MSFR Research Training Network
The legacy tool: Contributions from the Mexico site

This is an executive summary of the proposed contributions from Mexico to the MSFR Network. We contribute (already with many successful existing colaborations) to the scientific topics that are the core of the network, as we will outline below, but we envisage that the major participation of our site will be in the construction of the LEGACY TOOL as we will detail in what follows. Our site has a vigorous postgraduate programme, that no doubt will provide the appropriate setting for the training of young researchers.

We apply for the status of a Third Node in the MSFR Network and to train one PhD student and 1 postdoc.

1 Individual participants

There are several important sites in Mexico doing research in Astrophysics. The participants of this proposal, listed below in alphabetical order, are mainly staff from the INAOE (Instituto Nacional de Astrofísica, Optica y Electrónica, Tonantzintla, Puebla) and the IA-UNAM (Instituto de Astronomía of the Universidad Autónoma de México, Mexico City and Ensenada in Baja California), with some members also from the Universidad de las Americas and the Universidad Autónoma de Puebla (UDLA and BUAP, both in Puebla) and the Universidad de Guanajuato.

INAOE Astrophysics members are:

Itziar Aretxaga `itziar@inaoep.mx`
Compact supernovae; star formation in powerful radio galaxies; dust enshrouded massive star forming galaxies at low and high-z
Percentage of research time to be invested: **40%**

Alberto Carramiñana `alberto@inaoep.mx`
Pulsars and compact stars: high energy emission and evolution; gamma-ray bursts.
Percentage of research time to be invested: **10%**

Diego Malquori `malquori@inaoep.mx`
High energy processes in galaxies and handling of megadatasets.
Percentage of research time to be invested: **30%**

Y. Divakara Mayya `ydm@inaoep.mx`
Stellar populations in regions of recent star formation
Percentage of research time to be invested: **40%**

Guillermo Tenorio-Tagle `gtt@inaoep.mx`
Hydrodynamics and chemistry of violent star forming regions.
Percentage of research time to be invested: **30%**

Elena Terlevich (P.I. of the Node); `eterlevi@inaoep.mx`
Chemistry of regions of violent star formation; stellar fabric in galactic nuclei; interaction

between violent star formation, massive stars and the interstellar medium.
Percentage of research time to be invested: **70%**

Roberto Terlevich `rjt@inaoep.mx`
Violent star formation, Star formation in elliptical galaxies, Starburst-AGN connection, Observational cosmology, Fundamental planes, Supervised and unsupervised analysis of mega-datasets.
Percentage of research time to be invested: **70%**

Sergiy Silich `silich@inaoep.mx`
Structure, evolution, and energy budget of the interstellar medium powered by the action of massive stars, and star clusters. Mechanical feedback.
Percentage of research time to be invested: **40%**

The Computing Sciences group members are:

Aurelio López López Head of the Computing Science Department; `allopez@haro.inaoep.mx`
Natural language processing, computer perception, signal processing, information processing, speach recognition, encrypting
Percentage of research time to be invested: **50%**

Olac Fuentes Chavez `fuentes@inaoep.mx`
Pattern recognition and automated learning, image analysis
Percentage of research time to be invested: **60%**

Jesus A. González Bernal `jagonzalez@inaoep.mx`
Automated learning, data mining
Percentage of research time to be invested: **15%**

Jaime Muñoz Arteaga `jaime@inaoep.mx`
Man-machine interface, computer networks, databases.
Percentage of research time to be invested: **15%**

Gustavo Rodríguez Gómez `grodrig@inaoep.mx`
Numerical methods, computational methods
Percentage of research time to be invested: **15%**

The IA-UNAM (Mexico City and Ensenada) members are:

Luc Binette `binette@astroscu.unam.mx`
Processes of photoionization in high-z nebulae
Percentage of research time to be invested: **15%**

Jose Franco `pepe@astroscu.unam.mx`
Magneto-hydrodynamical star formation models, the interstellar medium
Percentage of research time to be invested: **10%**

Guillermo Garcia-Segura ggs@astrosen.unam.mx
Magneto-hydrodynamical star formation models, the interstellar medium
Percentage of research time to be invested: **20%**

Christophe Morisset Morisset@AstroScu.UNAM.mx
3D modelling of photoionization regions and massive stars.
Percentage of research time to be invested: **40%**

Margarita Rosado margarit@astroscu.unam.mx
Galactic HII regions, bubbles and superbubbles
Percentage of research time to be invested: **15%**

Mauricio Tapia mt@astrosen.unam.mx
Processes of star formation; infrared observations
Percentage of research time to be invested: **15%**

Computing scientists from Puebla Universities UDLA and BUAP:

Gerardo Ayala to be completed
Data mining
Percentage of research time to be invested: **15%**

Jose Luis to be completed
Database interfaces
Percentage of research time to be invested: **20%**

Genoveva to be completed
Database interfaces
Percentage of research time to be invested: **15%**

David Pinto to be completed
to be completed
Percentage of research time to be invested: **5%**

And from the University of Guanajuato:

Heinz Andernach heinz@abell.astro.ugto.mx
Megadatabases.
Percentage of research time to be invested: **15%**

2 Project Objectives

This section provides an outline for our highest priority project on which we are going to be working in the framework of the MSFR network.

We understand that in developing such a project is where our personal contribution to the network could be the greatest. Also, where the synergism with the rest of the network activities and

participants will be at its peak, and we will be able to produce en effect greater than the individual effects. The project consists of two parts: the first one is the managing of megadatasets for the astronomical comunity, the second one is the production and implementation of a tool that could be used by our colleagues to interpret the physics behind the data features (either images, spectra or modelled/simulated data): the *Legacy Tool*, available from a readily accessible site.

2.1 Megadatasets

In the last few years we have witnessed the coming of age of observational cosmology and a world-wide boom in projects aimed to the understanding of the evolution of the universe. Mexico with the Large Millimeter Telescope LMT/GTM that is being built in collaboration with UMASS, will be a key contributor in this area. At present most of the largest observational projects involve the collection of enormous data sets, but astronomers are presently inadequately prepared to cope with their analysis.

Within our group we will be able to provide the infrastructure to assemble interdisciplinary teams with the breath of expertise in computation, astrophysics, statistics, dynamics and visualization to exploit the anticipated massive data flow from future cosmology-related experiments, both ground and space based.

Recently there has been a huge increase in the amount of data available for astronomical research, due to a large number of new facilities coupled to technological advances in sensors and storage devices and the automation of data acquisition. As a consequence, astronomers will in the next few years generate calibrated data more rapidly than they can manage or analyze reliably.

New large surveys like the 2 degree Field (2dF), the Sloan digital sky survey (SDSS), 2MASS, VIRMOS and DEEP2 are revolutionizing the way we do astronomy, by adding enormous quantities of high quality data to that already available for research. To this sea of data, we should add the data bases of the recent generation of X-ray, optical, infrared and radio(mm) large telescopes with multiplexing capabilities that will also produce data in vast quantities. The wide-wavelength nature of these databases will provide unprecedented challenges in unifying optical, infrared, millimetre and X-ray observations, utilizing resources that include ground-based and space telescopes, and radio facilities such as the LMT/GTM (the large millimetre wave telescope presently under construction by INAOE and UMASS, the University of Massachusetts, in Cerro La Negra, Mexico, at 4600 meters elevation).

One cannot overemphasize the importance and complexity of this challenge. In the next five years the data available to research in databases will increase to the level of having thousands of parameters on tens of millions of astronomical objects,. i.e. MEGADATASETS. This complexity, will be compounded by the presence of measurement errors, biases and tendencies in the data. Another important dificulty is the lack of standardization across the different data bases. Different data sets have their own access sofware making very complicated the cross-correlation between the different datasets.

To advance under the current and continously growing data overload, there exists a need for applications to automatically and intelligently manage, query, visualize and analyze the whole space and variety of large data sets. This is even more important for countries like Mexico with restricted access to large observational facilities.

INAOE is well placed to play an important role in this new field. The size, quality and proximity of the Astrophysics and Computation Departments is rather unique among Astronomy departments and allows the fast and efficient development of specific interdisciplinary teams like ours. The collaboration with the other participant nodes will increase access to the many British and European databases plus expertise in cosmology.

Regarding teaching and training, the mixture of computation, statistics, visualization and scientific rigour, makes a powerful combination with wide applications across society, as well as for pure research.

2.1.1 Aims

We have selected a learning path that, starting from relatively modest aims points to ambitious targets.

We are interested in developing and comparing different statistical methods to analyze multiple Megadatasets.

Our main aims are multiple and include:

- Development of interface between the user and the different datasets based on remote analysis.
- Parameterization of features and division into sample and sub-sample outputs, approaching 100 gigabytes each.
- Development of visualisation tools.
- Unsupervised classification and determination of the different numbers of object classes present in the data.
- Searches for unusual objects, including the ability to detect new classes of objects.
- Search for regularities or correlations in the data.
- Support for supervised analysis.
- Support for analysis of simulated data.

The results of this analysis will as well provide new datasets for detailed follow-up work. In this respect, we are particularly interested in those topics that form the core of the research in the network, like the evolution of the properties of emission line galaxies, the physics of star formation in them, and their use to map the evolution and geometry of the Universe.

2.2 The legacy tool

The results of the multivariate analysis of the SDSS first release data will be used to select interesting samples of emission line galaxies to study :

- The ages of starbursts galaxies
- The chemical abundance distribution and its evolution with lookback time.
- Their use as distant estimators at high redshifts and the determination of cosmological parameters.

The follow-up studies will be carried out using large facilities like VLT and Gemini, and will consist of high S/N multislit spectroscopy of a moderate number of objects selected to begin with from the 2dF and SDSS surveys. Because these surveys are multi-fibre surveys in 3/4 meter class telescopes, the quality of the data will be substantially increased by observing with slits and in 8m class telescopes.

Different software packages that have been developed by members of the network (as well as others from the literature) exist and are (or can be made) available for the comunity to use, although they are in different environments, use different lenguages, scope of the distribution, state of documentation, etc.

INAOE is a multidisciplinary research institute (Astrophysics, Electronics, Optics and Computational sciences) which allows very close interbreeding between the different disciplines, very seldom found in other Astronomy Departments. In particular, the Megadatabase group already existing at the Institute (http://ccc.inaoep.mx/megadatasets.html) is especially qualified to perform the delicate job of putting together, understanding, documenting, homogenising and maintaining the *legacy tool* for the general use of the astronomical community as a whole.

2.3 Training

Both INAOE and the UNAM have a very vigorous postgraduate training programme. The nature of the INAOE is such that we offer programmes for postgraduate courses (Masters and PhDs) not only in Astrophysics but also in Optics, in Electronics and in Computational Sciences. In the last period (comprising 1999 through March 2003) we have graduated 270 postgraduate students, of which 25 were in Astrophysics. Computing Sciences is a new discipline at our Institute, and we have currently 53 postgraduate students working for their degrees in Astrophysics and in Computing Sciences at different levels in their courses.

The Institute of Astronomy of the National University (IA-UNAM) forms an equivalent number of Masters and PhDs in Astrophysics (some 10 per year) both in their headquarters at Mexico City and in the outpost at Ensenada (Baja California).

We already have three PhD students and four Master students working within our group of megadatasets. Although only half of them are Astrophysics students proper (the rest are Computing sciences students), their thesis are all related to Astrophysics topics. Several more students and postdocs are expected to be involved from the early stages, we hope that some will be co-directed by our group and by members of the collaboration that are the authors of the different existing codes.

2.4 Existing collaborations

Our proponents have long standing successful collaborations with several members of the network, some of them lead by the mexican colleagues, several leading to co-directed Thesis.

The sites with existing collaborations with INAOE and IA-UNAM are: IAC, IAA and Madrid (in Spain), Padova and Bologna (in Italy), IAP and Toulouse (in France), Geneva in (Switzerland), Cambridge (in the UK) to name a few.

2.5 Management experience

We have ample experience in the participation in, or the management of, large collaborations, for instance through the organization of the yearly Advanced Workshop in Astrophysics (Guillermo Haro) at INAOE since 1996, an event that attracts some 50 world experts in a field (generally related to violent star formation, galactic evolution, ISM, activity in galaxies) that live at the Institute for about a month, in a brain-storming effort to solve some pressing problems in the topic. At the end of the workshop, an international Conference (or an Advanced School) are organized on the same topic.

We are developing the Mexican Virtual Observatory, to form part of IVOA (International Virtual Observatory Alliance).

Many of the Mexican participants have been members of several pan- European and bi-lateral collaborations, and even invited members to some National collaboration (like Estallidos, in Spain).