) in image archives is straightforward, because only the position of the target is needed to compare with the known telescope pointing and instrument field. The largest astronomical observatories or their collaborating institutions provide simple web searching tools or more sophisticated services which allow searches of fixed objects only in their image archives. To search for fixed objects, in 2009 P. Erwin released the TELARCHIVE Python code[1] (which requires

* Corresponding author at: Isaac Newton Group (ING), Apt. de correos 321, E-38700, Santa Cruz de La Palma, Canary Islands, Spain.
E-mail addresses: ovidiu.vaduvescu@gmail.com, ovidiuv@ing.iac.es (O. Vaduvescu).

https://doi.org/10.1016/j.ascom.2019.100356
2213-1337/

Linux installation) allowing queries of a few remote collections of image archives. Data mining of moving objects and especially those having less accurate orbits becomes more complex, requiring the intersection in space and time of the orbit with the searched archives.

Thanks to their large field, few major photographic plate archives started to be used two decades ago in the first NEA data mining projects by D. Steel in Australia (AANEAS),[2] A. Boattini in Italy (ANEOPP, Boattini et al., 2001) and G. Hahn in Germany (DANEOPS).[3] Due to their initial tiny sizes, CCD cameras have been less appealing to the astronomical community for data mining. Only few major NEA surveys such as NEAT and Spacewatch (whose archives were integrated in *SkyMorph*[4]) or the modern Pan-STARRS (not providing a public tool) allow precovery searches of known asteroids and NEAs in their own archives.

Within the EURONEAR[5] project, since 2007 we data mined some major image archives to improve orbits of known NEAs, involving many amateurs and students in a few public outreach and educational projects. In 2010 we published the *Precovery* tool which allows searches of all know NEAs in a few existing and any other given instrument archive indexed in a simple ASCII metadata format. Using this tool with four powerful archives (CFHTLS-MegaCam, ESO/MPG-WFI, INT-WFC and Subaru-SuprimeCam) we improved the orbits of more than 400 NEAs searched through 800,000 images (Vaduvescu et al., 2009, 2011, 2012, 2013, 2017).

Similar work has been carried out recently for detection and data mining of asteroids in the Kilo-Degree Survey (KiDS) observed with the VST-OmegaCam (Mahlke et al., 2018), resulting in 20,221 Solar system objects (about half known and half unknown asteroids) detected in 346 sq.deg. Other authors proposed similar public asteroid data mining and citizen science projects. In 2012, S. Gwyn proposed the "MegaCam Archival Asteroid Search Verification" (MAASV) project focusing on the CFHT-MegaCam archive (Gwyn and Qi, 2012). Following our call for collaboration in 2010, the former EURONEAR member E. Solano proposed the SVO-NEA (Solano et al., 2014) citizen science tool to precover NEAs in the SDSS DR8, then later the similar SVO-ast project to measure NEAs and Mars-crossers in a collection of surveys (SDSS, UKIDSS, VISTA and VSS) (Solano et al., 2018).

Besides astrometry, image archives are valuable also for deriving photometry and physical properties of astronomical objects. Popescu et al. (2016, 2018) identified all known solar system objects imaged by VISTA Hemisphere Survey, obtaining near-infrared colors for 53,447 asteroids, including 57 NEAs, 431 Mars Crossers, 612 Hungaria asteroids, 51 382 main-belt asteroids, 218 Cybele asteroids, 267 Hilda asteroids, 434 Trojans and 29 Kuiper Belt objects.

In 2010 we introduced the *Mega-Precovery* project (Popescu et al., 2010), aiming to allow public searches of one or a few asteroids or NEAs in a large collection of image archives which aimed initially to include at least one million images. During the last years, the capabilities and databases of this project have been strengthened (Char et al., 2013; Popescu et al., 2014; Vaduvescu et al., 2013), and in this paper we present the latest version of *Mega-Precovery*. To our best knowledge, there is only another similar web-tool allowing searches of asteroids in a large collection of archive images, namely the "Solar System Object Image Search" (SSOIS)[6] hosted at CADC in Canada. This project was published first in 2012 by S. D. J. Gwyn and colleagues (Gwyn et al., 2012b,a), being focused first only on the CADC archive and enhanced later with many other archives (Gwyn et al., 2014; Gwyn, 2015; Gwyn et al., 2015, 2016). Another similar project to search for moving objects in image archives was "The Planetary Archive" (announced in 2014) (Penteado et al., 2014), but its outcome is unknown.

2. Mega-Archive

The largest astronomical observatories and related data centers provide *science archive collections* taken by their instruments and telescopes. These archives are served by common user interfaces allowing anybody to search and download images, spectra, catalogs and other data products. Following our first EURONEAR *Precovery* projects (Vaduvescu et al., 2009, 2011, 2013), in 2010 we joined the first three *instrument archives* (CFHTLS, ESO/MPG and ING/INT) to start building the *Mega-Archive* database (Popescu et al., 2010), with the aim to index at least one million images and add more archives later.

2.1. Archive format and index

To search for any fixed or moving objects, any archive collection must contain some basic information about the observed images. To index the instrument archives within our database, the following "meta-data" fields are essential to be included:

- **Image ID** — *the string needed for downloading the FITS image file*
- **Observing date and time** — *start of exposure in JD format (at least 5 decimals)*
- **Telescope pointings** — *J2000 equatorial coordinates RA (hours with decimals) and DEC (degrees with decimals)*
- **Exposure time** — *in seconds*
- **Filter** — *given as a string*
- **Targeted object or field** — *the name of the actual object or the observed field (string)*

The actual FITS images are linked remotely to the servers which store the archive collections, and they come either in raw original state or sometimes processed (ex from surveys). To avoid any duplicates due to possible processed images available for some surveys, the *Mega-Archive* indexes only the raw images from any given archive, without including any processed images. In the future we plan to add in the output list links to the processed images (whenever these are available), without including the processed images in the *Mega-Archive*.

The instrumental archives are summarized in the master ASCII index file `ArchiveLogs.txt`, which describes all instruments with the following information (one instrument in one line):

- **Collection/Telescope-Instrument** — *Three acronyms naming the instrument archive*
- **Web Address** — *Root internet address serving a given current image ID*
- **FOV** — *Field of view on sky (in sq.arcmin) of the instrument assumed a rectangle*
- **MPC** — *Minor Planet Centre observatory code where the instrument is located*
- **Width** — *Width of the field (in degrees) along RA*
- **Height** — *Height of the field (in degrees) along DEC*
- **Mag** — *Limiting visual magnitude (V-band) in one minute exposure, taking into account the telescope diameter*[7]
- **JD_START** — *Julian date of the first image in the archive*

1 http://www.mpe.mpg.de/erwin/code.
2 http://users.tpg.com.au/users/tps-seti/spacegd4.html.
3 https://web.archive.org/web/20171221091544/http://earn.dlr.de/daneops.
4 https://skys.gsfc.nasa.gov/cgi-bin/skymorph/mobs.pl.
5 http://www.euronear.org.
6 http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/ssois.

7 We adopt the following conventions: V = 19 for 0.4 m, V = 21 for 1 m, V = 23 for 2 m, V = 25 for 4 m, V = 26 for 8 m class telescopes and HST.

- **JD_END** — *Julian date of the last image in the archive*
- **Nr. imgs** — *Number of images in the archive indexed in the given period*

2.2. Instrument etendue and archive etendue

To characterize the efficiency of telescopes and instruments in survey work, the astronomers use the term *etendue* ($A\Omega$) defined as the product of the telescope collecting area A (expressed in square meters) and the surface on sky of the imaging camera Ω (in square degrees). To characterize the data mining efficiency of entire instrument archives, we propose the term *archive etendue* ($A\Omega\, A$) defined as the product of the etendue and the number of raw science images included in the given instrument archive.

2.3. Archive collections

We briefly present the six image archive collections included in *Mega-Archive* by **29 August 2019**.

The Canadian Astronomical Data Centre (CADC) archive was established in 1986 to store the Hubble Space Telescope (HST) images and later the Canada–France–Hawaii Telescope (CFHT) data. Today it manages data taken with many other North American telescopes. Since the beginning of our project we ingested some of the CFHT archives. Currently, the *Mega-Archive* includes about two million images from 29 instruments (visible and NIR) installed on 8 telescopes linked to the CADC collection. The CADC Advanced Search[8] allows multiple archive selections using a CGI form which presents a few selection menus. The CADC archive also allows programmatic queries, and for each instrument we used the following options to retrieve the needed meta-data for all available optical and NIR images (all other search options being left as default):

- *Obs. Constraints = SCIENCE DATA ONLY*
- *Band = OPTICAL + INFRARED*
- *Cal. Level = RAW STANDARD*
- *Data Type = IMAGE*
- *Obs. Type = IMAGING + OBJECT*
- *Instrument = Cycling one by one*
- *Temporal Constraints = Obs Date eventually split in intervals to avoid 30,000 lines limit*
- *Results page = Download CSV*

The European Southern Observatory (ESO) started in 1994 to archive NTT images (Albrecht and Benvenuti, 1994), and provides today one of the largest collection taken with major ESO telescopes in La Silla and Paranal. The *Mega-Archive* includes more than three million images from 15 instruments installed on 9 ESO telescopes. In order to query the ESO archive, we used the *ESO Science Archive Facility*[9] which allows to search and retrieve Raw Data using a CGI form showing two main selection blocks about the target/program and observing information. For each instrument we used the following options to search all available images (all other existing search options being left as default):

- *Return = ALL FIELDS*
- *Category = SCIENCE*
- *Type = OBJECT*
- *Mode = IMAGE*
- *Imaging = Cycling instruments one by one*
- *Start/End = Eventually split in intervals to avoid browser crashes*
- *Output preferences = ASCII CSV/DOWNLOAD*

Las Cumbres Observatory Global Telescope (LCOGT) provides since 2013 their Science Archive (Lister et al., 2013), storing all the images taken with the instruments and telescopes of this network. In 2017 we collected 21 available imagers from all LCOGT telescopes, treating each instrument mounted on any telescope as one distinct archive (due to the distinct locations and MPC codes of each telescope). Today *Mega-Archive* includes more than one million images from the entire LCOGT collection. The LCOGT Science Archive[10] could be manually or programmatically queried based on the following selections:

- *ObsType = EXPOSE*
- *Reduction Level = RAW*
- *Site = Cycling all*
- *Telescope = Cycling all*
- *Showing = 1000 records per page, then cycling all pages*

The Isaac Newton Group (ING) maintains since 2014 its ING science data archive based on telescope observing logs, the images being stored by the Cambridge Astronomy Survey Unit (CASU Astronomical Datacentre).[11] We used the ING facility to ingest in the *Mega-Archive* almost 1.5 million images taken with 8 instruments mounted at all 3 ING telescopes. Besides the night observing logs, the ING query form[12] allows an "Advanced search and browse" section for retrieval of meta-data and links to the available ING images sorted based on a few constraints. We used the following parameters to search all available ING images (all other available search options on this form being left as default):

- *Telescope = ALL*
- *Mode = IMAGING*
- *Frame Type = SCIENCE*
- *Dates Start/End = Eventually split in intervals to avoid 10,000 lines limit*
- *Output format = ASCII*

The U.S. National Optical Astronomical Observatory (NOAO) provides since 2007 the *NOAO Science Archive* (known also as the NOAO NVO Portal) (Miller et al., 2007) which allows VO-compliant access to imaging data taken with major American telescopes in Arizona and Chile. In 2012 we manually retrieved the first NVO instrument archives to ingest in *Mega-Archive*. Until today we included almost 5 million NVO images from 19 imaging instruments mounted on 10 telescopes. The NVO collection is available for searches from the NOAO Science Archive website.[13] Two options are possible, proprietary data access (requiring login) and general search for NOAO data (available for any other user). We used the second choice with the following options in the Simple Query Form:

- *Telescope and Instrument = Cycling all imagers one by one*
- *Observing Calendar Date = Eventually split in intervals to avoid 50,000 lines limit*

The Subaru-Mitaka-Okayama-Kiso Archive (SMOKA) was built in 2001 to store the images taken with major Japanese telescopes (Baba et al., 2002). It requires user registration and login before providing the FITS images, but the meta-data could be retrieved without registration. The *Mega-Archive* includes more than one million images from 17 instruments installed on 4

8 http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/search.

9 http://archive.eso.org/eso/eso_archive_main.html.

10 https://archive.lco.global.

11 http://casu.ast.cam.ac.uk/casuadc/ingarch/query.

12 http://www.ing.iac.es/astronomy/observing/inglogs.php.

13 http://archive.noao.edu.

Japanese telescopes. The SMOKA Archive Advance Search[14] provides meta-data selected based on the instruments and date intervals, allowing an output of maximum 20,000 rows in one query. We used the following parameters to search all available SMOKA images:

- *Instruments = Cycling one by one*
- *Observation Mode = IMAG*
- *Data Type = OBJECT + COMPARISON + TEST*
- *Obs Category = SCIENCE + ENGINEERING*
- *Output format = ASCII*
- *Observation Date = Eventually split in intervals to avoid 10,000 lines limit*
- *Output Columns = FRAMEID + DATE_OBS + OBJECT + FILTER + RA2000 + DEC2000 + UT_START + EXPTIME*

Three mosaic cameras (Subaru-HSC, Subaru-SuprimeCam and Kiso-KWFC) must be queried using mode *SHOT* (instead of *FRAME*) in order to provide one output line for each observed mosaic image (instead of many images representing many CCDs available for these large field cameras). That is, we do not include in *Mega-Archive* individual CCDs of mosaic cameras.

2.4. Other instrument archives

Additionally to major archive collections, we indexed two other instrument archives and surveys:

The Wide Field Imager (WFI) of the Anglo-Australian Telescope (AAT) was ingested in the *Mega-Archive* in 2012. We considered important this instrument due to its large etendue, although this archive includes only about 5000 images observed in the period 2000–2006. To manually collect the meta-data we used the former AAT Data Archive form[15] which was recently integrated in a modern portal serving more instrument archives.[16]

The Sloan Digital Sky Survey (SDSS-III DS9 release) was included in 2013, being queried via the DR9 Science Archive Server.[17] The *Mega-Archive* includes almost one million images from the SDSS-III DR9 archive.

2.5. Migration to the SQL architecture

The classic format of the *Mega-Archive* used between 2010 and 2017 was ASCII (txt files), each observation being stored in one line with columns defining data presented in Section 2.1 (separated by the "|" character), and each instrument archive being stored in one file.

Following the migration to the actual EURONEAR dedicated server, in July 2017 the old *Mega-Archive* ASCII database migrated to a SQL relational database. This change was necessary in order to keep the search time as short as possible, due to the growing number of images and instrument indexes contained in the *Mega-Archive*. We already presented the architecture of *Mega-Archive* and its connection to the new EURONEAR tools as a flowchart in another recent paper (Curelaru et al., 2019). Thanks to the new SQL architecture, the searches of *Mega-Archive* have become faster for Solar system objects (taking a few minutes necessary to build accurate ephemerides needed to search the entire archive of 15 million observations for one asteroid) and extremely fast for fixed objects (taking only few fractions of a second for one target).

2.6. Mega-Archive daily update crawler

Major observatories ingest in their archives observations and images on a daily basis, which makes these databases appealing for searches related to very rapid time-domain phenomena. In other cases, dedicated surveys become ingested in archives much later, following the image reduction, project completion or expiration of proprietorship periods (typically one year following the observing date). Nevertheless, some newly discovered objects could rapidly raise attention of the astronomical communities, society and mass-media, such as a potential impactor asteroid, the past imminent impactors 2008 TC3, 2014 AA and 2018 LA, or the interstellar object 'Oumuamua. Sometimes such objects become very rapidly invisible even in largest telescopes, thus data mining remains the only possibility to prevent loss and obtain more information. In these cases, data mining of **up to date** existing archives for precovery observations (preferably closer in time to discovery) could bring crucial information regarding the objects' nature, orbital classification and probability of virtual or imminent impacts.

A *web crawler* (known also as a spider), is an Internet robot tool which systematically browses the World Wide Web with the aim to create entries to build or update a search engine index (e.g. Olston and Najork, 2007). For automated update of the *Mega-Archive*, in 2014 we designed and implemented a crawler to check major archive collections on a daily basis. This PHP tool includes some scripts to query the ESO, CADC, SMOKA, ING and LCO collections, crawling the programmable interfaces published by each server (which very rarely need some minor changes of format). The only collection not able for automate crawling is the NVO archive, which we update manually on a yearly basis. The daily update tool starts automatically in cron every midnight and typically takes less than one hour to update all instrument archives in the *Mega-Archive*. The SQL database holding the instrument archives is updated with the new information, and also the master `ArchiveLogs.txt` file is updated (changing the number of images and the observed interval, needed for comparison with next day crawling). Each instrument archive is searched at once, checking only the most recent data (ingested in the collection since the previous midnight/day). Most problems (possible malfunctions due to servers or internet connection) could be traced and corrected next day by the admin who checks the logs listing the operations with all instrumental archives crawled in all collections, but in practice such system errors very rarely happen.

Some instruments and therefore archives have become dormant or were retired completely, and in this case the End Date column in the Instrument Selection table show earlier dates. Other instruments continue to be active, but they observe depending on the scheduling periods, and in this case the End Date is closer or coincides with current date (2018 or 2019). Some archives (such as ESO) publish up to date meta-data (therefore the End Date match the current date), while other archives (such as ING) do not publish any information beyond proprietorship period due to safety reasons (thus the End Date will usually be one year before the current date). Other few collections do not allow automate crawling (such as NVO), thus in this case the End Date will simply represent the date of our last manual crawling.

2.7. Mega-archive instruments comparison

As of **29 August 2019**, the *Mega-Archive* **includes 111 instruments adding 15,797,519 images**, listed in Table 1. The columns list several parameters, namely the *archive etendue* ($A\Omega\, A$, in square meters by square degrees), the field of view of the instrument (FOV, in square arcminutes), the archival date interval, and

14 https://smoka.nao.ac.jp/fssearch.jsp.

15 Old server http://site.aao.gov.au/arc-bin/wdb/aat_database/observation_log/make.

16 New server https://datacentral.org.au/archives.

17 https://dr9.sdss.org/fields.

Table 1
15,797,519 raw images from 111 instrument archives are indexed in the *Mega-Archive* **by 29 August 2019**.

Instrument archive	*AΩ A*	FOV	Start Date	End Date	Images
AAT-WFI	14 831	1089.0	2000-08-21	2006-02-05	4 453
CADC/APASS	42 242	31 329.0	2010-04-11	2014-06-02	121 350
CADC/CFHT-aobir	184	1.4	1997-12-10	2011-11-23	45 108
CADC/CFHT-aobvis	6	1.4	1996-05-04	1999-03-30	1 573
CADC/CFHT-CFHTIR	1 396	17.6	2001-01-10	2005-11-17	27 996
CADC/CFHT-Megacam	2 502 061	3600.0	2003-02-22	2019-08-28	245 782
CADC/CFHT-MOCAM	421	225.0	1995-05-28	1995-11-29	662
CADC/CFHT-REDEYE	241	4.4	1993-02-04	1998-09-03	19 312
CADC/CFHT-WIRCam	447 422	466.6	2005-11-18	2019-08-13	339 129
CADC/CTIO-CPAPIR	19 575	348.2	2005-02-13	2019-07-22	114 534
CADC/CTIO-CPAPIRVIS	1 623	348.2	2008-01-16	2019-05-06	9 498
CADC/DAO-E2VCCD	10 611	253.6	2008-02-11	2019-08-28	60 490
CADC/GeminiN-GMOS	12 258	30.5	2001-08-14	2019-07-03	28 106
CADC/GeminiN-GNIRS	196	1.4	2004-09-05	2018-01-30	9 491
CADC/GeminiN-NIRINIR	10 613	3.2	2002-02-22	2019-07-03	228 837
CADC/GeminiN-NIRIVIS	101	3.2	2004-06-28	2015-10-07	2 181
CADC/GeminiS-Flamingos2	26 737	36.0	2013-10-10	2019-07-03	51 886
CADC/GeminiS-GMOS	13 423	30.5	2003-02-28	2019-07-03	30 776
CADC/HST-ACS	1 277	11.7	2002-04-02	2018-03-19	88 290
CADC/HST-NICMOS-IR	114	0.8	1997-03-21	2008-09-10	113 867
CADC/HST-NICMOS-VIS	2	0.8	1997-07-23	2008-08-21	1 730
CADC/HST-WFC3/NIR	505	4.9	2009-07-05	2019-08-28	82 915
CADC/HST-WFC3/Vis	716	7.3	2009-07-13	2019-08-28	79 493
CADC/HST-WFPC2NIR	15	9.0	1994-03-05	2008-04-06	1 366
CADC/HST-WFPC2OPT	1 803	9.0	1993-12-20	2009-05-12	162 072
CADC/HST-WFPCOPT	41	9.0	1989-11-30	1993-12-03	3 691
CADC/Mt.Stromlo-MACHO	114 404	1794.4	1992-07-21	2002-11-19	197 868
CADC/UKIRT-Michelle	160	1.4	2001-10-04	2004-04-23	35 337
CADC/UKIRT-UFTINIR	1 871	3.2	1999-10-16	2011-07-18	183 298
ESO/3.6m-TIMMI2	305	1.9	2004-05-08	2006-06-28	63 817
ESO/MPG-WFI	150 380	1149.1	1999-04-15	2019-08-28	144 959
ESO/NTT-EMMI	3 834	83.2	2004-03-17	2008-04-01	17 541
ESO/NTT-SOFI	18 889	24.2	2006-03-30	2019-08-23	296 951
ESO/NTT-SUSI2	1 366	30.5	2004-04-02	2008-12-29	17 057
ESO/VISTA-VIRCAM	26 263 195	4726.6	2009-10-16	2019-08-28	1 668 327
ESO/VLT-EFOSC2	26 347	16.6	2004-07-03	2019-08-24	109 323
ESO/VLT-FORS1	23 860	46.0	1999-01-23	2009-03-26	35 852
ESO/VLT-FORS2	139 744	46.0	1999-10-30	2019-08-26	209 977
ESO/VLT-HAWKI	114 688	56.2	2007-08-01	2019-08-25	140 829
ESO/VLT-ISAAC	19 153	6.4	1999-03-01	2013-12-13	208 325
ESO/VLT-NACO	5 093	0.9	2001-12-02	2019-08-25	381 728
ESO/VLT-VIMOS	1 134 679	841.8	2002-10-30	2018-03-24	93 108
ESO/VLT-VISIR	137	0.1	2004-05-11	2015-11-22	73 068
ESO/VST-OmegaCam	1 369 786	3664.5	2011-04-01	2019-08-28	288 773
ING/INT-PFCCD	221	121.0	1993-01-13	2014-05-17	1 492
ING/INT-WFC	954 969	1169.6	2002-01-02	2018-08-20	666 501
ING/JKT-JAG	2 158	100.4	2002-01-03	2003-08-01	98 553
ING/WHT-ACAM	25 459	63.7	2009-06-10	2018-08-27	103 887
ING/WHT-LDSS	1 276	100.4	1993-03-18	2000-03-24	3 302
ING/WHT-LIRIS	39 975	18.1	2004-03-03	2018-07-03	572 398
ING/WHT-PFIP	20 328	256.6	1993-12-03	2016-09-22	20 582
ING/WHT-WHIRCAM	13	1.0	1995-02-10	1999-06-28	3 174
LCO/CTIO-0M4–09	320	582.1	2017-12-04	2019-08-28	16 511
LCO/CTIO-1M0–04	1 887	697.0	2014-05-12	2019-08-28	12 495
LCO/CTIO-1M0–05	16 152	697.0	2014-06-01	2019-08-28	106 963
LCO/CTIO-1M0–09	987	697.0	2014-05-31	2019-08-28	6 533
LCO/GOL-1M0–02	213	169.5	2014-05-02	2019-06-08	5 803
LCO/HLK-0M4–04	295	582.1	2016-03-31	2019-08-28	15 184
LCO/HLK-0M4–06	1 508	582.1	2016-03-31	2019-08-28	77 695
LCO/HLK-2M0–01	5 674	100.4	2014-05-04	2019-08-28	64 788
LCO/MDO-1M0–08	10 343	697.0	2014-05-02	2019-08-27	68 493
LCO/OT-0M4–10	397	582.1	2015-06-03	2019-08-28	18 909
LCO/OT-0M4–14	1 931	582.1	2015-02-13	2019-08-28	91 845
LCO/OT-0M4-XX	54	582.1	2015-02-15	2015-06-02	2 577
LCO/SAAO-1M0–10	17 643	697.0	2014-05-01	2019-08-28	113 914
LCO/SAAO-1M0–12	10 240	697.0	2014-05-01	2019-08-28	66 117
LCO/SAAO-1M0–13	15 676	697.0	2014-05-01	2019-08-28	101 211
LCO/SED-0M8–01	513	125.3	2014-05-02	2018-02-14	36 883
LCO/SSO-0M4–03	670	582.1	2015-10-15	2019-08-28	34 545
LCO/SSO-0M4–05	495	582.1	2015-10-02	2019-08-28	25 492
LCO/SSO-1M0–03	3 515	169.5	2014-05-01	2019-08-06	95 712
LCO/SSO-1M0–11	2 424	169.5	2014-05-01	2019-08-28	64 351
LCO/SSO-2M0–02	3 216	100.4	2014-05-01	2019-08-28	36 730
NVO/Blanco-DECam	21 873 144	17 424.0	2012-09-12	2019-07-11	451 473

(*continued on next page*)

Table 1 (*continued*).

Instrument archive	$A\Omega A$	FOV	Start Date	End Date	Images
NVO/Blanco-ISPI	30 117	104.0	2006-06-05	2014-04-10	104 106
NVO/Blanco-MOSAIC2	322 956	1388.2	2004-08-11	2012-02-20	83 665
NVO/Blanco-NEWFIRM	126 046	785.1	2010-07-03	2011-05-19	57 738
NVO/Bok-CCD	197 417	4844.2	2015-01-08	2017-06-04	35 610
NVO/CTIO0.9m-CCD	24 679	182.2	2000-04-04	2017-05-17	826 261
NVO/CTIO1.3m-ANDICAM	4 524	5.8	2000-04-04	2017-06-01	2 141 898
NVO/CTIO1m-Y4KCam	22 409	392.0	2006-12-14	2014-06-02	263 817
NVO/KPNO2.1m-CCD	859	29.2	2010-09-03	2014-07-02	35 940
NVO/Mayall-Misc	64	0.4	2007-02-21	2015-06-29	70 005
NVO/Mayall-MOSAIC	109 052	1296.0	2004-09-01	2013-02-07	32 998
NVO/Mayall-NEWFIRM	260 610	785.1	2007-06-30	2013-02-04	130 171
NVO/SOAR-OptImg	19 150	27.2	2005-11-07	2017-06-14	203 377
NVO/SOAR-SAM	581	9.0	2014-03-05	2017-05-23	18 689
NVO/SOAR-Spartan	7 929	25.4	2011-08-16	2017-05-06	90 331
NVO/WIYN0.9m-MOSAIC	11 534	3478.6	2004-11-25	2010-03-29	20 232
NVO/WIYN0.9m-S2KB	11 705	416.2	2004-09-09	2014-11-17	171 614
NVO/WIYN3.5m-MiniMosaic	2 887	100.4	2004-12-18	2013-04-20	12 989
NVO/WIYN3.5m-WHIRC	6 009	10.9	2008-04-15	2017-05-19	249 238
SDSS-DR9	131 306	140.8	1998-09-19	2009-11-18	938 046
SMOKA/KANATA-HOWPol	12 826	225.0	2008-12-06	2018-02-23	116 138
SMOKA/KANATA-HONIR	3 764	100.4	2014-03-05	2018-02-24	76 380
SMOKA/Kiso-1kCCD	1 123	155.8	1993-02-26	2000-10-24	30 004
SMOKA/Kiso-2kCCD	70 609	2498.0	1998-09-08	2012-02-27	117 640
SMOKA/Kiso-KWFC	415 904	15 620.0	2012-04-02	2018-02-14	110 815
SMOKA/Okayama-ISLE	1 219	17.1	2006-11-15	2017-04-03	92 268
SMOKA/Okayama-KOOLS	64	21.8	2008-01-04	2016-01-10	3 822
SMOKA/Okayama-OASIS	85	16.2	1998-08-14	1998-12-13	6 843
SMOKA/Subaru-CAC	45	0.7	1999-01-06	2000-06-21	4 318
SMOKA/Subaru-CISCO	8 046	3.7	1999-01-12	2007-04-09	148 791
SMOKA/Subaru-COMICS	472	0.5	1999-12-14	2018-01-29	62 079
SMOKA/Subaru-FOCAS	6 367	36.0	2000-02-02	2018-01-02	12 057
SMOKA/Subaru-HSC	6 749 712	8100.0	2014-03-26	2018-02-18	56 805
SMOKA/Subaru-ICRS	2 395	1.0	2000-09-22	2018-02-05	156 921
SMOKA/Subaru-Kyoto	61	3.9	2004-04-08	2015-09-24	1 065
SMOKA/Subaru-MOIRCS	31 804	28.2	2004-06-11	2018-02-07	76 826
SMOKA/Subaru-SCExAO	405	0.2	2000-01-22	2008-07-15	119 955
SMOKA/Subaru-SuprimeCam	1 143 209	918.5	1999-01-05	2017-05-30	93 228

the total number of raw science images indexed to date. Using these data, we could compare the archives and assess their use for data mining.

Fig. 1 plots the histogram counting the total number of archives versus the number of images in each archive. The first three instruments are actually three surveys, namely CTIO1.3m (mostly 2MASS), VISTA and SDSS-DR9, followed by CTIO0.9m, INT-WFC, WHT-LIRIS, Blanco-DECam, VLT-NACO and CFHT-WIRCam.

Based on the etendue ($A\Omega$), the most powerful survey facilities in the present *Mega-Archive* are Subaru-HSC, Blanco-DECam, VISTA-VIRCAM, Subaru-SuprimeCam, VLT-VIMOS, CFHT-MegaCam, Bok and VST-OmegaCam. These and the following archives are plotted in the histogram in Fig. 2. The analysis is extended in Fig. 3 which presents the histogram counting the number of archives versus the *archive etendue* ($A\Omega A$). Based on this factor, the most productive facilities to date are VISTA-VIRCAM, Blanco-DECam, VST-OmegaCam, Subaru-HSC, CFHT-MegaCam, VST-OmegaCam, Subaru-SuprimeCam, VLT-VIMOS and INT-WFC. Quite many near infrared (NIR) instruments populate Figs. 1 and 3 due to their fast cadence of images explained by short exposure times and very rapid readouts compared to the other instruments in visible.

Most exposures are shorter than 80s, probably owing to the fast cadence of NIR instruments and other fast time-domain science, as one can observe the histogram in Fig. 4. This makes *Mega-Archive* and data mining very appealing for asteroid science, as most targets should have stellar-like aspect (instead of longer trails) which results in easily measurable astrometry and photometry.

The number of images as a function of the observing time is shown in Fig. 5. The great majority of the images in the **current** *Mega-Archive* were collected since 2005. Apparently, the plot is leveling around one million images taken every year. The last three years show lower numbers due to ingestion delays and proprietorship embargoes which make later images invisible to some collections and the *Mega-Archive*.

3. Data mining tools

Mega-Archive was designed in 2010 for asteroid searches and it was improved since 2017, following the migration of EURONEAR to the new private dedicated server. We will present next the applications accessing the *Mega-Archive*.

3.1. Mega-Precovery for moving objects

In 2010 we introduced *Mega-Precovery* to search for Solar System Objects using the *Mega-Archive* collection (Popescu et al., 2010). Our project targeted mostly the precovery and recovery of observations for PHAs and VIs. The algorithm and a flowchart was presented in Vaduvescu et al. (2013). During last years, several major improvements were added to the project, significantly extending its functionality. These are described below.

The original version (2010–2016) resided on the old EURONEAR IMCCE server. It was embedded in the old wikiplugin PHP environment and used the old ASCII archives format. The code migrated to the new EURONEAR server where it is preserved (under the Older Tools section) as version v.1.[18] The second version (2016–2017) updated the code to pure PHP 7 environment installed in the new EURONEAR server, and converted the *Mega-Archive* to the relational SQL database. The code has been

18 http://www.euronear.org/tools/megaprecoveryV1.php.

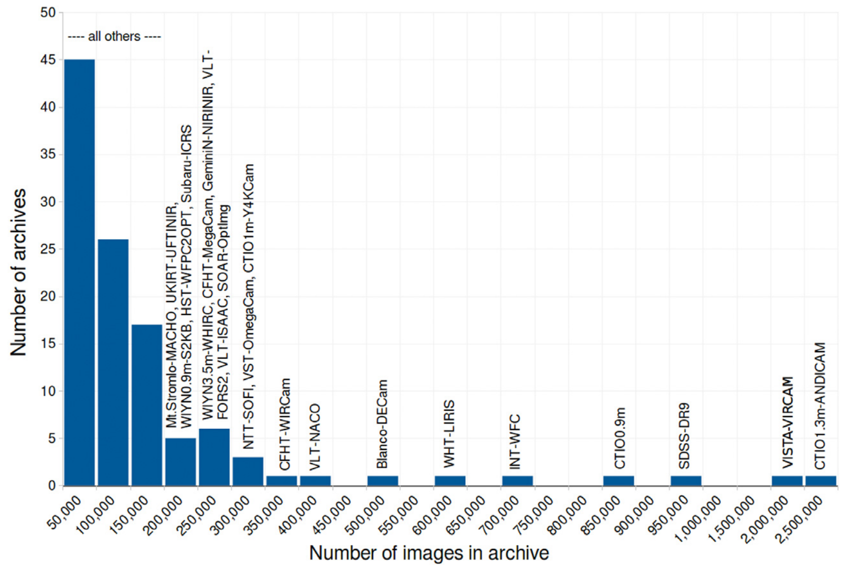

Fig. 1. The number of instrument archives versus the number of images in the *Mega-Archive*. Only the most populated instruments are labeled, and the step of the X axis is changed above $X = 1,000,000$ to be able to accommodate the most prolific two archives.

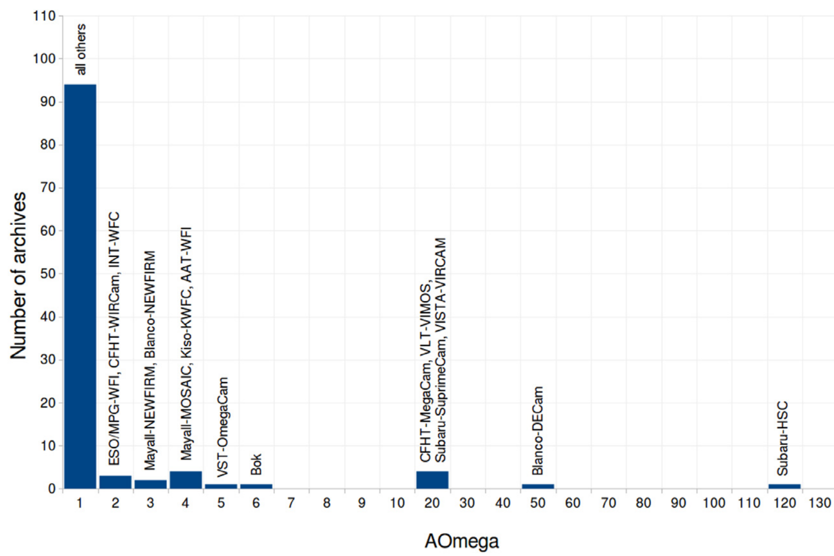

Fig. 2. The number of instrument archives versus the etendue ($A\Omega$) in the *Mega-Archive*. Only the most powerful instruments are labeled and the step on the X axis is changed above $A\Omega = 10$.

preserved as version v.2.[19] The actual version 3 (released in October 2017) has implemented two options for the calculus of the ephemerides of the searched object, allowing three options for the input which are offered as three different pages linked from the main EURONEAR Data Mining Tools.[20] Moreover, additionally to the standard web form, the actual *Mega-Precovery* v.3 allows programmable queries via HTTP commands.

3.1.1. *Mega-Precovery from designations*

This is our first classic search for one or a few known asteroids or comets, based on the name, number or designation.[21] The user can choose between two ephemerides generators, either using the classic Miriade server[22] (Berthier et al., 2009) (available in *Mega-Precovery* v.1, v.2 and v.3) or the *OrbFit*[23] software (Orbfit Consortium, 2011) installed only in *Mega-Precovery* v.3. The SsODNet service at IMCCE[24] (Berthier et al., 2007) is used to check object designations.

3.1.2. *Mega-Precovery from orbit*

If the object does not have an official name, number or designation, but its orbital elements have been derived, then these can be input into *Mega-Precovery*.[25] The required elements are the semimajor axis a, the eccentricity e, the inclination i, the ascending node Ω, the argument of periapsis ω, the mean anomaly M, and the epoch *MJD*.

19 http://www.euronear.org/tools/megaprecoveryV2.php.
20 http://www.euronear.org/tools.php.
21 http://www.euronear.org/tools/megaprecdes.php.
22 http://vo.imcce.fr/webservices/miriade.
23 http://adams.dm.unipi.it/orbfit.
24 http://vo.imcce.fr/webservices/ssodnet.
25 http://www.euronear.org/tools/megaprecorb.php.

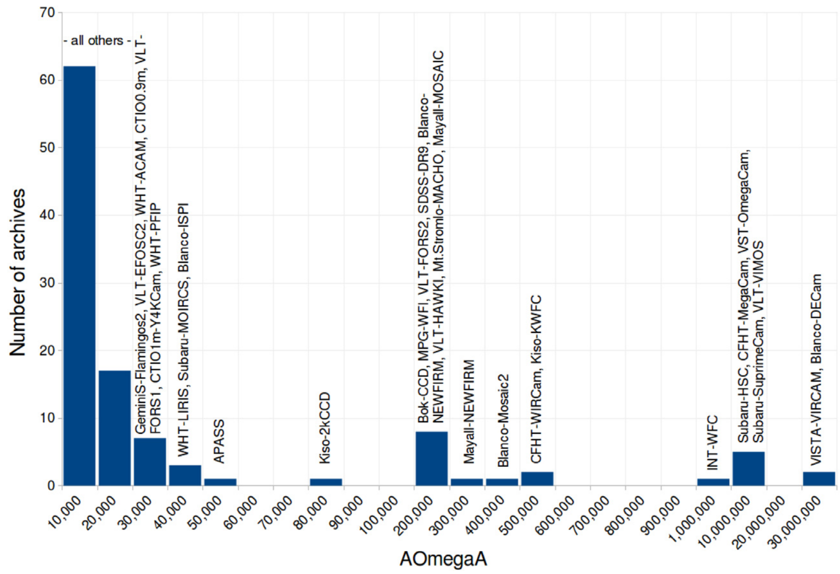

Fig. 3. The number of instrument archives versus the *archive etendue* ($A\Omega A$) in the *Mega-Archive*. Only the most powerful instrument archives are labeled across the variable step for the X axis (changed at $X = 100,000$ and $X = 1,000,000$), to be able to accommodate the most prolific archives.

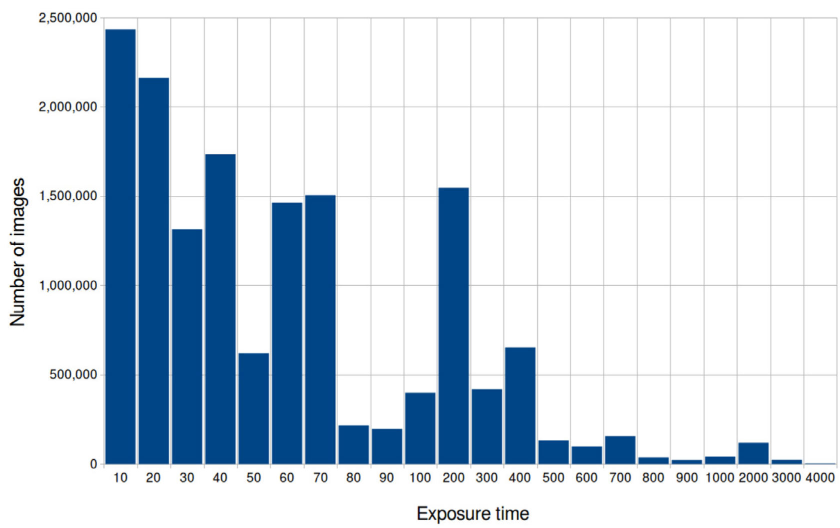

Fig. 4. The number of images versus the exposure time (in seconds) in the entire *Mega-Archive*. The step on the X axis is changed twice by one order for the last cuts, to be able to accommodate all exposures.

3.1.3. Mega-Precovery from observations

If the object was discovered very recently or if the connection with the ephemerides server is broken, then the target can be searched using an input consisting in a block of observations in MPC format.[26] In this case, *OrbFit* is run locally on the EURONEAR server, either using a single step ephemerides model (given at least 3 observations taken in one or few nights following discovery) or a three steps model (defined by blocks of discovery and follow-up data, first opposition data, and multi-opposition data).

All three search options are provided with a few advanced options: All instruments or only some archives selected by the user could be searched at once in *Mega-Precovery*. The computational interval can be constrained in an interval around the observations (to speed up the search of the entire *Mega-Archive*) or left default to cover all archives. The safety search border allows some flexibility due to telescope pointings (sometimes inaccurate, affected by small dithering during multiple run exposures, or due to the pointing not matching the camera center). The default value of the safety border is 0.02°, which means that the search could allow any target outside the image by up to 1.2′ to be included in the results. The running mode allows the user to choose between fast geocentric ephemerides (where all ephemerides and steps are calculated at geocentre) and slow topocentric mode (where each integration step is calculated for the topocentric observatory position (recorded as MPC code in `ArchiveLogs.txt` where the current searched instrument is hosted — this is important only for very close flyby observations of PHAs or VIs where topocentric correction could become important). Finally, three output options are provided: HTML, simple text (formatted as a table), and CSV (fields separated by coma). The text and CSV formats were added to facilitate the portability of the results via copy/paste to other tools,

The *Mega-Precovery* output columns are: the archive name (in the archive/telescope-instrument format), the image ID (tentatively linked in HTML to the collection server to retrieve the

26 http://www.euronear.org/tools/megaprecobs.php.

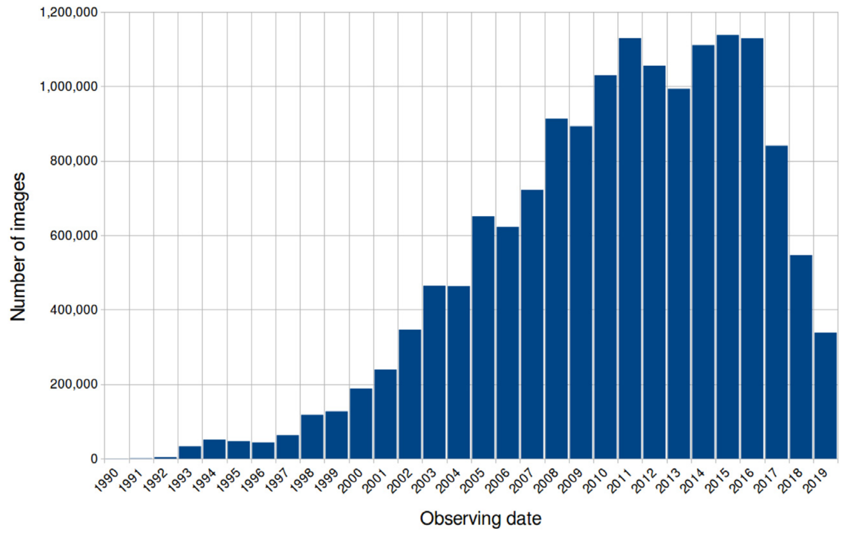

Fig. 5. The number of images indexed in the *Mega-Archive* versus the observing date (year).

FITS image), the observed time (YYYY/MM/DD HH:MM:SS UT format), exposure time (in seconds), the expected object magnitude (V when available), the telescope limiting magnitude (rough limit given the aperture, based on conventions mentioned in Section 2.1), filter and targeted object, angular distance to the field pointing (in degrees, followed by a percentage compared with the circular FOV), and a link to a plot showing the camera overlay (only for available mosaic cameras) marking the expected position of the searched object.

3.2. MASFO for fixed objects

In July 2017 we introduced the MASFO[27] tool to allow *Mega-Archive Search for Fixed Objects* given their J2000 RA and DEC coordinates. One or a few objects (entered in successive lines) could be searched given the approximating V magnitude of the object needed to compare with the limiting magnitude of the telescope (as given in Section 2). The output consists in the same columns as *Mega-Precovery*, including a link to the CCD overlay plot for most mosaic cameras included in *Mega-Archive*.

3.3. MASDS for double stars

Following some data mining work in double stars using the OmegaCam archive (Curelaru, 2017), in 2018 we implemented some tools for data mining the entire *Mega-Archive* for known double stars given by the Washington Double Stars Catalog (WDS) ID or the WDS Discoverer ID. This tool is named *Mega Archive Search of Double Stars* (acronym MASDS).[28] For observational planning we provide another tool named *WDS Filter Datamining*[29] which allows as input the sky area (α/δ box limits), stars' magnitudes, separation, existing observations and discoverer. Both these tools are presented fully in a recent paper (Curelaru et al., 2019).

3.4. FindCCD for moving and fixed objects

Survey telescopes and larger field mosaic cameras are increasing in number, and *Mega-Archive* presently includes about 20 such powerful imaging instruments, with Subaru-HSC (104 CCDs) and Blanco-DECam (62 CCDs) leading the list. Ideally data mining tools need to point their users directly to the exact CCDs of such a mosaic camera possibly holding their fixed or mobile targets.

In 2013 we provided *Find CCD Subaru*[30] to search for known NEAs in the SuprimeCam archive (Vaduvescu et al., 2017). In 2016 we extended this work to other few cameras, namely VST-OmegaCam, INT-WFC, VISTA-VIRCAM, CFHT-MegaCam, Blanco-DECam and Subaru-HSC. This tool is named *FindCCD*,[31] being developed for NEA searches. It queries the SkyBoT server[32] (Berthier et al., 2006) to calculate positions of the known NEAs located in the field, and the NEODyS-2 server[33] (Chesley and Milani, 1999) to overlay the uncertainty ellipses for poorly observed NEAs.

In March 2018 we released *FindCCD for Double Stars*[34] to identify the particular CCD of few major mosaic cameras possible to hold known double stars from the Washington Star Catalog (Curelaru et al., 2019).

In June 2018 we deployed *FindCCD for Fixed Objects*[35] to search for stars, galaxies or other fixed objects given as J2000 α/δ coordinates. Seven mosaic cameras are supported by August 2019 and we plan to extend the list soon.

4. Use case scenarios

We exemplify next few research case scenarios employing the EURONEAR data mining tools.

4.1. Precovery of one-opposition VIs and NEAs using Mega-Precovery

Orbits of poorly known NEAs and VIs could be greatly improved based on data mining, by searching precovery or recovery serendipitously apparitions at second or third oppositions.

First, we checked recently the three options of *Mega-Precovery* (from Designation, Orbit and Observations, using MPC orbits and observations database) by comparing the search of the one-opposition VI 2018 RQ4 in the Blanco-DECam archive. The results match each other (10 candidate images), proving the accuracy of

27 http://www.euronear.org/tools/masfo.php.
28 http://www.euronear.org/tools/dstars/masds.php.
29 http://www.euronear.org/tools/dstars/wdsfilter/wdsfilter.php.
30 http://www.euronear.org/tools/findsubaruccd.php.
31 http://www.euronear.org/tools/findccdaster.php.
32 http://vo.imcce.fr/webservices/skybot.
33 https://newton.spacedys.com/neodys.
34 http://www.euronear.org/tools/dstars/findccddstars.php.
35 http://www.euronear.org/tools/findccdfixed.php.

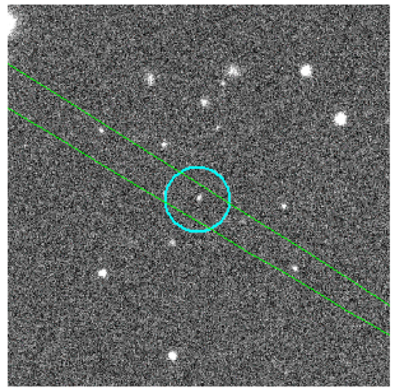

Fig. 6. Use case scenario of *Mega-Precovery from Designation* — $2' \times 2'$ crop from the SDSS image `frame-r-008096-5-0072.fits` taken at 15 Oct 2009 showing the precovery of the one-opposition NEA 2018 VN3 (discovered by EURONEAR in 2018, marked by cyan circle) inside the NEODyS uncertainty region (overlaid by an elongated green ellipse).

orbital calculation using *OrbFit*, and the resulting astrometry was reported to MPC part of another project (Vaduvescu et al., 2019).

Second, we used *Mega-Precovery from Designation* to improve orbits of the 8 one-opposition NEAs discovered by EURONEAR.[36] One of the relevant cases is 2018 VN3, which we precovered in the SDSS-DR9 image archive. Our search allowed to prolong the observed arc from 4 months to 20 years. We present below the strategy to obtain the astrometric data, and the case can serve as a method for similar data mining studies.

In the first step we considered the 15 Oct 2009 precovery (closest in time with the discovery arc), consisting in 3 candidate images (`frame-X-008096-5-0072.fits` where X stands for the 3 filters *rig* — in the y band the asteroid being invisible). According to NEODyS, for the observed date and time, the ephemeride was affected by about $8'$ uncertainty region. By overlaying in DS9 the uncertainty region as an ellipse over the FITS image, then blinking the 4 candidate images, we could spot easily the target $2.5'$ away from the predicted position. In Fig. 6 we give a crop of the `frame-r-008096-5-0072.fits` SDSS image showing the target (marked by cyan circle) inside the NEODyS uncertainty (overlaid by an elongated green ellipse).

In the second step, we considered the earlier 23 Nov 1998 precovery but we could not find the target in the candidate images, probably due to the larger uncertainty based on the actual 2018 four-month orbit. By considering *Mega-Precovery from Orbit* and the improved orbital elements fitted with *Find_Orb* software of Bill Gray,[37] we predicted some accurate 1998 position (affected only by a small sub-arcsec uncertainty), allowing to find the right frames holding the target (3 images `frame-X-000308-6-0152` with X standing for the same 3 filters *rig*). We reported both these 2009 and 1998 precoveries of 2018 VN3 to MPC.

4.2. Asteroid colors and lightcurves using Mega-Precovery

Image surveys or archives could be searched to derive photometric data (colors), lightcurves and spin periods of asteroids and NEAs using *Mega-Precovery*. We could exemplify with the NEA 2014 NL52 discovered by EURONEAR which showed signs of fast rotation, being followed with the INT and proving indeed very fast rotation ($P = 4.5$ min). Using *Mega-Precovery from Designation* and selecting only the INT-WFC archive, one could find 225 images ($V = 20.6$, TEXP $= 20$ s) observed at 30 July 2014 used to derive the published period (Vaduvescu et al., 2015).

Many candidate images could be found upon searching many NEAs using *Mega-Precovery* in the entire *Mega-Archive*, nevertheless few sets are actually feasible to derive lightcurves and rotation periods. Nevertheless, some faster (few minutes) or longer cadence surveys (few dozen minutes or hours) could be feasible for providing lightcurve data, such as the K2 Kepler mission (29.4 min cadence) which resulted in photometric constraints of 924 MBAs and rotational periods of 26 MBAs (Szabó et al., 2015, 2016). Asteroid colors could also provide spectral properties, and actually such studies were derived for NEAs and Mars-crossing asteroids using Sloan data (Carry et al., 2016). Moreover, sparse data (few days or weeks cadence) taken from ground or space could be data mined for physical properties and rotation periods of asteroids and other Solar System objects using one-band or multi-band observations, employing modern inversion algorithms (Cellino et al., 2009; Durech et al., 2009; Saha and Vivas, 2017; Muinonen et al., 2015), which could have applications to important space missions (such as Hipparcos, Gaia, Kepler, TESS, Plato) and to important future surveys (especially the LSST).

We plan to insert new collections and archives in our *Mega-Archive* and to tailor *Mega-Precovery* to allow extraction of photometric data and lightcurves of asteroids, comets and other objects.

4.3. Deep imaging of dwarf galaxies using MASFO

Extragalactic and stellar researchers could take advantage of *Mega-Archive* to search for fixed objects using the *MASFO* tool. One such Ph.D. student working in galaxies had to select recently near-infrared images taken with VISTA-VIRCAM survey for about 100 targets (Pinter 2019, paper in preparation). Thanks to the SQL architecture of the database, each query took less than 1s to identify typically 100–200 images (more than 1 TB data) which were downloaded and reduced using THELI,[38] in spite of the complicated geometry of the VISTA camera which includes 16 chips with large gaps between.

4.4. Predicting the exact location in mosaic cameras with FindCCD

While *Mega-Precovery* searches for moving or fixed objects falling inside the rectangular area delimited by the box defined by the margins in α and δ of the mosaic cameras, *FindCCD* predicts the chip in which the target is actually falling or its position inside any gap. Fig. 7 exemplifies of the search of a galaxy in one VIRCAM image which predicts the target located inside chip number 12.

36 http://www.euronear.org/discoveries.php.

37 https://www.projectpluto.com/find_orb.htm.

38 https://www.astro.uni-bonn.de/theli.

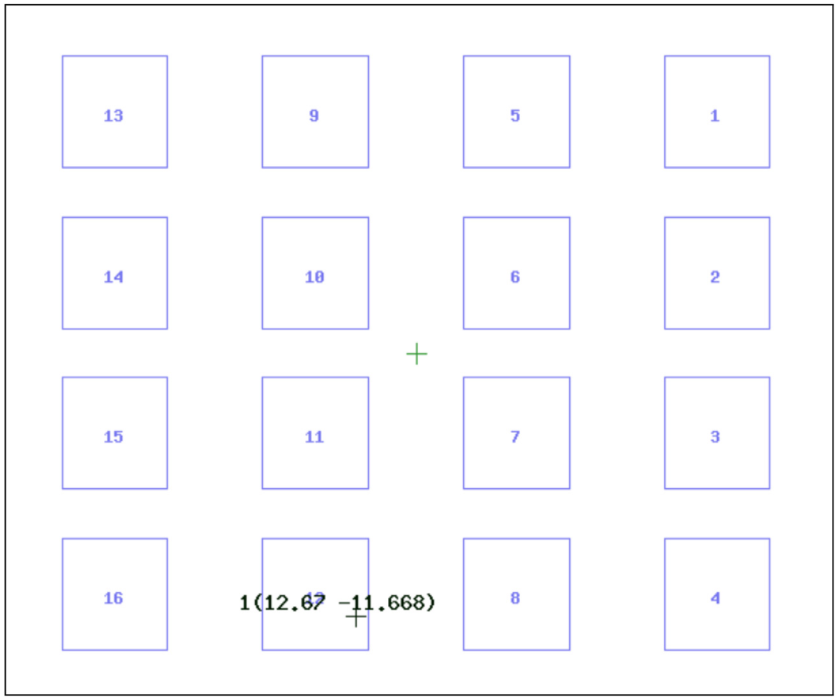

Fig. 7. *FindCCD for Fixed Objects* plot showing the CCD holding a galaxy of coordinates 12:40:03.1 -11:40:04 overlaid on the VISTA-VIRCAM mosaic image `VCAM.2013-01-21T07:52:15.676`.

4.5. Removing virtual impactors with Mega-Precovery (VIMP project)

There are about one thousand VIs known to date, and many of them have extremely poor orbits based on very small arcs observed following discovery (typically just a few days, sometimes only two nights of data). Such objects could be data mined using a tool similar to *Mega-Precovery* but focused on VIs whose ephemerides usually present very large uncertainties. In 2015 we started this project acronym *VIMP* by developing the Python code, and since Nov 2018 we datamine the entire Blanco-DECam archive with the aim to improve and eliminate VIs from the risk lists. To date, more than 30 VIs were precovered or recovered, and about half were removed from at least one risk list (Vaduvescu et al., 2019).[39] In Fig. 8 we give a use case of *VIMP* and *FindCCD* used to search the poorly observed VI 2015 BS516 in one Blanco-DECam image, possibly to fall in many CCDs which cover the large uncertainty region (in red) extracted from NEODyS.

5. Comparison with other services

5.1. SSOIS

Since 2012, the only service and database similar to *Mega-Precovery* has been the Canadian *Solar System Object Image Search* (SSOIS, known before as SSOS)[40] (Gwyn et al., 2012b,a). This service was focused first on the MegaCam, CFHT, Gemini and other image archives hold at CADC, and later added some ESO telescopes, Subaru and others).

As of 2017, the SSOIS database harvested meta-data for 32 million images based on 78 instruments. Nevertheless, their collection seems to be passed by *Mega-Archive* which includes to date 111 instruments and more than 15 million images. This apparent inconsistency could be explained because besides the raw images, SSOIS actually includes for some archives few levels of processed images which artificially seem to increase the numbers. This fact could be probed for example for the DECam outputs which is 4 times more numerous for SSOIS compared with *Mega-Precovery*. While SSOIS includes the NEAT, WISE and ALMA archives, *Mega-Archive* harvested two more collections, namely the Japanese SMOKA (18 instruments) and the LCOGT (19 instruments).

SSOIS introduced first a few options to search solar system objects (by designation, orbit, ephemerides and arc — first three being implemented by us later) and a few alternatives for ephemerides generators (CADC, JPL, MPC and Lowell — while we are using European services provided by IMCCE and NEODyS). Also, SSOIS was the first to provide direct links to resulted CADC images (thanks to images being stored locally).

While the SSOIS interface includes all 4 search options in the same page which becomes cumbersome to use, *Mega-Precovery* offers 3 separated pages to ease searches and eventually link to other internal or external services. The results (image counting) and response (execution time) appear similar for the SSOIS and our services, but sometimes SSOIS takes longer to query archives which include raw and calibrated images (for example, four times longer to query DECam).

Smaller PHAs and VIs could flyby closer to Earth and become visible only for a very short time windows (weeks or days), and in this sense *Mega-Precovery* implemented two search models: geocentric (allowing fast searches) and topocentric (running slower, based on ephemerides calculated to different locations which hold the instruments in the archives).

5.2. SkyMorph and TELARCHIVE

Early in the 90s, *SkyMorph*[41] has been another similar service developed by NASA (JPL, GSFC and STScI), allowing searches

39 http://www.euronear.org/publications/VIMP-EPSC-DPS-2019.pdf.

40 http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/ssois.

41 https://skys.gsfc.nasa.gov/cgi-bin/skymorph/mobs.pl.

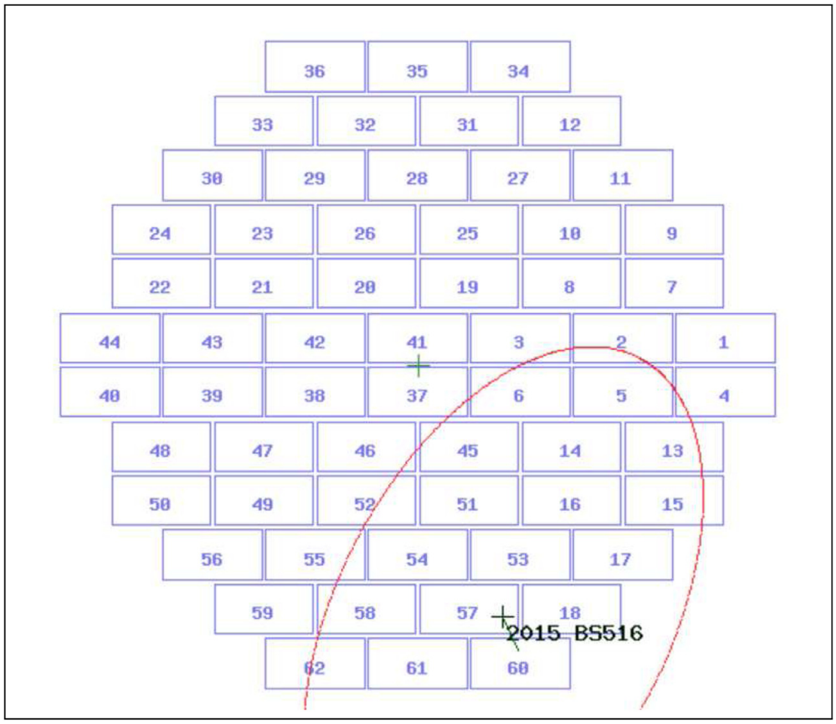

Fig. 8. *FindCCD* plot showing the Blanco-DECam overlay of the image `c4d150213_085354_ooi_g_v1` and the uncertainty position of the poorly observed VI 2015 BS516 (bordered in red) covering many CCDs.

based only on orbital elements (Lawrence et al., 1998; Pravdo et al., 1998). It was focused first on the NEAT survey archive (1996–2006), adding later a few plates archives DSS (1948–1956, 1972–1988), DSS2 (1984–1997), POSSI (1948–1956), USNO (1978–1999), and later some HST images and the Spacewatch survey (2003–2008). Since 2010 at least, *SkyMorph* does not appear to be maintained anymore. We plan to insert soon Pan-STARRS, NEAT, Spacewatch and the DSS in our *Mega-Archive*.

While SSOIS does not provide any tool to datamine fixed objects, since 2017 we take advantage of our *Mega-Archive* by offering *MASFO* and *MASDS* for data mining stars, galaxies and objects beyond our Solar system. Although most major observatories offer their interfaces for such searches only for their own archives, to our knowledge there is only one other service serving simultaneously many archives, namely TELARCHIVE software[42] (which needs local installation), besides *MASFO*.

6. Conclusions and recommendations

The EURONEAR *Mega-Archive* project started in 2010 by joining meta-data of the first three instrument archives (CFHTLS, ESO/MPG and ING/INT), becoming the first such server for data mining asteroids and NEAs. In the same year we introduced *Mega-Precovery* to search *Mega-Archive* images for one or a few known asteroids or comets given by designations.

During past years we added other archives, mainly based on six archive collections (CADC, ESO, ING, SMOKA, NVO and LCOGT) and by 2018 *Mega-Archive* has indexed 15 million images. In 2014 we implemented a crawler for automate query of five collections (except for NVO which does not allow programmatic queries) for daily update of the *Mega-Archive*. In 2016 we released the second version of *Mega-Archive* and *Mega-Precovery* which migrated to the new EURONEAR server and adopted the new SQL database architecture. Since Oct 2017 the third actual version of *Mega-Precovery* introduced two options for calculus of the ephemerides and three input options (search by designation, based on orbit, and observations).

During last few years we designed other tools aimed to take advantage of *Mega-Archive* for science other than NEAs. In 2017 we introduced MASFO tool to search the *Mega-Archive for Fixed Objects*. In 2018 we implemented the MASDS tool for *Mega-Archive Search for Double Stars*. To specify exactly the exact CCD of major mosaic cameras included in the *Mega-Archive*, in 2016 and 2018 we built the graphical tools *FindCCD* and *FindCCD for Fixed Objects* to overlay the moving or fixed targets over seven mosaic cameras, plotting also the uncertainty ellipse for poorly observed NEAs. We include a few use case scenarios and we compare our data mining tools with other few similar services, mainly the Canadian SSOIS. In the near future we plan to grow the *Mega-Archive* and to improve *Mega-Precovery* and other related data mining tools.

We have already entered into the big data epoch, where present and future surveys will provide huge amount of imaging data valuable for data mining and time-domain astronomy. In this sense, we recommend to the IAU, specifically Division B (Facilities, Technologies and Data Science) and Commissions B2 (Data and Documentation) and B3 (Astroinformatics and Astrostatistics) to adopt a common format and recommend to all astronomical observatories to use it for indexing and storing their science images, and make them available for programmable queries.

On a cosmic scale, NEAs pose real threat to mankind, and there are cases when faint or/and very fast moving virtual or imminent impactors could not be recovered even by the largest telescopes which typically are lacking immediate observing time. In such cases, tools able to datamine very recent archives could become crucial to precover such objects, improve their orbits, assess the risks and eventually eliminate the impact threats. In this sense, we recommend to astronomical observatories to index their

[42] http://www.mpe.mpg.de/~erwin/code.

entire available observations (completely up to date) even though some images are considered the proprietorship of a given PI who could decide to collaborate per request in making available such images, whenever such scenario will raise.

Acknowledgments

This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. This work is based on data obtained from the ESO Science Archive Facility. The SMOKA archive is operated by the Astronomy Data Center and the National Astronomical Observatory of Japan, being based on data collected at Subaru Telescope, Okayama Astrophysical Observatory, Kiso observatory (University of Tokyo) and Higashi-Hiroshima Observatory. This research uses services or data provided by the Science Data Archive at NOAO. NOAO is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the National Science Foundation. The first manual search of the NVO archive was performed by the EURONEAR collaborator Farid Char (University of Antofagasta, Chile). This paper makes use of data obtained from the Isaac Newton Group of Telescopes Archive which is maintained as part of the CASU Astronomical Data Centre at the Institute of Astronomy, Cambridge. The ING archive was built from the observing logs during a summer project by the student Vlad Tudor (former ING student and EURONEAR collaborator). This work makes use of the Science archive and observations from the LCOGT network. The first archive search and retrieval was performed by the EURONEAR collaborators Ioana and Adrian Stelea (Bucharest Astroclub, Romania). MASDS for Double Stars tool has made use of the Washington Double Star catalogs maintained at the U.S. Naval Observatory. This research has made use of Miriade, SkyBoT and SsODNet VO servers developed at IMCCE, *Observatoire de Paris*. We thank Jerome Berthier for his continuous support regarding the access to these services. MP acknowledges support from the AYA2015-67772-R (MINECO, Spain). We thank the anonymous referee whose suggestions helped us to improve the paper.

References

Albrecht, M.A., Benvenuti, P., 1994. The ESO science archive: A pilot implementation. In: Albrecht, M., Pasian, F. (Eds.), European Southern Observatory Conference and Workshop Proceedings. In: European Southern Observatory Conference and Workshop Proceedings, vol. 50, p. 15.

Baba, H., Yasuda, N., Ichikawa, S.-I., Yagi, M., Iwamoto, N., Takata, T., Horaguchi, T., Taga, M., Watanabe, M., Ozawa, T., Hamabe, M., 2002. Development of the subaru-mitaka-okayama-kiso archive system. In: Bohlender, D.A., Durand, D., Handley, T.H. (Eds.), Astronomical Data Analysis Software and Systems XI. In: Astronomical Society of the Pacific Conference Series, vol. 281, p. 298.

Berthier, J., Hestroffer, D., Carry, B., Vachier, F., Lainey, V., Emelyanov, N.V., Thuillot, W., Arlot, J.-E., Ephemerides Service, I., 2009. Miriade: A service for solar sytem objects ephemerides in the vo framework.. In: European Planetary Science Congress 2009. p. 676.

Berthier, J., Thuillot, W., Vachier, F., Arlot, J.E., Lainey, V., 2007. SsODNEt: the solar system object database network project. In: European Planetary Science Congress 2007. p. 488.

Berthier, J., Vachier, F., Thuillot, W., Fernique, P., Ochsenbein, F., Genova, F., Lainey, V., Arlot, J.-E., 2006. Skybot, a new VO service to identify solar system objects. In: Gabriel, C., Arviset, C., Ponz, D., Enrique, S. (Eds.), Astronomical Data Analysis Software and Systems XV. In: Astronomical Society of the Pacific Conference Series, vol. 351, p. 367.

Boattini, A., D'Abramo, G., Forti, G., Gal, R., 2001. The Arcetri NEO precovery program. Astron. Astrophys. 375, 293–307. doi:10.1051/0004-6361:20010825.

Carry, B., Solano, E., Eggl, S., DeMeo, F.E., 2016. Spectral properties of near-Earth and Mars-crossing asteroids using Sloan photometry. Icarus 268, 340–354. doi:10.1016/j.icarus.2015.12.047, arXiv:1601.02087.

Cellino, A., Hestroffer, D., Tanga, P., Mottola, S., Dell'Oro, A., 2009. Genetic inversion of sparse disk-integrated photometric data of asteroids: application to hipparcos data. Astron. Astrophys. 506 (2), 935–954. doi:10.1051/0004-6361/200912134.

Char, F., Vaduvescu, O., Popescu, M., 2013. Exploring the solar system with Mega-Precovery data mining server: strengthening our database and functionalities. Bol. Asoc. Argentina Astron. Plata Argentina 56, 411–414.

Chesley, S.R., Milani, A., 1999. Neodys: an online information system for near-Earth objects. In: Bulletin of the American Astronomical Society. In: Bulletin of the Astronomical Society, vol. 31, p. 1117, (1999).

Curelaru, L., 2017. Data mining for double stars on VLT survey telescope image archive. J. Double Star Obs. 13, 207–215.

Curelaru, L., Popescu, M., Vaduvescu, O., 2019. A set of data mining tools for double stars measurements using the EURONEAR Mega-Archive. J. Double Star Obs. 15, 3–14.

Durech, J., Kaasalainen, M., Warner, B.D., Fauerbach, M., Marks, S.A., Fauvaud, S., Fauvaud, M., Vugnon, J.M., Pilcher, F., Bernasconi, L., Behrend, R., 2009. Asteroid models from combined sparse and dense photometric data. Astron. Astrophys. 493 (1), 291–297. doi:10.1051/0004-6361:200810393.

Gwyn, S.D.J., 2015. Aggregating metadata from multiple archives: a non-VO approach. In: Taylor, A.R., Rosolowsky, E. (Eds.), Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV). In: Astronomical Society of the Pacific Conference Series, vol. 495, p. 413.

Gwyn, S.D.J., Hill, N., Kavelaars, J.J., 2012a. SSOS: A moving-object image search tool for asteroid precovery. Publ. Astron. Soc. Pac. 124, 579. doi:10.1086/666462, arXiv:1204.4182.

Gwyn, S., Hill, N., Kavelaars, J., 2012b. SSOS: A moving object image search tool for asteroid precovery at the CADC. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: Astronomical Society of the Pacific Conference Series, vol. 461, p. 789, arXiv:1111.3364.

Gwyn, S., Hill, N., Kavelaars, J., 2014. Solar system object image search: A precovery search engine. In: Muinonen, K., Penttilä, A., Granvik, M., Virkki, A., Fedorets, G., Wilkman, O., Kohout, T. (Eds.), Asteroids, Comets, Meteors 2014.

Gwyn, S., Hill, N., Kavelaars, J., 2015. Solar system object image search: A precovery search engine. IAU Gen. Assem. 22, 2223906.

Gwyn, S.D.J., Hill, N., Kavelaars, J., 2016. Solar system object image search: A precovery search engine. In: Chesley, S.R., Morbidelli, A., Jedicke, R., Farnocchia, D. (Eds.), Asteroids: New Observations, New Models. In: IAU Symposium, vol. 318, pp. 270–273. doi:10.1017/S1743921315008704.

Gwyn, S.D.J., Qi, J., 2012. MAASV: Megacam archival asteroid search and verification. In: Asteroids, Comets, Meteors 2012. In: LPI Contributions, vol. 1667, p. 6030.

Lawrence, K.J., Pravdo, S.H., Rabinowitz, D.L., Helin, E.F., McGlynn, T.A., Seelos, I., 1998. SkyMorph - an online program for accessing images from the near-Earth asteroid tracking program. In: Bulletin of the American Astronomical Society, Vol. 30. p. 1036.

Lister, T., Walker, Z., Ciardi, D., Gelino, C.R., Good, J., Laity, A., Swain, M., 2013. The LCOGT science archive and data pipeline. In: American Astronomical Society Meeting Abstracts #221. In: American Astronomical Society Meeting Abstracts, vol. 221, p. 240.18.

Mahlke, M., Bouy, H., Altieri, B., Verdoes Kleijn, G., Carry, B., Bertin, E., de Jong, J.T.A., Kuijken, K., McFarland, J., Valentijn, E., 2018. Mining the kilo-degree survey for solar system objects. Astron. Astrophys. 610, A21. doi:10.1051/0004-6361/201730924, arXiv:1711.02780.

Miller, C.J., Gasson, D., Fuentes, E., 2007. The NOAO NVO portal. Found. Trends Inf. Retr. 4 (3), 175, URL http://infolab.stanford.edu/~olston/publications/crawling_survey.pdf.

Muinonen, K., Wilkman, O., Cellino, A., Wang, X., Wang, Y., 2015. Asteroid lightcurve inversion with lommel-seeliger ellipsoids. Planet. Space Sci. 118, 227–241. doi:10.1016/j.pss.2015.09.005.

Olston, C., Najork, M., 2007. Web crawling. In: Shaw, R.A., Hill, F., Bell, D.J. (Eds.), Astronomical Data Analysis Software and Systems XVI. In: Astronomical Society of the Pacific Conference Series, vol. 376, p. 625.

Orbfit Consortium, 2011. OrbFit: Software to Determine Orbits of Asteroids. Astrophysics Source Code Library, arXiv:1106.015.

Penteado, P.F., Trilling, D., Szalay, A., Budavári, T., Fuentes, C., 2014. The planetary archive. In: AAS/Division for Planetary Sciences Meeting Abstracts #46. In: AAS/Division for Planetary Sciences Meeting Abstracts, vol. 46, p. 214.09.

Popescu, M., Licandro, J., Carvano, J.M., Stoicescu, R., de León, J., Morate, D., Boacă, I.L., Cristescu, C.P., 2018. Taxonomic classification of asteroids based on MOVIS near-infrared colors. Astron. Astrophys. 617, A12. doi:10.1051/0004-6361/201833023, arXiv:1807.00713.

Popescu, M., Licandro, J., Morate, D., de León, J., Nedelcu, D.A., Rebolo, R., McMahon, R.G., Gonzalez-Solares, E., Irwin, M., 2016. Near-infrared colors of minor planets recovered from VISTA-VHS survey (MOVIS). Astron. Astrophys. 591, A115. doi:10.1051/0004-6361/201628163, arXiv:1605.05594.

Popescu, M., Vaduvescu, O., Birlan, M., 2010. MEga-Precovery, a dedicated project for data mining worldwide image archives for poorly known asteroids. J. Sci. 2010 Inst. Méc. Céleste Calc. Éphémérides 95.

Popescu, M., Vaduvescu, O., Char, F., Curelaru, L., Euronear Team, 2014. Mega-precovery and data mining of near-Earth asteroids and other solar system objects. In: Muinonen, K., Penttilä, A., Granvik, M., Virkki, A., Fedorets, G., Wilkman, O., Kohout, T. (Eds.), Asteroids, Comets, Meteors 2014.

Pravdo, S.H., Helin, E.H., Rabinowitz, D.L., Lawrence, K.J., McGlynn, T.A., 1998. Skymorph: A 3-dimensional sky archive. In: American Astronomical Society Meeting Abstracts. In: American Astronomical Society Meeting Abstracts, vol. 192, 55.06.

Saha, A., Vivas, A.K., 2017. A hybrid algorithm for period analysis from multiband data with sparse and irregular sampling for arbitrary light-curve shapes. Astron. J. 154 (6), 231. doi:10.3847/1538-3881/aa8fd3, arXiv:1709.10156.

Solano, E., Rodrigo, C., Carry, B., Cortés-Contreras, M., 2018. SVO-Ast: A citizen-science project to identify NEAs and Mars crossers using the virtual observatory. In: European Planetary Science Congress, Vol. 12. EPSC2018–180.

Solano, E., Rodrigo, C., Pulido, R., Carry, B., 2014. Precovery of near-Earth asteroids by a citizen-science project of the Spanish Virtual Observatory. Astronom. Nachr. 335, 142. doi:10.1002/asna.201311888, arXiv:1302.5375.

Szabó, R., Pál, A., Sárneczky, K., Szabó, G.M., Molnár, L., Kiss, L.L., Hanyecz, O., Plachy, E., Kiss, C., 2016. Uninterrupted optical light curves of main-belt asteroids from the k2 mission. Astron. Astrophys. 596, A40. doi:10.1051/0004-6361/201629059, arXiv:1609.02759.

Szabó, R., Sárneczky, K., Szabó, G.M., Pál, A., Kiss, C.P., Csák, B., Illés, L., Rácz, G., Kiss, L.L., 2015. Main-belt asteroids in the K2 engineering field of view. Astron. J. 149 (3), 112. doi:10.1088/0004-6256/149/3/112, arXiv:1501.05967.

Vaduvescu, O., Conovici, M., Popescu, M., Sonka, A., Paraschiv, A., Lacatus, D., Tudorica, A., Hudin, L., Curelaru, L., Inceu, V., Zavoianu, D., Cornea, R., Toma, R., Asher, D.J., Hadnett, J., Cheallaigh, L.Ó., 2017. Data mining of near-Earth asteroids in the Subaru Suprime-Cam archive. Astronom. Nachr. 338, 527–535. doi:10.1002/asna.201713296, arXiv:1704.03936.

Vaduvescu, O., Curelaru, L., Birlan, M., Bocsa, G., Serbanescu, L., Tudorica, A., Berthier, J., 2009. EURONEAR: Data mining of asteroids and near Earth asteroids. Astronom. Nachr. 330, 698. doi:10.1002/asna.200811198, arXiv:0906.5030.

Vaduvescu, O., Hudin, L., Tudor, V., Char, F., Mocnik, T., Kwiatkowski, T., de Leon, J., Cabrera-Lavers, A., Alvarez, C., Popescu, M., Cornea, R., Díaz Alfaro, M., Ordonez-Etxeberria, I., Kamiński, K., Stecklum, B., Verdes-Montenegro, L., Sota, A., Casanova, V., Martin Ruiz, S., Duffard, R., Zamora, O., Gomez-Jimenez, M., Micheli, M., Koschny, D., Busch, M., Knofel, A., Schwab, E., Negueruela, I., Dhillon, V., Sahman, D., Marchant, J., Génova-Santos, R., Rubiño-Martín, J.A., Riddick, F.C., Mendez, J., Lopez-Martinez, F., Gänsicke, B.T., Hollands, M., Kong, A.K.H., Jin, R., Hidalgo, S., Murabito, S., Font, J., Bereciartua, A., Abe, L., Bendjoya, P., Rivet, J.P., Vernet, D., Mihalea, S., Inceu, V., Gajdos, S., Veres, P., Serra-Ricart, M., Abreu Rodriguez, D., 2015. First EURONEAR NEA discoveries from La Palma using the INT. Mon. Not. Roy. Astron. Soc. 449 (2), 1614–1624. doi:10.1093/mnras/stv266.

Vaduvescu, O., Popescu, M., Comsa, I., Paraschiv, A., Lacatus, D., Sonka, A., Tudorica, A., Birlan, M., Suciu, O., Char, F., Constantinescu, M., Badescu, T., Badea, M., Vidican, D., Opriseanu, C., 2013. Mining the ESO WFI and INT WFC archives for known near Earth asteroids. Mega-Precovery software. Astronom. Nachr. 334, 718–728. doi:10.1002/asna.201211720, arXiv:1301.6902.

Vaduvescu, O., Popescu, M., Curelaru, L., Danila, B., 2019. Eliminating virtual impactors with Mega-Precovery (VIMP). In: European Planetary Science Congress and Division of Planetary Society. In: European Planetary Science Congress and Division of Planetary Society.

Vaduvescu, O., Popescu, M., Tudorica, A., Conovici, M., Birlan, M., Badescu, T., Sonka, A., Lacatus, D., Paraschiv, A., Curelaru, L., Ursache, F., Vidican, D., Dascalu, M., Opriseanu, C., Grigore, V., 2012. Data mining the Suprime-Cam archive for near Earth asteroids. In: Asteroids, Comets, Meteors 2012. In: LPI Contributions, vol. 1667, p. 6235.

Vaduvescu, O., Tudorica, A., Birlan, M., Toma, R., Badea, M., Dumitru, D., Opriseanu, C., Vidican, D., 2011. Mining the CFHT Legacy survey for known near Earth asteroids. Astronom. Nachr., 332, 580–589. doi:10.1002/asna.201011550, arXiv:1107.2249.