Combining ‘big science’ r sources and standards. The case of Gravitational Waves and GRB search

By Luisa De Biagi*, Fabio Garufi PhD**

** Consiglio Nazionale delle Ricerche (CNR)*
Informative and Documental Systems Office.
CNR Central Library ‘G. Marconi’
National Referring Centre for Grey Literature

***University of Naples “Federico II”, Dept. of Physics*

Scientific communities and OA data-sets

Large OA data-sets do represent an added value, granting opportunities for scientific discovery but…..

Understanding the context in which data-sets are collected it's crucial to:

- properly frame question and results;
- Elaborate right conclusions

So → scientists need to work closely with scientific data 'hunters' if they intend to use huge composite data-sets

Multi-messenger astrophysics

- Besides traditional astronomy, based on the detection of light coming from cosmic objects, other forms of radiation can give valuable information on the astrophysical sources.

- Different scientific communities look at the different forms of radiation coming from outer space: Cosmic rays, gamma rays, cosmic neutrinos, gravitational waves…

- The data produced by the various detectors are very heterogeneous, due to the peculiarity of both the detectors and the sources, and their integration can be troublesome.

The Gamma Ray Bursts (GRB)

- **Gamma Ray bursts (GRB) are the most energetic events in the known Universe. They were initially detected by military satellites monitoring USSR nuclear experiments, and nowadays are observed both by satellites and high quote ground detectors**.
- The frequency of these events is of the order of one per day.
- GRB sources are still unknown but it is believed they are produced in *Supernova* shock waves and coalescence of binary stars and black holes.
- The data flow of some MB/hour is peaked on the event trigger and, depending on the detector – satellite or ground detector – point in different location in space or integrate a large portion of visible sky.
- GRB data are stored in FITS (Flexible Image Transport System). It is primarily designed to store scientific data sets consisting of multidimensional arrays (images) and 2-dimensional tables organized into rows and columns of information

Gravitational Waves (GW)

- Gravitational Waves are produced, as a consequence of the General Theory of Relativity, by non spherically symmetric variation of mass.
- They are ripples in the space-time fabric that propagate at the speed of light and interact very weakly with matter, resulting in a very difficult detection. In fact the have not been detected directly, although indirect evidences exist.
- The sources of GW are rotating neutron stars, coalescing binary stars and merging black holes, that generate continuous waves, and supernova explosion that generate impulsive waves.
- Ground based GW detectors as VIRGO in Italy and the LIGO detectors in the USA, produce a continuous data flow of ~10 MB/s that are stored on the online data archive for a limited period and then on grid data centers.
- The GW data are organized with the LIGO/VIRGO Data Frame Format for interferometric gravitational wave detectors (IGWD). The predominant type of data stored in frames is time series data of arbitrary duration. It is possible, however, to encapsulate in frame structures other types of data, e.g., spectra, lists, vectors or arrays, etc.

Mixing GW and GRB data

- The joint analysis of GRB and GW data to search for possible correlation in the signals, due to a common origin implies the access and the management of both type of data.
- GRB satellite data are often publically available on e.g. the HEASARC data archive: http://heasarc.gsfc.nasa.gov/docs/archive.html.
- The NSF has asked LIGO Laboratory to prepare a Data Management Plan for the open release of LIGO data, including full strain data.
- This is envisioned to be similar to open data models like NASA's HEASARC; In response, LIGO Lab convened the LIGO Open Science Center (LOSC)

Mixing GW and GRB data

The use of the full Virgo data remains limited to:

- The Virgo collaboration, Other GW partners projects like LIGO, Detailed MOU with joint data analysis and publications…

→ ……..but work is in progress…

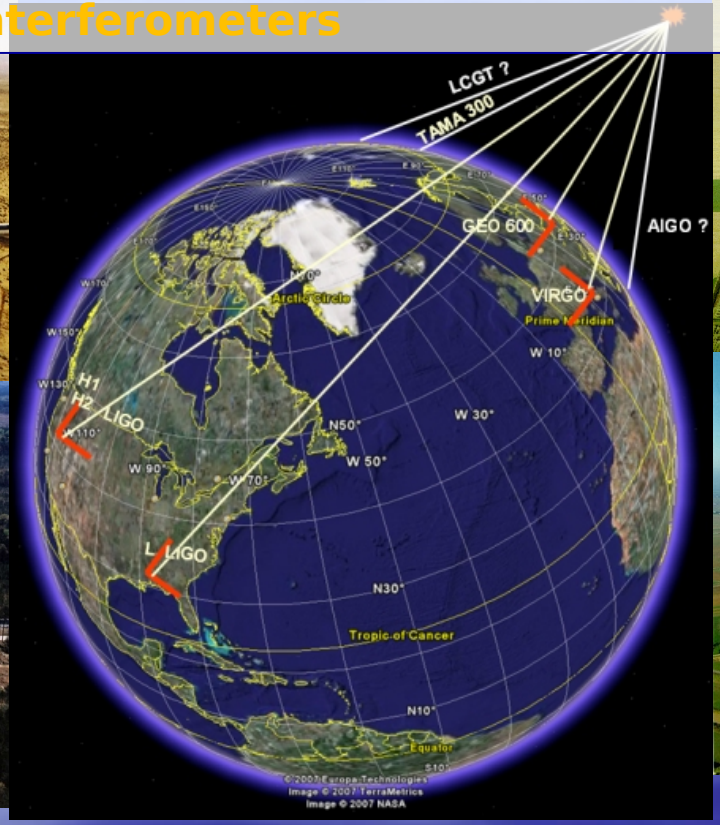
- GW and GRB data are usually sampled at very different rates. Meaningful GRB data integrate gamma flux on the detector with a time resolution of the order of the second (1 Hz) in different energy bands; present GW data are meaningful in the 5-1000 Hz range (0.2 – 0.001 s).

Worldwide Interferometer Network

A network of 4 (5) GW interferometers

LIGO – Hanford, WA

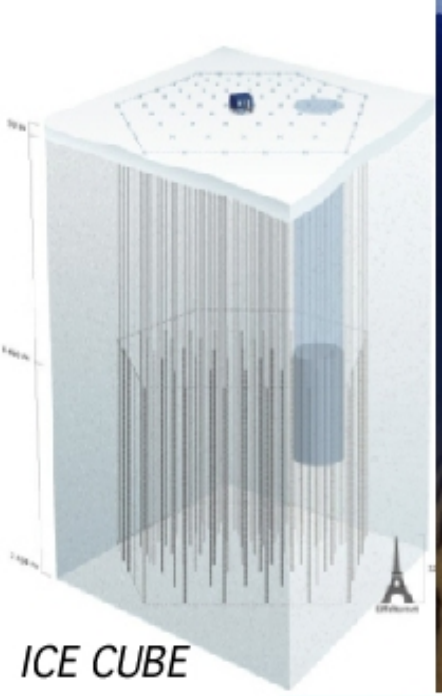
LIGO – Livingston, LA

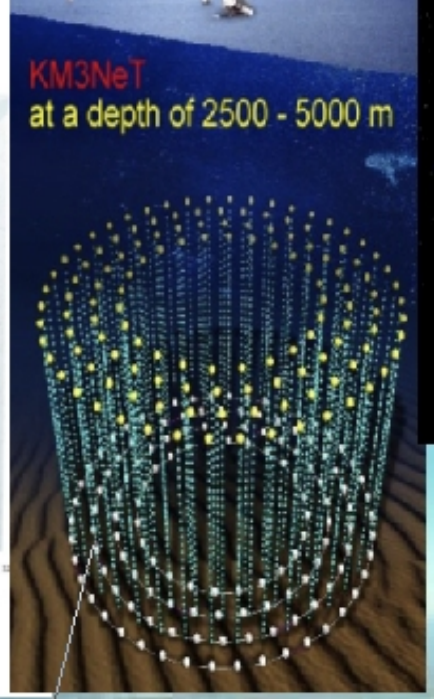


GEO600, Hannover, D

VIRGO, Pisa, Italy

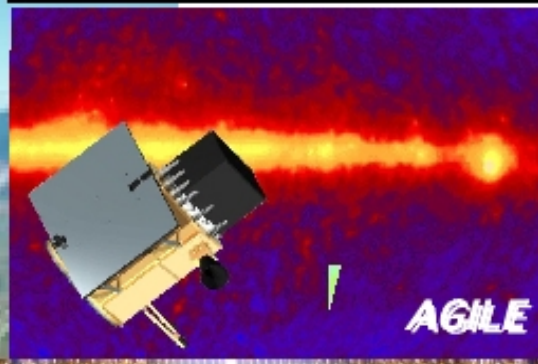
Virgo and LIGO Scientific Collaboration (LSC) signed a MoU for data exchange data analysis and publication policy. They act as a single multi-detector observatory.

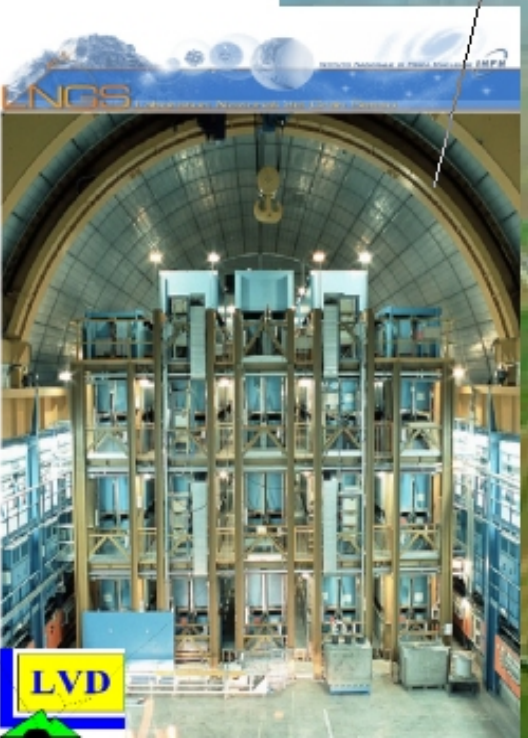

ICE CUBE


KM3NeT
at a depth of 2500 - 5000 m


GLAST - FERMI


SWIFT
UVOT
BAT
XRT


AGILE


LNGS
LVD


スーパーカミオカンデ
Super-Kamiokande

o tool for sharing astrophysical data

ArXiV. Org: OA repository provided by Cornell University Library gathering e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics → an electronically and 'interactive' action started in 1991, born from the old practice of distributing preprints of accepted articles in High energy Physics within HEP community by post.Now → more than 875000 e-papers/e-prints mainly in computer science statistics and maths → a well known repository where where researchers and scientists can deposit or find articles and papers published in Green OA. http://arxiv.org/year/astro-ph/13

N.B. → Detailed categories within astrophysics, updated entries in real time (new, recent, current month)

astro-ph **papers** *entries:* **10461** + *1428 in 2013 (update October 2013)*

Now improved with a mobile application for I-Phone granting access to the repository with a quick and easy search → extensive search engine (keywords or filters as title, author, abstract and ID.)

Other useful OA repositories

- CERN Document Server. An extensive database of bibliographic records and full-text documents, useful to people working in particle physics and related areas. Covers preprints, articles and other documents. http://cdsweb.cern.ch/
- IOP eprintweb.org, a free e-print service, provided by arXiv (for contents). Enhanced features in navigation, searching, personalization and presentation. Reference linking across the entire content, and enhanced searching on all key fields, including institutional address.
- KEK Information Service System for Preprints (KISS): KEK publications and other preprints stored in KEK library, in the fields of accelerators, high energy and nuclear physics, synchrotron radiation, health physics, and computer science. http://www-lib.kek.jp/KISS/kiss_prepri.html
- The SPIRES-HEP database, run by the Stanford Linear Accelerator Center(SLAC) http://www.slac.stanford.edu/spires/hep : the 1st to be served over the web (since 1991), comprehensively covering literature in HEP (High Energy Phys.) and supplying also special databases for jobs (employment database), Institutions and recent HEP experiments.

Main geospatial Big Data initiatives

Launched to increase capabilities to processing geospatial data:

- the European Commission's Big Data Public Private Forum http://big-project.eu/
- the US National Science Foundation's Big Data Science & Engineering (http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm)
- the US Office of Science and Technology Policy's (OSTP) Big Earth Data Initiative (BEDI) http://www.whitehouse.gov/blog/2013/04/19/taking-pulse-our-planet-new-strategy-earth-observations.
- recent ESA/ESRIN "Big Data from Space" event addressed challenges posed by policies for dissemination, data search, sharing, transfer, mining, analysis, fusion and visualization http://www.whitehouse.gov/blog/2013/04/19/taking-pulse-our-planet-new-strategy-earth-observations
- **OPENGEOSPATIAL Consortium: http://www.opengeospatial.org/**