

# **Where the Earth Meets the Sky**

**Thursday 27 May 2021 - Friday 28 May 2021**

**online**

## **Scientific Rationale**

Looking to the sky...

We are at a tipping point in Astrophysics where fundamentally new science will soon be enabled by extremely large telescope surveys that will capture the light of billions of galaxies near and far. For example the Large Synoptic Survey Telescope in Chile (starting operations in 2023) will collect 30 Tb of data per night in the visible spectrum. Compared to a standard extragalactic survey this is more than a factor 100 increase, which will make traditional software completely unfeasible. On top of that, recent work has shown that statistical tools commonly used to interpret observations are sub-optimal, e.g. they do not properly take into account the heterogeneity of data taken with very different facilities (in terms of wavelength coverage, image resolution, background noise, etc.). ...while keeping feet on the ground.

At the same time, biologists and other experts in medical science are witnessing an analog “big data” revolution in their research field. Health records are now electronically stored for an unprecedented number of patients. For example the Massachusetts Institute of Technology released to the scientific community more than 350,000 chest X-rays. Another noticeable example is the Demographic and Health Surveys Program that provides medical and socio-economic information of millions of individuals >90 countries. Nonetheless, the potential of these surveys is yet to be fully exploited as both data processing and interpretation are still plagued by major issues like e.g. the lack of interoperability between different data sets. Why Astrophysics and Biomedical Sciences together?

As illustrated above, researchers in the two domains face analogue obstacles. To overcome them, both communities are making a dramatic effort to catch up with advanced statistical and machine learning (ML) methods, or develop new ones. An interdisciplinary approach will increase the pace and extend the reach of such innovation. The proposed workshop will set a common ground for ideas exchange and training within the **four areas of interests** listed below.

## Images

The number of telescope images, as well as digital medical scans, is increasing exponentially. Visual classification (of either galaxy types or e.g. cardiothoracic indices) is a time-consuming task, and human eyes will soon be replaced by ML algorithms.

## Surveys

Extremely large surveys all pose the same challenges in terms of data mining. Another similarity is the expensive pre-processing that is required for cleaning and harmonizing heterogeneous pieces of information into the desired format.

## Models and Inference

A crucial task in any scientific domain is to devise theoretical models able to describe observed trends and correlations. The subsequent, and even harder step is to fit those models to real data. New software based e.g. on Bayesian inference can achieve such goals in a broad range of contexts, from reconstructing the formation history of galaxies to estimating variations in the mortality rate of African children.

## Literature

Scientific journals have the important role of recording and circulating results within the whole community. Unfortunately finding the desired information has become increasingly difficult as the number of published articles overcomes the time researchers have to read. On-line repositories like arXiv.org and PubMed can be explored with a new generation of search engines, or more elaborated algorithms for text mining that are even able to generate new hypotheses.