

Archivi astronomici di oggi...e che immaginiamo tra 20 anni?

Riccardo Smareglia

*INAF, Istituto Nazionale di Astrofisica
Osservatorio Astronomico di Trieste*



Abstract. In astronomia, scienza osservativa, l'esperimento (osservazione) è unico ed irripetibile in quanto legato al trascorrere del tempo. In questo contesto i dati acquisiti assumono un particolare valore soprattutto per ciò che può essere ricercato successivamente. Inoltre, la moderna astronomia si sta concentrando su grandi progetti (SKA, CTA, EST, ELT, LSST per fare degli esempi) il cui prodotto sarà una mole di dati difficile da gestire e soprattutto da archiviare. Il rischio è una perdita "di memoria" dell'evento, contrapposta ad una affannosa ricerca della cattura "dell'attimo fuggente". Sotto questo profilo, un'analisi della situazione esistente, di quella prevista nel prossimo decennio, ed una, non facile, estrapolazione dei possibili sviluppi nei 20 anni, verrà esposta ed analizzata.

Mentre la produzione e l'analisi dei dati provenienti da un singolo telescopio potrebbe non essere un problema, anche se si parla di decine, se non centinaia di TeraByte al giorno, le nuove frontiere dell'astronomia sono sempre più indirizzate alle cross-correlazioni di dati provenienti da più telescopi ed alle osservazioni multibanda utili per la creazione della SED (Spectral Energy Distribution) di ogni singolo oggetto. Questo porta ad avere la necessità di accedere a decine se non centinaia di archivi distribuiti sul pianeta. Affinché questo possa essere possibile due sono i punti essenziali che possono inficiare l'efficienza della ricerca scientifica: l'accesso ai dati ed il loro trasferimento. Relativamente al primo punto, ossia la facilità di accesso al dato e la sua interoperabilità, (senza soffermarci sul problema della qualità), un occhio di riguardo verrà dato all'Osservatorio Virtuale (VO), sviluppato all'interno dell'IVO (International Virtual Observatory Alliance). Il suo scopo è quello di sviluppare standard e strumenti che permettano l'interoperabilità dei dati e quindi la loro facile ricerca e confronto (cross-correlazioni e creazioni di SED per esempio). Questo sforzo potrebbe però essere pesantemente limitato dal collo di bottiglia della rete. Si vede quindi che la rete, e la sua efficienza, sono elementi essenziali che possono influenzare profondamente la qualità della scienza prodotta nei prossimi decenni. Quindi reti veloci che colleghino i generatori dei dati, il sistema di elaborazione degli stessi e l'osservatore, stanno diventando una priorità pari allo sviluppo degli strumenti stessi.

1. Astronomia: Scienza Osservativa!

In astronomia, scienza osservativa, l'esperimento (osservazione) è unico ed irripetibile in quanto legato al trascorrere del tempo. In questo contesto i dati acquisiti assumono un particolare valore soprattutto per ciò che può essere ricercato successivamente. È anche vero che, per fortuna, c'è tanto da osservare, ma in ogni caso anche se

nuovi strumenti e nuove tecnologie permetteranno di ottenere dati sempre più raffinati, l'utilizzo del dato "storico" permette di sco-

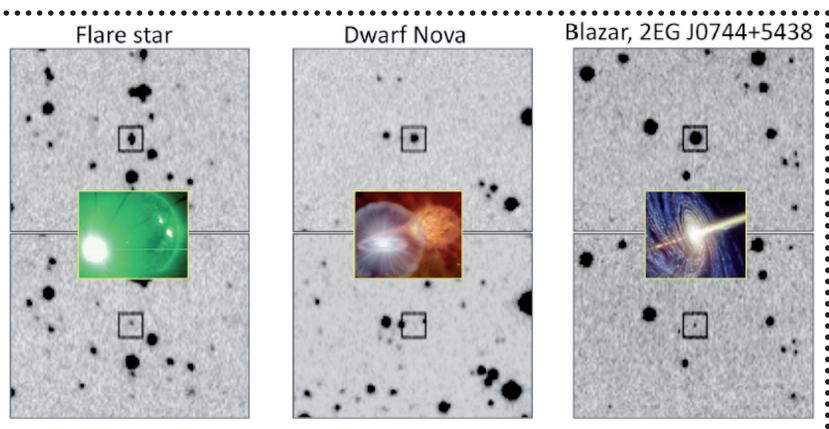


Fig. 1 Gli oggetti transienti hanno un aspetto molto simile nelle immagini astronomiche, ma possono rappresentare fenomeni fisici assai differenti

prire effetti come i “Transienti”. Si tratta di oggetti celesti che sono variabili e che quindi hanno una luminosità diversa a seconda del periodo in cui vengono osservati. Questa diversa luminosità può essere dovuta a diversi fenomeni fisici, come stelle variabili, doppie, stelle Flare, Nove nane, stelle Blazar, per non parlare delle famose SuperNovae.

È inoltre importante ricordarsi che i telescopi sono solitamente posizionati in luoghi remoti, se non su satelliti, mentre gli scienziati, gli archivi ed i centri calcolo sono di solito distribuiti su molte sedi. Solo l’INAF (Istituto Nazionale di Astrofisica) è distribuito su 19 sedi su 12 città oltre ad avere diverse *facility* osservative tra cui il telescopio nazionale Galileo (TNG), i radio telescopi di Medicina ed il radio telescopio SRT. Tutte le strutture italiane sono collegate tramite la rete GARR.

2. Astrofisica: una scienza ricca di dati

Come si è detto, per fortuna c’è molto da osservare e, tramite i telescopi terrestri e spaziali, a tutte le lunghezze d’onda. Negli ultimi anni si è passati dall’osservazione del singolo oggetto a grandi *Survey*. Una *Survey* è un’osservazione sistematica di tutto o parte della sfera celeste, ad una o più lunghezze d’onda, in modo da ottenere un’osservazione omogenea. Se a questo ottimizzazio-

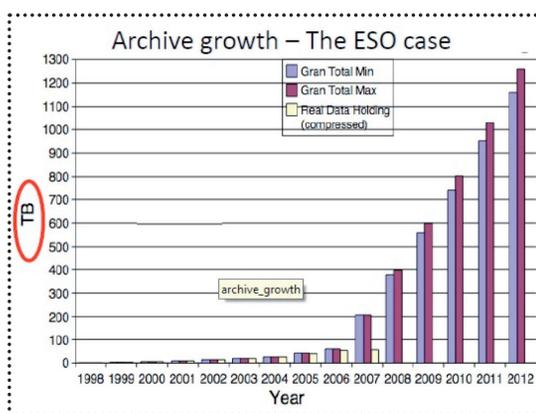


Fig. 2 L’occupazione dell’archivio ESO nel tempo

ne dell’osservazione aggiungiamo gli sviluppi tecnologici che ci permettono di lavorare con telescopi e strumentazioni sempre più grandi ed efficienti, otteniamo una quantità di dati complessi ed eterogenei che al momento hanno dimensioni intorno ai 100 Terabyte a survey e che possono creare cataloghi con centinaia di milioni di sorgenti. Ciò porta ad un notevole aumento dei dati archiviati, come ad esempio si può vedere nel grafico dell’archivio dell’ESO (European Southern Observatory) di cui l’Italia è membro.

Ma se già gli attuali telescopi distribuiti in giro per il mondo producono una quantità di dati difficili da gestire, gli astronomi stanno pensando a strumenti sempre più performanti. Un esempio può essere il telescopio LSST (Large Synoptic Survey Telescope) ed il CTA

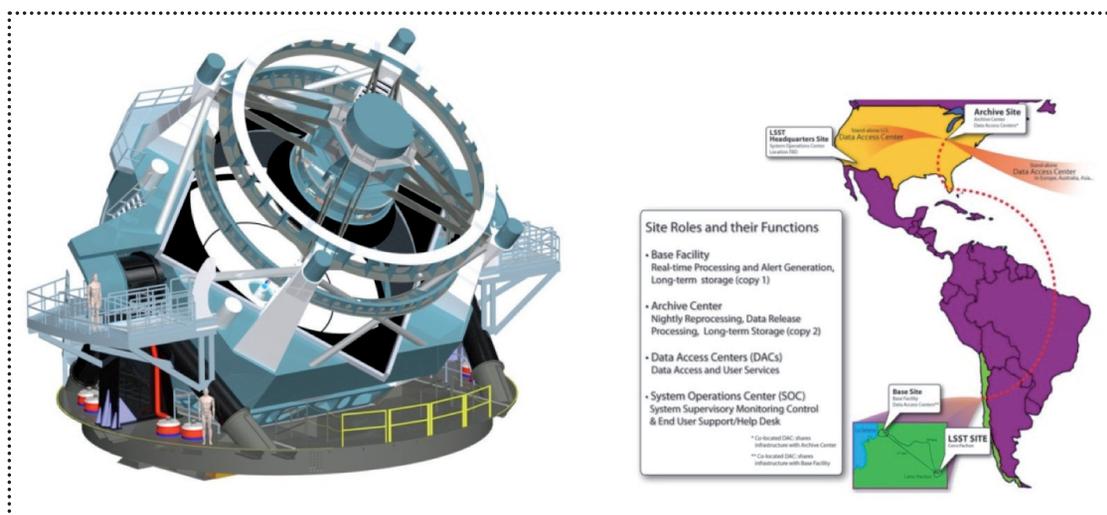


Fig. 3 Il telescopio LSST, la sua posizione (Cile) e la posizione dell’archivio (Ohio)

(Cherenkov Telescope Array).

Il LSST è un telescopio con uno specchio da 8.4 mt (come gli specchi dei telescopi dell'ESO) che però è dedicato ad eseguire una Survey di tutto il cielo, con lo scopo dichiarato di "visitare almeno 1000 volte la sfera celeste in 10 anni!". Si prevede che possa entrare in funzione tra il 2015 ed il 2020, con una "camera" che produrrà immagini di dimensioni di circa 6 Giga Pixel. Tutto ciò porterà ad avere un volume di dati prodotti pari a circa 20TB a notte, che in 10 anni porterà ad una quantità di dati *raw* (cioè dati non calibrati) pari a 60 PetaByte, ed un catalogo non inferiore ai 15 PB.

L'altro esempio riguarda il progetto CTA (Cherenkov Telescope Array), che vuole costruire la nuova generazione di strumenti per la ricerca di raggi gamma da alta energia con base a terra. Si parla di circa 150 telescopi (da 24, 12 e 6 metri) posizionati in due siti nei due emisferi per coprire tutta la sfera celeste. Rispetto ai precedenti telescopi per le alte energie dedicati a singoli esperimenti, il CTA sarà un osservatorio che servirà una vasta comunità di utenti. Infatti al momento nel progetto sono coinvolti oltre 800 scienziati di 25 paesi. Anche in questo caso il flusso di dati previsto è di 1-10 GB/secondo per totale di 1-10 PB/anno. In più l'analisi dei dati da questo tipo di telescopio necessita l'utilizzo di pesanti simulazioni Monte Carlo.

Un'altra fonte di grandi quantità di dati

sono anche le simulazioni teoriche: dal singolo evento (l'evoluzione di una supernova) alla evoluzione dell'universo su grande scala. Queste simulazioni richiedono spesso il confronto con i dati osservativi.

La ricerca astrofisica si sta evolvendo:

- nel passato le osservazioni erano "puntuali" ed omogenee, occupavano qualche MByte fino a qualche GByte e si indirizzavano verso pochi e "semplici" oggetti con pochi parametri per la loro identificazione;
- al presente si lavora su grandi survey omogenee (circa dai 10 ai 100 TB)che includono grandi esempi di oggetti (da 10^6 a 10^9 oggetti) con dozzine di parametri fisici;
- in un futuro, neanche tanto lontano, si parlerà di archiviare survey di cieli federati, a diverse lunghezze d'onda, dove i PetaByte saranno la base di ogni archivio e che implicheranno l'osservazione di tutto il cielo nel suo insieme con centinaia di parametri da confrontare per ogni singolo oggetto.

Ci si trova, e ci si troverà sempre più, di fronte ad uno scenario distribuito di generatori di dati (telescopi, simulazioni, centri di calcolo e scienziati distribuiti sull'intero globo terrestre, dove la scienza necessita di lavorare in modo interattivo su tutte le lunghezze d'onda utilizzabili (scienza *multi-wavelength*). Per far fronte a ciò già da 10 anni gli astronomi stanno lavorando al cosiddetto Osservatorio Virtuale (o Virtual Observatory: VO).

Il Virtual Observatory (VO) è una collezione

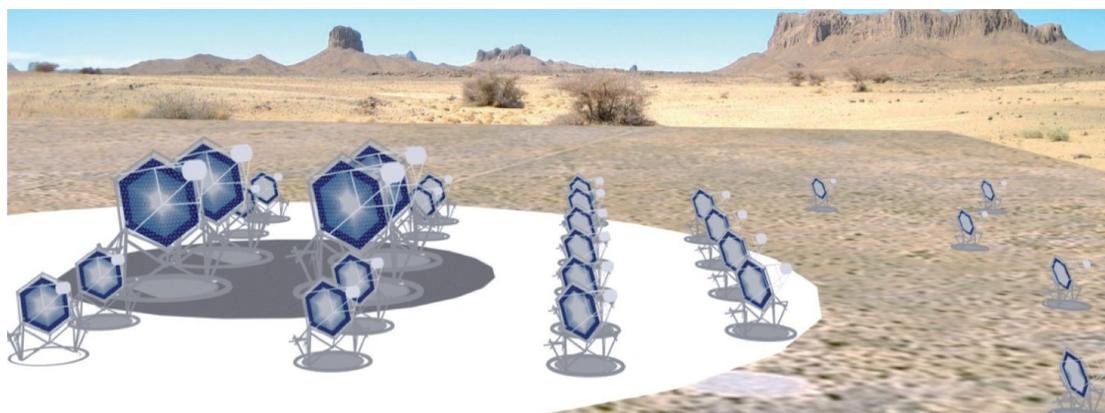


Fig. 4 Un disegno preliminare del CTA



Fig. 5 IVOA con i loghi dei membri

ne di archivi astronomici e strumenti software interoperabili che tramite Internet crea un ambiente di sviluppo scientifico nel quale sviluppare complessi programmi scientifici. Il VO viene sviluppato nei suoi vari aspetti da una alleanza internazionale IVOA (International Virtual Observatory Alliance) che raggruppa le varie organizzazioni nazionali che lavorano nel VO. In particolare l'Italia è rappresentata da VObs.it.

Il VO lo si può anche definire come “un nuovo tipo di organizzazione scientifica nell’era dell’abbondanza dell’informazione”. I suoi pregi sono:

- è intrinsecamente distribuito, ma web-centrico: Non esiste un unico *data center*, ma i dati rimangono dove ci sono le competenze tecniche e scientifiche per un loro efficiente primo utilizzo;
- al passo con lo sviluppo tecnologico: Il VO si basa sulle tecnologie informatiche e web più avanzate e punta decisamente verso il

web 2.0. Infatti un grosso sforzo all’interno del VO è stato fatto nella semantica, in modo da poter facilmente integrare modelli di dati eterogenei indipendentemente dalla disciplina e dalla lunghezza d’onda considerata;

- è “astronomia democratica”: viene rilasciato allo scienziato uno potente strumento di accesso ai dati, focalizzando il suo lavoro sulla ricerca di nuove frontiere da esplorare: identificare facilmente lunghezze d’onda in cui andare, aumentare le osservazioni e sviluppare nuovi strumenti o ricercare interazioni tra i dati esistenti tramite *data mining* etc.

Scopo dell’ IVOA è quello di sviluppare un sistema trasparente e di facile utilizzo da parte dell’utente, senza che questo debba preoccuparsi delle problematiche tecniche che lo compongono. Se si va a infatti a analizzare gli standard che l’IVOA ha sviluppato, e che sta ancora sviluppando, questi rappresentano un bosco di tecnologie e protocolli in cui solo gli sviluppatori esperti sanno destreggiarsi, con i quali la comunità astronomica non necessita di confrontarsi.

Proprio l’uso del VO potrà dare un’accelerazione allo sviluppo della scienza *multi-wavelength* tramite un facile accesso alla SED (Spectral Energy Distribution) cioè all’analisi delle caratteristiche di un oggetto su tutte le lunghezze d’onda.

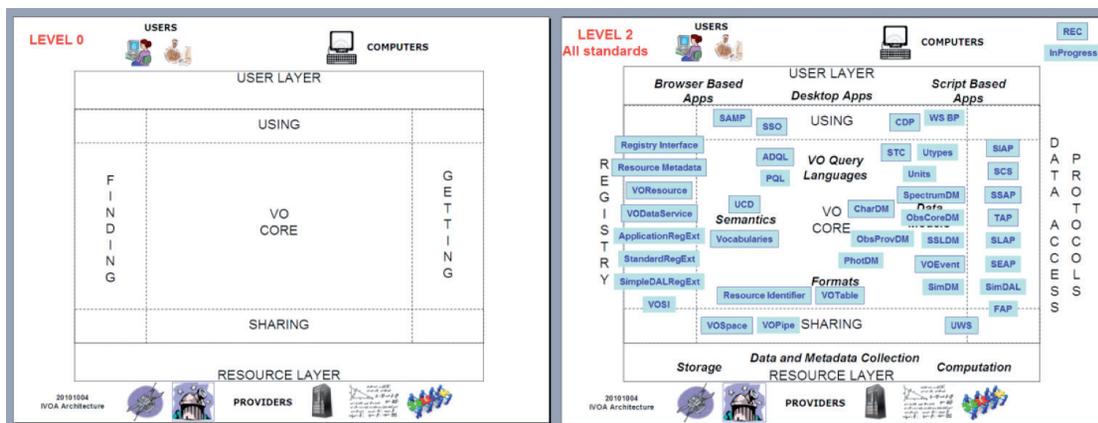


Fig. 6 Il Livello 0 e 2 del disegno del VO



Fig. 7 L'immagine di una galassia vista a diverse lunghezze d'onda

3. Conclusioni

Ci troviamo di fronte ad una sfida in cui scienziati, sistemi di calcolo e “fornitori di dati”(telescopi terrestri e centri di calcolo) saranno sempre più distribuiti e Petabyte (o ExaByte) di dati potranno essere prodotti senza il problema dello storage. Ad unire tutti questi attori ci sarà la rete, sempre più protagonista in questo nuovo scenario. Sarà la rete in grado di permetterci l'acquisizione e l'analisi di questo Tsunami di dati in arrivo?

Riferimenti bibliografici

- [1] INAF: <http://www.inaf.it>
- [2] LSST: <http://www.lsst.org>
- [3] CTA: <http://www.cta-observatory.org>
- [4] IVOA: <http://www.ivoa.net>
- [5] VObs.it: <http://vobs.astro.it>



Riccardo Smareglia

smareglia@oats.inaf.it

Laureato presso l'Università degli studi di Trieste nel 1987. Dal 1991 è dipendente presso l'Osservatorio Astronomico di Trieste.

La sua attività di ricerca si è sempre svolta nel campo del data handling e degli archivi di dati astronomici. Partecipa a diverse attività legate all'Osservatorio Virtuale (IVOA) sin dalla sua creazione nel 2000 e, dal 2006 al 2010, è stato responsabile italiano di due progetti finanziati dalla Comunità Europea (EURO-VO DCA ed EURO-VO AIDA). Dal 2004 è responsabile della creazione e gestione del “Centro per gli Archivi Astronomici” (IA2) il quale gestisce i dati dei telescopi ottici italiani o a partecipazione italiana come TNG ed LBT.