

DATAMININGUL, O NOUĂ METODĂ DE A OBȚINE DATE ÎN ASTRONOMIA OBSERVAȚIONALĂ

Lucian CURELARU*, Ovidiu VĂDUVESCU**

Key words: datamining, EURONEAR, asteroids, doublestars, amateur astronomers.

Odată cu inventarea și evoluția CCD-urilor în anii '70, observatoarele astronomice au început să treacă rapid de la fotografia clasică la cea digitală, prima cameră CCD fiind folosită în 1974, pe un telescop de 20 cm, pentru a fotografia Luna¹. În paralel, dezvoltarea infrastructurii IT a făcut ca aceste imagini colectate de marile observatoare să nu mai fie păstrate în dulap, ele fiind stocate în arhive electronice gestionate prin intermediul bazelor de date. Grație internetului și, în special, a evoluției la internetul de bandă largă, un pas natural și necesar pentru accesul ușor al cercetătorilor la imagini a fost ca arhivele de imagini să fie făcute accesibile prin internet, în cadrul site-urilor observatoarelor sau pe site-uri comune ale unor grupuri de observatoare asociate. Astfel, majoritatea imaginilor din aceste arhive devin accesibile în mod gratuit oricui, la circa un an de la achiziție, ceea ce a creat șansa astronomilor amatori de a contribui la cercetarea științifică folosind imagini luate de marile telescoape ale lumii.

Această abordare a făcut ca, în prezent, să fie disponibile gratuit zeci de milioane de imagini luate cu telescoapele cele mai mari din lume în ultimele trei decenii. Firește că odată cu această disponibilitate a început să apară, deși destul de modest deocamdată, o nouă abordare în astronomie, numită „datamining”. Ideea de bază este simplă și anume: dacă dorim o imagine cu o anumită zonă de cer sau care să conțină un obiect cu anumite coordonate, dar nu suntem constrânși să observăm obiectul respectiv la o anumită dată, atunci vom putea căuta imagini care, deși au avut ca țintă alt obiect, ar putea să conțină, întâmplător, și obiectul căutat. Deși la prima vedere pare puțin probabil să se întâmple așa ceva, trebuie să ținem seama de numărul foarte mare de imagini conținute în arhive, cât și de faptul că unele instrumente au câmpuri foarte largi și sunt exploatate continuu (de multe ori, în modul “survey”) în toate nopțile cu cer utilizabil. În plus, mai există avantajul că aceste imagini sunt produse de instrumente cu dimensiuni și calități excepționale, amplasate în cele mai bune locații de pe glob, imaginile rezultate fiind în cele mai multe cazuri excelente din punct de vedere al seeingului (condițiile meteo) dar și al magnitudinii limită.

Faptul că putem utiliza imagini de calitate din aproape orice zonă a cerului, achiziționate la mai multe momente de timp, deschide oportunități nebănuite pentru obținerea unor rezultate de valoare, cu costuri extrem de mici (practic, zero costuri instrumentale). De amintit aici că, în general, construcția dar și întreținerea instrumentelor mari implică și costuri foarte mari, ceea ce face ca timpul de observație pe aceste instrumente să fie greu de obținut, traducându-se, în general, indirect, într-un cost mare per fiecare imagine produsă. Pe de altă parte, identificarea și utilizarea imaginilor existente implică doar muncă de căutare și, în unele cazuri, pentru un câștig de viteză de lucru, muncă de dezvoltare de scripturi sau software suport, ceea ce, însă, chiar tradus în costuri, duce la cifre incomparabil mai mici.

Ideea de a utiliza imagini deja existente, acolo unde e posibil, aduce pe lângă avantajul costului redus și avantajul observării la alte momente de timp a unor obiecte astronomice, aspect care poate fi important în anumite domenii ale astronomiei unde dorim să urmărim evoluția în timp a unui obiect sau fenomen. Firește că abordarea are și unele dezavantaje, cum ar fi riscul de a nu găsi o imagine potrivită care să acopere o anumită zonă și să corespundă tehnic (timp de expunere, filtre utilizate etc.) pentru un anumit scop, risc, însă, minimizat de numărul imens de mare de

* Astronom amator Brasov, Romania (<http://luci.astroclubul.org>), colaborator EURONEAR

** ING/IAC La Palma, Spain, UCV/Romania, fondator EURONEAR

¹ http://astro-canada.ca/les_cameras_ccd-ccd_cameras-eng

imagini disponibile, număr aflat într-o continuă creștere. În mod clar, un alt dezavantaj este imposibilitatea de a obține observații în prezent.

La capitolul avantaje, însă, am putea adăuga faptul că forma electronică a acestor arhive de imagini permite foarte ușor o abordare parțial sau total automatizată a căutărilor și chiar a unor prelucrări. Acest potențial a fost observat și exploatat în ultimii 12 ani și în cadrul proiectului de cercetare a asteroizilor EURONEAR (EUROPEAN Near Earth Asteroid Research), proces care comportă, pe de o parte, creșterea numărului imaginilor accesibile din arhive, iar pe de alta dezvoltarea de softuri și interfețe de căutare specifică. Pentru a ilustra mai bine ce implică datamining-ul, vă vom prezenta ca exemplu modul în care este exploatată în linii mari această abordare în cadrul proiectelor EURONEAR.

Una din primele idei de la care s-a pornit a fost identificarea într-o arhivă electronică a unui anumit instrument cu potențial, a imaginilor candidat care pot să conțină toți asteroizii NEAs (Near Earth Asteroids), cu scopul de a le îmbunătăți orbita prin adăugarea mai multor observații decât cele existente. Monitorizarea asteroizilor NEA este importantă în primul rând datorită riscului de impact cu Pământul prezentat de unii dintre aceștia. Deși la prima vedere această căutare într-o singură arhivă de imagini nu sună prea complicat, această sarcină poate implica enorm de multă muncă de rutină pentru fiecare rezultat, dacă se face manual, întrucât obiectul căutat se află într-o continuă și rapidă mișcare pe cer, ceea ce face necesară intersecția în timp și spațiu a imaginilor cu orbita asteroidului căutat. Este, însă, simplu de construit un script care să parcurgă arhiva de imagini și să calculeze efemerida asteroidului căutat pentru momentul de timp al fiecărei imagini, după care să verifice dacă rezultatul se suprapune cu imaginea. În felul acesta, munca de rutină urmează să fie făcută de calculator, care va rula câteva zeci sau sute de mii de efemeride pe care le va compara cu tot atâtea câmpuri existente în acea arhivă. Primele astfel de aplicații au fost dezvoltate de EURONEAR începând cu 2007, acestea fiind capabile să identifice peste 500 de NEAs în patru mari arhive, cu imagini luate de telescoapele și instrumentele MPG-WFI 2.2 m, INT-WFC 2.5 m, CFHT-Mega Cam 3.6 m și Subaru-Suprime Cam 8.2 m, proiecte care au generat patru publicații științifice internaționale, cu aportul a peste 20 de astronomi amatori din România.

Acest început, care s-a dovedit de succes, a generat imediat o serie de idei de dezvoltare, idei puse treptat în practică și care au făcut ca EURONEAR să ajungă în timp un dezvoltator de aplicații și pentru alte proiectele de datamining. Astfel, recent, am dezvoltat aplicații de datamining pe obiecte fixe, între care galaxii și stele duble, în unele cazuri existând și mici adaptări specifice domeniilor respective, cum ar fi o integrare cu Washington Double Star Catalogue pentru partea de stele duble.

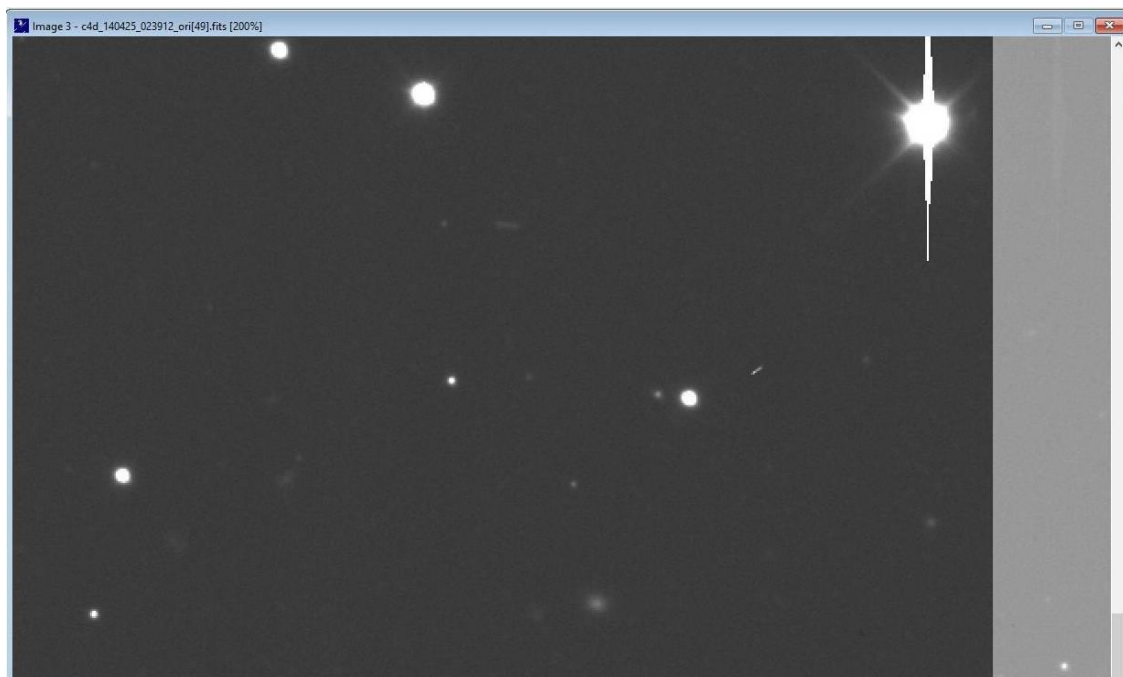
Revenind la domeniul NEAs, al doilea pas important a fost acela de a extinde aplicațiile de căutare pentru a accesa într-o singură căutare nu doar o singură arhivă, ci o colecție întreagă de arhive care să conțină în același format mai multe instrumente. Această idee este, aparent, simplu de implementat, însă a pus ceva probleme, în principal, generate de faptul că arhivele de imagini, aflate încă la începuturi, nu respectă nici un fel de standard, fiecare observator astronomic organizându-și datele cum a crezut de cuviință, în vreme ce IAU sau inițiativa globală VO (Virtual Observatory) încă nu au recomandat sau impus nici un astfel de standard. În lipsă de orice standard, EURONEAR a inventat pentru uz propriu un index de imagini cu format fix pentru toate arhivele, index care să conțină datele minime necesare despre imagini din orice arhivă disponibilă care să permită o căutare ușoară a unor perechi coordonate-timp. În aprilie 2009, acest index a fost numit EURONEAR Mega-Archive, și s-a sperat, la început, să includă cel puțin 1 milion de imagini. Acest index de imagini (meta-data) conține informații simple dar esențiale, cum ar fi data și lungimea expunerii pentru fiecare imagine, filtrele utilizate, coordonatele centrului acesteia, instrumentul cu care a fost realizată imaginea, arhiva unde se află și un ID după care poate fi identificată imaginea. Începând din 2014, am dezvoltat scripturi automate care interoghează zilnic un număr de arhive de imagini și actualizează acest index EURONEAR. Diversitatea arhivelor, însă, a făcut ca acest sistem de actualizare să necesite scripturi specifice pentru fiecare arhivă în parte, ceea ce înseamnă că indexul nu conține date de la absolut toate arhivele din lume ci doar de la cele pentru care a fost realizată o integrare. În Mega-Archive s-au integrat, în primul rând, arhivele conținute în câteva colecții mari

(europeană ESO, canadiană CADAC, americană NVO, japoneză SMOKA, europeană ING) dar și alte câteva instrumente sau survey-uri considerate importante pentru NEAs. În prezent, Mega-Archive conține 111 instrumente și un total de aproape 15 milioane de imagini (sau, mai corect, informații despre imagini) Firește că activitatea de extindere la cât mai multe arhive din lume este continuă, eficiența utilizării Mega-Archivei fiind proporțională cu numărul de imagini conținute.

Cum, însă, activitatea de datamining nu implică doar identificarea de imagini ci și analiza lor, în timp au fost dezvoltate și alte aplicații care să asiste astronomul în activitatea de datamining. Un exemplu este aplicația Find CCD pentru instrumentele mozaic active la câteva mari telescoape din lume, aplicație care ne ajută să identificăm exact sub-imaginea CCD în care se găsește obiectul căutat.

Toate aceste aplicații amintite mai sus sunt accesibile direct pe site-ul EURONEAR și pot fi utilizate gratuit².

Firește că odată cu dezvoltarea aplicațiilor au apărut și proiecte în care aceste aplicații au fost utilizate și au produs rezultate, în general, pe partea de astrometrie de asteroizi și, recent, pe partea de astrometrie de stele duble. Pe partea de asteroizi, dorim să amintim cele patru publicații științifice de datamining³ dar și utilizarea actuală a aplicației MASFO⁴ pentru studierea galaxiilor pitice eliptice, în cadrul unei teze de doctorat din Romania. Pe partea de stele duble, au fost desfășurate două proiecte care au produs astrometrie pentru peste 100 de stele duble⁵, din care majoritatea fiind duble insuficient observate⁶.



Imaginea nr. 1 - Asteroidul 2014 HJ198⁷

²<http://www.euronear.org/tools.php>

³<http://www.euronear.org/publications.php>

⁴<http://www.euronear.org/tools/masfo.php>

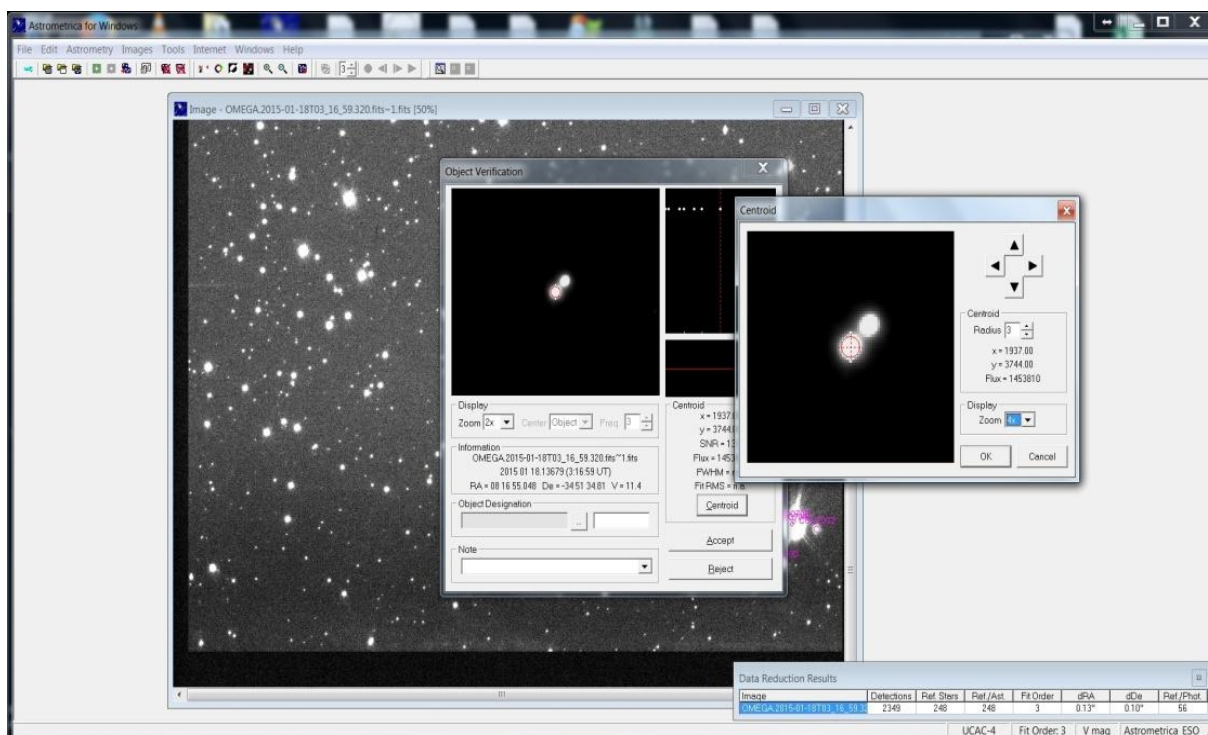
⁵http://www.jdso.org/volume13/number2/Curelaru_207-215.pdf

⁶ http://www.jdso.org/volume15/number1/Curelaru_3_14.pdf

⁷Dâra vizibilă în partea de mijloc-sus a imaginii este asteroidul 2014 HJ198 (magnitudine aparentă 20), un virtual impactor identificat în 2018 în cadrul proiectului de datamining EURONEAR VIMP, pe o imagine produsă de telescopul Victor Blanco (4 m) utilizând camera DECam (Dark Energy Camera). Imaginea alăturată a surprins întâmplător asteroidul, cu două zile înainte ca el să fie descoperit oficial pe alte imagini realizate tot de telescopul Blanco.

De asemenea, în prezent, este în derulare un proiect numit VIMP care identifică imagini ce ar putea produce noi poziții pentru asteroizi din categoria Virtual Impactorilor (VIs), definiți ca asteroizi NEA care au un risc nenul, deși extrem de mic, de a lovi în viitor Pământul. Firește că probabilitatea este mică, iar orbita multora dintre aceștia este destul de imprecisă. Tocmai din acest motiv identificarea de noi imagini de pe care să se extragă poziții noi este importantă pentru îmbunătățirea calității orbitelor și, implicit, clarificarea nivelului real de risc pe care îl au aceste obiecte.

De menționat, însă, faptul că datamining-ul este un domeniu destul de nou și surprinzător de puțin abordat în acest moment, dar în creștere în ultimul deceniu, în astrofizică, și care are cu siguranță un potențial foarte mare în viitorul apropiat. Aplicațiile și proiectele EURONEAR dar și ale altor foarte puțini „jucători” pe zona de datamining (cum ar fi proiectul canadian SSOIS⁸) sunt cu siguranță doar un început, existând extrem de multe idei de dezvoltare pe această direcție, idei care sperăm că atât noi cât și ceilalți „mineri” le vom pune în practică în viitorul apropiat.



Imaginea nr. 2 - Steaua dublă COO 70 AB⁹

Pe plan internațional, atât EURONEAR Mega-Archive cât și aplicațiile de datamining prezintă un mare potențial de a crește ca volum și flexibilitate în căutări, iar în acest sens invităm orice astronom amator interesat să ni se alăture în dezvoltarea de soft, includerea de noi arhive instrumentale, precum și în minarea efectivă de asteroizi apropiați de Pământ.

Iată, deci, că, datorită dezvoltării în zona electronicii, a IT-ului și a Internetului, am ajuns la un vis poate nesperat de înaintașii noștri, și chiar a co-autorilor acestui articol, aflați acum 20-30 de ani în faza de pionierat ca astronomi amatori. Astăzi putem face astronomie de calitate producând rezultate utile și de precizie mare, chiar și fără să avem un instrument. Cu un simplu calculator cu acces la internet și ceva cunoștințe de astronomie putem produce date valoroase de pe imagini de o

⁸ <http://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/en/ssois/>

⁹ Măsurarea stelei duble COO 70 AB cu componente egale, de magnitudine aparentă 9, având o separare de 2.25 secunde de arc, pe o imagine realizată de telescopul VLT Survey Telescope (2.6 m diametru) și camera Omega CAM. Imaginea realizată în 2015 a fost identificată în cadrul unui proiect de datamining de stele duble pe arhiva telescopului, proiect desfășurat la sfârșitul anului 2016.

calitate excepțională, luate cu cele mai importante instrumente existente în lume, contribuind cu o picătură în vastul ocean al cunoașterii umane.

Bibliografie:

- 1.Văduvescu, O. et al.: *Datamining of asteroids and near Earth asteroids*, Astronomische Nachrichten, 2009
- 2.Văduvescu, O. et al.: *Mining the CFHT Legacy Survey for known near Earth asteroids*, Astronomische Nachrichten, 2011
- 3.Văduvescu, O. et al.: *Mining the ESO WFI and INT WFC archives for known Near Earth Asteroids. Mega-Preccovery software*, Astronomische Nachrichten, August 2013
- 4.Văduvescu, O.; et al.: *Datamining of near-Earth asteroids in the Subaru Suprime-Cam archive*, Astronomische Nachrichten, June 2017
- 5.Curelaru, L.: *Datamining for Double Stars on VLT Survey Telescope Image Archive*, Journal of Double Stars Observer, April 2017
- 6.Curelaru, L., Popescu, M., Văduvescu, O.: *A Set of Datamining Tools for Double Stars Measurements Using the EURONEAR Mega-Archive*, Journal of Double Stars Observer, January 2019.

DATAMINING, A NEW METHOD OF COLECTING DATA IN OBSERVATIONAL ASTRONOMY

Datamining is a relatively new approach in astronomy, aiming to extract data from already existing images produced by the best telescopes and instruments in the world. Besides the first science aim for which they were initially acquired by only one principal investigator, the images produced often contain other important objects than the targeted ones, and for some of them is important to collect some data. This potential could be exploited also in a quasi-automatic manner by building helper targeted applications for the currently existing images archives. We present some examples of such approaches developed in the frame of EURONEAR project, among which near Earth asteroids, double stars, other fixed objects and Mega-Archive collection are briefly reminded here.

We invite any interested amateur astronomer to join our international projects in software development of our existing and new datamining tools, new instrument insertion in the Mega-Archive or the effective search of near Earth asteroids in the existing archives.