European Union's FP7 Programme
DG Connect
Directorate C: Excellence in Science
Unit C1: e-Infrastructure

Deliverable D6.6

Report on Large facilities and Remote Instrumentation

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

Periodical Progress Report

ELCIRA Deliverable: CODE – Title of the document

Document Full Name	**ELCIRA DO6.6 Report on Large facilities and Remote Instrumentation**
Date	**2014-07-23**
Activity	**6. Fostering the Collaboration between EU and LA Researchers (RedCLARA)**
Lead Partner	**RedCLARA**
Document status	**Ready**
Classification Attribute	**Public**
Document link	

Abstract: This document reports the large facilities and remote instrumentation inventoried in Latin America and used in behalf of the development and reinforcement of the science, the research and the academy of the latinamerican region and all over the world.

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

COPYRIGHT NOTICE

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

DELIVERABLE ROUTE

	Name	Member/Activity	Date	Responsible
From	Herbert Hoeger	SCALAC/ Team member	July 23, 2014	RedCLARA
	Emilia Serafín	RedCLARA/Communities	July 23, 2014	RedCLARA
Revised by	Luis Núñez	RedCLARA/ARM	July 23, 2014	RedCLARA
Aproved by	Florencio Utreras	RedCLARA/CEO	July 24, 2014	RedCLARA

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

TABLE OF CONTENTS

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

1.1 INTRODUCTION

Data-intensive scientific analysis is a completely new way of doing science. How to deal with large datasets is still in evolution and has a long way to go. All disciplines, either physical, life sciences and humanities are becoming increasingly data-driven and data intensive. This is happening mainly due to technological advances in information networks, computing capacity, big instruments, penetration by sensors in all areas, as well as increasingly collaborations among researchers. Big Data Science requires interdisciplinary skills in which computer scientists, statisticians and other experts combine their knowledge to create new techniques, tools and methodologies, shifting from a hypothesis-driven to a data-driven way to analyse the increasing datasets. Astronomy is the most pioneer data-driven science and its communities are early adopters and creators of multiple discovery environment incorporating strategies and tools to manipulate and analyse huge amounts of data.

Large data productions are usually carried out by global collaborations, i.e., multinational science groups that generate large volumes of data, geographically distributed and maintained only during the project life cycle. Most of these data is never published and, when the collaborations end, many is lost or stashed away in national (or international) reservoirs that have nothing to do with their origins. Production decisions, approximations and provenance are buried in a huge electronic correspondence to which no-one has access.

A similar path is followed by small data producers scattered around the globe. Both large and small data producers face the same problems in knowledge cataloguing, preservation and dissemination. It is imperative to plan and build repositories that store data as they emerge and to retain the history of the decisions and criteria that generate them[i]. Starting the century several multilateral organizations and planners in Europe and the United States generated technical reports to encourage the preservation of important scientific data collections[ii]. Recently, most of these recommendations have rooted as national and multinational initiatives for general policies concerning data curation[iii]. However, many of these recommendations have not permeated into the producing communities and/or to the collection custodians in these countries. The situation is even worse in Latin America where we are still not convinced by, or at least aware of, the new paradigms in the production and dissemination of scientific knowledge, and consequently, only a low-level use of Information and Communication Technology (ICT) awareness has been incorporated in teaching and research.

Moving in this direction, we feel the need to identify instruments, resources, data bases that are or may potentially be accessed, via Internet or advanced research and education networks, and shared with other researchers and research groups. This sharing may well be under conditions/restrictions, but certainly will boost cooperation within Latin America and between Latin America and Europe. Knowing where data sources and remote instruments are (or could be) will also provide an idea for the capacity planning of future e-infrastructure.

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

1.2 A METHODOLOGY TO IDENTIFY OTHER INSTRUMENTS AND DATA SOURCES

To identify instruments and data sources, various approaches were used:

1. A survey of scientific instruments (see appendix and next section), equipments or devices designed to search, acquire, measure, observe and store reproducible and verifiable data, as well as datasets and repositories with the possibility of remote use: operated/accessed remotely via Internet.
2. Internet searches for this type of resources.
3. Direct contact/interviews with National Research and Educational Network (NREN) personal.

1.3 SURVEY ANALISYS TO IDENTIFY RESOURCES AND DATA SOURCES

The survey conducted can be found in the following URL's:

1. English version: http://encuestas.redclara.net/index.php?sid=67821&lang=en
2. Spanish version: http://encuestas.redclara.net/index.php?sid=52525&lang=es

In short the questions in the survey were related to:

1. Institution and resource or data source identification.
2. Connectivity: Internet and/or education and research networks, and bandwidth.
3. Possibility of remote access.
4. Who has access to it and access conditions.
5. Data organization and data size.
6. Services provided.

In total we got 23 answers, 21 in Spanish and 2 in English. The Excel file *Instruments.xlxs* provides a summary with the organization that owns the resource/data, its location, a description of the resource/data, its purpose, the URL and the type of users that have access to it:

1. There are 6 related to computing resources.
2. Two repositories of scientific articles and academic work, one of them integrates the Mexican higher education repositories, and the other one integrates repositories from eight Latin American countries: Argentina, Brazil, Chile, Colombia, Ecuador, El Salvador, Mexico, Peru and Venezuela.
3. 13 instruments, 8 of them generate data constantly (they are data sources), the other 5 generate data on demand, i.e., particular data for specific studies.
4. One collects and organizes data, a virtual observatory.
5. One is a scientific visualization wall.
6. Five are only for local users, not open to non members of the organization and with no immediate plans to open them.
7. The other 18 are basically available by project submission and approval or by being affiliated to the organization.

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

1.4 VIRTUAL COMMUNITIES, BIG INSTRUMENTS AND CONSOLIDATED DATA SOURCES IN LATINAMERICA

There are several active world scale virtual research communities from Latin American institutions. These communities share data, either streaming them out from big instruments operated by a multi-institutional consortia, such as particle accelerators or telescopes, or produced by individual institutions or research groups. Datasets coming from High Energy Physics experiments and Astroparticle Observatories are non-public or not open, they are mainly reserved to be used by the members of the consortia. This situation contrasts visions and actions of the astronomical communities where most of their relevant and important datasets are open and freely available form online service centres. Among these unique data preservation and dissemination initiatives we could mention: the SAO/NASA Astrophysics Data System (ADS[iv]); the Centre de Données astronomiques de Strasbourg (CDS[v]), the NASA/IPAC Extragalactic Database (NED[vi]) and the International Virtual Observatory Alliance[vii].

1.4.1 CERN COMMUNITY IN LATIN AMERICA

CERN[viii], the European Organization for Nuclear Research, is the world's largest particle physics centre and one of Europe's first joint ventures for research and high-tech activities. The CERN scientific programme is mainly based on the operation of the Large Hadron Collider and its four experiments: ALICE[ix], ATLAS[x], CMS[xi] and LHCB[xii].

These four main experiments generate a huge amount of data that has to be processed/analysed remotely by hundreds of institutions around the world. The data distribution models used by these experiments is (originally) based on the MONARC model[xiii], a hierarchical, tier model with a central data store: Tier0. The central data system maintains copies of all data, which are replicated at a number of geographically distributed Tier1's. Each Tier1 has a number of Tier2's which provided additional computing resources, obtaining data from the associated Tier1[xiv]. Despite that ATLAS and CMS are now moving away from the MONARC model to a mesh model, however the current model will keep working worldwide for the coming years[xv].

In Latin America there are 23 partner institutions (see *Instruments.xlxs*) from 7 countries (ar, br, cl, co, cu, mx, pe) contributing to CERN main experiments and there are 6 Tier2 distributing and synchronising datasets with the other Tier architecture. The institutions providing data and computing resources are:

1. Universidad Nacional de la Plata[xvi] (UNLP), La Plata, Argentina;
2. Universidade de São Paulo[xvii] (USP), Sao Paulo Brazil;
3. Centro Brasileiro de Pesquisas Físicas[xviii] (CBPF), Rio de Janeiro, Brazil;
4. Sao Paulo Research and Analysis Center[xix] (SPRACE) Sao Paulo Brazil;
5. Universidad Técnica Federico Santa María[xx] (UTFSM), Valdivia, Chile;
6. Universidad Nacional Autónoma de México[xxi], UNAM, México DF, México

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

In addition to datasets these institutions provide Grid computing recourses to analyse the data. The monitoring of the Grid operations for Latin America is carried out by ROC-LA[xxii], an organisation, supported by the Centro Latino-Americano de Física (CLAF[xxiii]), lead by UNAM in cooperation with CBPF, UTFSM and UniAndes[xxiv].

1.4.2 ASTROPARTICLE OBSERVATORIES IN LATINAMERICA

There are two main Astroparticle instruments installations in Latinamerica:

1. The Pierre Auger Cosmic Ray Observatory[xxv] is one of the biggest astroparticle installations in the world. Located at Pampa Amarilla in western Argentina, it is operated by a consortium of 100 institutions from 18 countries. There are 23 Latin-American universities, from 4 countries (ar, bo, br, and mx) that are part of this consortium. The data recorded at the experiment site is preserved and disseminated by the Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules[xxvi] (CC-IN2P3) at Lyon France. This centre also provides a computational grid installation for simulations.
2. The Large Aperture Gamma Ray Burst Observatory, LAGO[xxvii] is a recent collaboration that comes from the association of Latin American astroparticle researchers. It is a network of ground-based detectors motivated by the experience of the Pierre Auger Observatory. There are 22 institutions from 9 Latin American countries (ar, bo, co, ec, gt, mx, pe and ve) involved in this collaboration. LAGO is building a distributed network of data repositories, but now all the data is recorded in 4 operational sites are preserved at Universidad Industrial de Santander[xxviii], Bucaramanga-Colombia.

Other installation will become an important data source for the Astroparticle community in the near future. The HAWC (High Altitude Water Cherenkov Gamma Ray Observatory[xxix]), on the flanks of the Sierra Negra volcano near Puebla, Mexico, is under construction supported by a cooperation from government agencies from Mexico and Unite State.

1.4.3 LATIN AMERICA ASTRONOMICAL OBSERVATORIES

As we have mentioned before, Astronomy is a pioneer in data-driven science and astronomical communities are becoming early builders and adopters of distributed digital data curating environments. Astronomy has several important contributions to the present era of data-driven science, adopting a unified taxonomy, vocabulary and coded definition of metrics and units; incorporating peer reviewed data carefully collected with rigorous statistical standards and integrating tractable data provenance, generating innovative schema to disseminate their available data[xxx].

For years this community has been building a knowledge platform that has become a revolution in the way astronomers use data, providing a role model to other disciplines on how technology can be used to improve the quality and effectiveness of scientific enquiry. These unique data services involve: the SAO/NASA Astrophysics Data System[xxxi], the Centre de Données astronomiques de Strasbourg[xxxii] and the NASA/IPAC Extragalactic Database[xxxiii]; the International Virtual Observatory Alliance (IVOA[xxxiv]). the public data releases from individual astronomical projects; the preprint dissemination from the arXiv-astro-ph server[xxxv]; and the rapid dissemination of results

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

made possible by forward-thinking journals. Most of the data repositories are open-access compliant and enthusiastically promote the Open Data Movement. In addition to the data services, the IVOA offers the community a toolkit of online software for data analysis[xxxvi].

Latin America hostess several of the most important astronomical observatories of the world. Most of these installations are international collaborations, operated by academic institutions from Europe, Japan and United States. All these observatories maintain local data repositories and most of them are member of IVOA. We have identified 21 main astronomical sites where research groups from 50 academic institutions of 5 countries collaborate (see *Instruments.xls*). It is worth mentioning that only Argentina, as a country initiative, is collecting and preserving its astronomical data through the Nuevo Observatorio Virtual Argentino (NOVA[xxxvii]). This national initiative is encouraging the participation of all astronomical institutions in Argentina to promote the generation and integration of information technology with special emphasis on the statistical analysis of data and management of astronomical images. Argentina contributes to IVOA and the integration of local data with global standards is agreed and coordinated by the alliance.

1.4.4 ADVANCED COMPUTATIONAL SERVICES IN LATINAMERICA

In this section we will mention some relevant projects co-financed by the European Commission related to setting up Grid e-Infrastructures in LA, cooperation among institutions and research groups, and High Performance Computing. A good number of Latin American universities or institution are participating or have participated in them. All these e-Infrastructures end up being shared resources and potential computing and data sources. On top of these e-Infrastructures particular application can be implemented and as well as known cpu consuming and big data generating applications for weather forecast, chemistry, genomics, physics, to mention some. All of them lie in the spectrum of instruments and data sources this work is looking for. Among these projects we have:

1. **EELA**[xxxviii]: *E-Infrastructure shared between Europe and Latin America*, went from January 2006 to December 2007. The project built a Grid e-Infrastructure between Europe and Latin America. It was a test e-Infrastructure intended for dissemination, training, learning and setting up a network of people involved with technical, decision making, and organizational issues.
2. **EELA-2**: *E-science grid facility for Europe and Latin America*. The project ran from April 2008 to March 2010. It was the continuation of EELA but now the e-Infrastructure ended up as a production quality Grid. Dissemination, training and learning in Grid technologies kept being a priority as well as maintaining and increasing the network of people.
3. **GISELA**[xxxix]: *Grid Initiatives for e-Science virtual communities in Europe and Latin America*. A project between September 2010 and August 2012 focusing on supporting virtual communities and running the Grid a-Infrastructure. Learning and training was carried out but in a lesser scale than in EELA and EELA-2. This project also had the mission to transfer the e-Infrastructure built during EELA, EELA-2 and GISELA to CLARA.

During the GISELA project, a number of institutions signed an agreement to provide continuity to the e-Infrastructure handed over by GISELA when it ended. The institutions that signed this

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

agreement where: CEDIA[xl] (Ecuador), CUDI[xli] (Mexico), RENATA[xlii] (Colombia), UNAM[xliii] (Mexico), and UNIANDES[xliv] (Colombia). Later RedCONARE/CeNAT[xlv] (Costa Rica) and InnovaRed[xlvi] (Argentina) joined the agreement.

This agreement let to the launch of the **SCALAC**[xlvii]: *Servicio de Cómputo Avanzado para América Latina y el Caribe*, initiative, aiming at providing computing resources and support to research groups in Latin America. Since the starting of SCALAC on the 1st of March 2013, a task force coordinated by UNAM has been working to guarantee the interoperation of the e-Infrastructure. Several centres in Europe and LA are supporting SCALAC, among them we can mention:

1. **BSC**[xlviii]: Centro de Supercomputación de Barcelona, Spain
2. **CeCalCULA**[xlix]: Centro Nacional de Cálculo Científico Universidad de Los Andes, Venezuela
3. **CEDIA**: Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado, Ecuador
4. **CESUP**[l]: Centro Nacional de Supercomputação, Brazil
5. **CETA-CIEMAT**[li]: Centro Extremeño de Tecnologías Avanzadas, Spain
6. **CIEMAT**[lii]: Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
7. **CUDI**: Corporación Universitaria para el Desarrollo de Internet CUDI, México
8. **INNOVA|RED**[liii]: Red Nacional de Investigación y Educación de Argentina, Argentina
9. **LNCC**[liv]: Laboratório Nacional de Computação Científica, Brazil
10. **RedCLARA**[lv]: Cooperación Latino Americana de Redes Avanzadas, Internacional
11. **CeNAT** Centro Nacional de Alta Tecnología, Costa Rica
12. **RENATA**: Red Nacional Académica de Tecnología Avanzada, Colombia
13. **SC3UIS**[lvi]: Supercomputación y Cálculo Científico UIS, Colombia
14. **SINAPAD**[lvii]: Sistema Nacional de Procesamiento de Alto Desempeño a través del Laboratorio Nacional de Computación Científica, Brazil
15. **UNAM**: Universidad Nacional Autónoma de México, México
16. **UNRC**[lviii]: Universidad Nacional de Río Cuarto, Argentina

Another project tightly related to computing recourses is **RISC**[lix]. The RISC web site defines the project as "The RISC project aims at deepening strategic R&D cooperation between Europe and Latin America in the field of High Performance Computing (HPC) by building a multinational and multi-stakeholder community that will involve a significant representation of the relevant HPC R&D European and Latin American actors (researchers, policy makers, users)." Besides aiming at setting up HPC resources, the targeted research areas of this project may also offer a good source of data for analysis.

The institutions involved in this projects can be found in the attached Excel file: *Instruments.xls*, in total there are 96, all potentially capable or already providing computing resources as well as access to instruments and data via Internet. As an example, CIEMAT, LNCC, SC3UIS and CERN, provide computing resources and data and they filled out the survey.

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

1.4.5 AGRICULTURE, BIOINFORMATICS AND BIODIVERSITY

1.4.5.1 BIOINFORMATICS AND BIOMEDICINE

Communities like High Energy Physics and Astronomy acquire their data from big instruments and have a significant tradition to curate and disseminate those datasets through data repository services. Bioinformatics, as in the rest of the world is emerging as the third thematic area to provide data sources and build international consortia to curate/maintain them. Other areas like agriculture and biodiversity, depending to procure their data upon remarkable fieldwork effort, lack the tools and infrastructure to manage the growing amounts of data generated by new forms of instrumentation. The lack of an integrated framework for managing these types of scientific data presents significant barriers not only to those scientists conducting the research, but also to those who would subsequently reuse the datasets. This situation is changing slowly as thematic communities are getting aware of the paramount importance of use and reuse of data to generate new knowledge.

In Latin America there is growing interest to develop national bioinformatics initiatives. This is evident by several recently created national institutes in the region. Particularly, we could mention

1. Instituto de Investigación en Biomedicina de Buenos Aires[lx], (Biomedicine Research Institute of Buenos Aires), a partner Institute of the Max Planck Society (IBioBA-MPSP) at Buenos Aires, Argentina;
2. Bioinformatics Lab (Labinfo[lxi]) of the National Laboratory of Scientific Computation of the Ministry of Science and Technology (MCT), research campus in Petrópolis Brazil.
3. Nodo Nacional de Bioinformática (National Bioinformatics Node[lxii]) at Universidad Nacional Autónoma de México in México DF, México.
4. Centro de Bioinformática y Biología Computacional[lxiii] at Manizales-Colombia.

The Bioinformatics in Latin America is increasingly gaining international reputation[lxiv] and its has a growing global visibility. Six Latin American countries (ar, br, cl, co, cr and mx) are nodes of the EMB.net[lxv] (a world-wide network, spread over 29 countries, that rises awareness, develops and supports databases/tools for biotechnology and bioinformatics-related research areas). There are three important data sources developed, hosted and maintained by regional institutions. They are the Star STING server[lxvi], Regulatory networks in gamma-proteobacteria database[lxvii] and Cancer-Test (CT) database[lxviii], It is also worth mentioning the participation of Argentina in the European Molecular Biology Laboratory[lxix], EMBL, as an as associate member state.

1.4.5.2 AGRICULTURE AND BIODIVERSITY

In agriculture and biodiversity there are several institutes that provide information products and data sources. Among them, we could highlight the following research institutes with worldwide contributions:

1. Centro de Referência em Informação Ambiental (Center for Reference and Environmental Information), CRIA[lxx], in Brazil;

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

2. Laboratorio Nacional de Genómica y Biodiversidad (National Laboratory of Genomics for Biodiversity), LANGEBio[lxxi], at Centro de Investigacion y de Estudios Avanzados del IPN in México;
3. Instituto Nacional de Biodiversidad (National Institute of Biodiversity), INBio[lxxii], at Costa Rica
4. International Potato Center, CIP[lxxiii], a worldwide organization with headquarters in Lima-Perú.
5. Centro Internacional de Agricultura Tropical, CIAT at Colombia[lxxiv]
6. The Smithsonian Tropical Research Institute, STRI[lxxv], Institute in Panamá

It is worth mentioning the participation of CIP in a worldwide data preservation and dissemination initiative through CGIAR (formerly the Consultative Group on International Agricultural Research[lxxvi])

1.4.6 SOCIAL SCIENCES DATA SOURCES

The e-research and data driven social science research is just emerging in Latin America, mostly promoted by the accessibility to the e-Infrastructure[lxxvii]. Digital repositories (including institutional repositories) are becoming potential interesting data sources for social & sociological studies[lxxviii]. In Latin America there are at least two relevant Digital repository networks that could become important data sources in the near future, They are:

1. SciELO (Scientific Electronic Library Online[lxxix]). A cooperative bibliographic database and a digital library of open access journals. It started originally in Brazil but now is a network of 12 countries (ar, br, cl, co, cr, cu, es, mx, pe, ve and za) and three other (bo, py and uy) are on the way to join.
2. Red Federada de Repositorios Institucionales de Publicaciones Científicas: LAReferencia[lxxx]. It is network of institutional repositories of more than 100 Latin American universities and research institutions from 9 countries (ar, br, cl, co, ec, mx, pe, sv and ve).

There are other data sources which are interesting for the social science like Comisión Económica para América Latina y el Caribe CEPAL[lxxxi]

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

1.5 APPENDIX: THE SURVEY

Scientific Instruments

Survey of scientific instruments: equipments or devices designed to search, acquire, measure, observe and store reproducible and verifiable data, and the possibility of remote use: operated remotely via Internet.

Thank you so much for your time!

This is a survey of scientific instruments: equipments or devices designed to search, acquire, measure, observe and store reproducible and verifiable data, and the possibility of remote use: operated remotely via Internet.

We understand that these instruments can be used in general "campaigns" that produce large amounts of data that is shared by a community. The data is shared through repositories or repository networks. Terrestrial and space telescopes and particle accelerators are among the most typical examples.

There are other tools which can be used almost uniquely, whether teleoperated or through very specific use proposals. The generated data is usually not placed in repositories to be shared.

This survey wants to explore what instruments exist and their possible use spectrum (between open and restricted).

There are 20 questions in this survey

Information about the scientific instrument

The following questions are related to connectivity and the use of the instrument.

1 Please identify your organization. *

Please write your answer here:

2 Please identify the instrument *

Please write your answer here:

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

3 Is the instrument connected to the Internet? *

Please choose the appropriate response for each item:

	NO	<10 Mbps	10 to 100 Mbps	100 to 1000 Mbps	> 1000 Mbps
Commercial	○	○	○	○	○
Advanced	○	○	○	○	○

Mbps = Mega bits per second

Advanced Internet is a network, used primarily for research, education and innovation, which allows the transmission of large volumes of information at high speeds and integrates networks of universities, research centres, and public and private high level agencies, providing access to resources and applications that would be very difficult to reach by Commercial Internet channels due to its low availability and saturation among others.

CLARA (Latin American Cooperation of Advanced Networks) develops and operates the only Advanced Internet network in Latin American integrating 15 countries, and interfaces with advanced international networks in North America, Europe, Asia and Africa, including Internet2, GEANT2, CANARIA, APAN and TERNA.

Commercial Internet is the name commonly used to distinguish Traditional Internet from Advanced Internet. Commercial Internet allows access to Internet services including Google, YouTube, Twitter, Face book, among others.

4 If the instrument can NOT be used remotely, are there plans to allow remote use?

Please choose **only one** of the following:

- ○ NO
- ○ Between the next 6 months
- ○ In the course of a year
- ○ In more than one year

5 For the case in which the instrument IS used remotely, from where is it used ?

Please choose **all** that apply:

- ☐ The country where the instrument is located
- ☐ Latin American countries different from where the instrument is located
- ☐ Europe
- ☐ USA
- ☐ Asia
- ☐ Other (specify):: _____

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

6 If the instrument is NOT available for research groups that are NOT part of your organization, are there plans to allow its use?

Please choose **only one** of the following:

- ○ NO
- ○ Between the next 6 months
- ○ In the course of a year
- ○ In more than one year

7 For the case in which the instrument IS available to researchers outside your organization, are there restrictions depending on whether these groups are from public or private institutions?

Please choose **only one** of the following:

- ○ NO
- ○ YES (specify): [ ]

8 Is there a web address where the requirements for using the instrument can be found?

Please choose **only one** of the following:

- ○ NO
- ○ YES, URL: [ ]

9 Under what form could other groups access your instrument?

Please choose **all** that apply:

- ☐ An enrolment process with your organization
- ☐ Submitting projects
- ☐ Pay per use
- ☐ Other (specify):: [ ]

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

10 Are there requirements, in terms of the research lines, that these groups should follow?

Please choose **only one** of the following:

- ○ NO
- ○ YES (specify): _____

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

Data Generated by the Instrument

The following questions are related to the data generated by the instrument, if it does.

11 Approximately how much data is generated per month?

Please choose **only one** of the following:

- ○ <10 MB/month
- ○ 10 to 100 MB/month
- ○ 100 to 1000 MB/month
- ○ > 1000 MB/month

MB = Mega Bytes (1 byte = 8 bits)

12 Where and how much data is stored?

Please choose the appropriate response for each item:

	<10 MB/month	10 to 100 MB/month	100 to 1000	MB/month	> 1000 MB/month
Locally	○	○	○	○	○
Remotely in the country where the instrument is located	○	○	○	○	○
Remotely elsewhere	○	○	○	○	○

13 Is there any data curation done?

Please choose **all** that apply:

- ☐ NO, stored raw
- ☐ Customized standardization/classification
- ☐ Standardization/classification according to international standards
- ☐ They are organized in databases or data repositories
- ☐ They incorporate metadata

14 What kind of access to your data is provided?

Please choose **only one** of the following:

- ○ They are open access
- ○ They require authorization for use

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

15 Is there information (website URL) on the accessible data (and metadata) organization?

Please choose **only one** of the following:

- ○ NO
- ○ YES, URL: [ ]

16 Do you do data processing?

Please choose **all** that apply:

- ☐ Distributed
- ☐ Centralized
- ☐ Through a Science Gateway
- ☐ Other (specify):: [ ]

17 Do you do/require data visualization?

Please choose **all** that apply:

- ☐ Local
- ☐ Remote
- ☐ Other (specify):: [ ]

18 Do you collaborate with other researchers/groups to handle your data?

Please choose **only one** of the following:

- ○ NO
- ○ YES, with whom? [ ]

A project of the Seventh Framework Programme (FP7)

This project is funded by the European Commission

A project implemented by RedCLARA

Generated services

The following questions are associated with the services generated by the instrument, if any.

19 Do you use web interfaces to provide these services?

Please choose **only one** of the following:

- ○ NO
- ○ YES, URL: [ ]

20 What services do you provide?

Please choose **all** that apply:

- ☐ Access to the instrument by booking time
- ☐ Course on how to use the instrument
- ☐ Course on remote manipulation of the instrument
- ☐ Course to demonstrate an application that uses the instrument
- ☐ Citizen science projects
- ☐ Public access to data
- ☐ Other (specify):: [ ]

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

[i] See a classical discussion of this problem of how small labs and institutes are facing the data deluge problems in C Borgman, J Wallis, and N Enyedy.(2007) **Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries** *Int J Digit Libr,* **7,** 17-30.
[ii] Arzberger, P., Schroeder, P., Beaulieu, A., et al. (2004) **Science and government: an international framework to promote access to data**. *Science*, **303**:1777–1778; Simberloff, D., Barish, B. C., Droegemeier, K. K., et al. (2005) **Long-lived digital data collections: enabling research and education in the 21st century**. Technical Report NSB-05-40, National Science Foundation, Washington DC, USA; and Lyon, L. (2007) **Dealing with data: roles, rights, responsibilities and relationships**. Consultancy Report, UKOLN, University of Bath, UK.
[iii] See **Report of The European Commission Public Consultation on Open Research Data** (July 2013) http://ec.europa.eu/research/science-society/document_library/pdf_06/report_2013-07-open_research_data-consultation.pdf and **Digital Research Data Sharing and Management** Committee on Strategy and Budget National Science Board NSF Technical Reports (2011) http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf
[iv] http://www.adsabs.harvard.edu/ and
[v] http://cdsweb.u-strasbg.fr/
[vi] http://nedwww.ipac.caltech.edu/
[vii] http://ivoa.net
[viii] http://home.web.cern.ch
[ix] A Large Ion Collider Experiment http://aliceinfo.cern.ch
[x] A Toroidal LHC Apparatus http://atlas.web.cern.ch
[xi] Compact Muon Solenoid http://cms.web.cern.ch
[xii] Large Hadron Collider beauty https://lhcb.web.cern.ch/
[xiii] http://monarc.web.cern.ch/MONARC/
[xiv] See CERN Tier distribution sites http://wlcg-rebus.cern.ch/apps/topology/ and a map for the CERN Tier2 sites http://gstat2.grid.sinica.edu.tw/gstat/gstat/geo/openlayers#/WLCG_TIER/2
[xv] See Witt, S de, and J W Huang. 2014. **"Tier 1 Evolution in Response to Experiment Data Model Changes in LCG.**" *Journal of Physics: Conference Series* 513 (4): 042010. doi:10.1088/1742-6596/513/4/042010, and references therein.
[xvi] http://www.unlp.edu.ar/
[xvii] http://www5.usp.br/
[xviii] http://www.cbpf.br/
[xix] http://www.sprace.org.br/
[xx] http://www.utfsm.cl/
[xxi] http://www.unam.mx
[xxii] http://www.e-science.unam.mx/rocla.jsp
[xxiii] http://www.claffisica.org.br
[xxiv] http://www.uniandes.edu.co/
[xxv] http://www.auger.org/
[xxvi] http://cc.in2p3.fr/
[xxvii] http://lagoproject.org/
[xxviii] http://www.uis.edu.co
[xxix] http://www.hawc-observatory.org
[xxx] Gray, J., Szalay, A. S., Thakar, A. R., et al. **Online scientific data curation, publication, and archiving**. Technical Report MSR-TR-2002-74, Microsoft Research, Redmont, USA, 2002, pp. 1–6. http://research.microsoft.com/pubs/64568/tr-2002-74.pdf
[xxxi]http://www.adsabs.harvard.edu/

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

[xxxii] http://cdsweb.u-strasbg.fr/
[xxxiii] http://nedwww.ipac.caltech.edu/
[xxxiv] http://www.ivoa.net/
[xxxv] http://arxiv.org/archive/astro-ph
[xxxvi] See Djorgovski, S. G., Williams, R. (2005) **Virtual observatory: from concept to implementation**. In From Clark Lake to the Long Wavelength Array: Bill Erickson's radio science, N. Kassim, M. Perez, M. Junor, P. Henning (eds.), ASP Conference Series, Vol. 345, San Francisco, USA, pp. 517–530; and Berriman, G Bruce, Robert J Hanisch, and T Joseph W Lazio. (2012) **The Role in the Virtual Astronomical Observatory in the Era of Massive Data Sets**. Observatory Operations: Strategies 8448 (September).
[xxxvii] http://nova.conicet.gov.ar
[xxxviii] http://www.eu-eela.eu/first-phase.php
[xxxix] http://www.gisela-grid.eu/
[xl] http://www.cedia.org.ec
[xli] http://www.cudi.mx
[xlii] http://www.renata.edu.co
[xliii] http://www.unam.mx
[xliv] http://www.uniandes.edu.co/
[xlv] http://www.cenat.ac.cr/
[xlvi] http://www.innova-red.net/
[xlvii] https://comunidades.redclara.net/wiki/scalac
[xlviii] http://www.bsc.es
[xlix] http://www.cecalc.ula.ve
[l] http://www.cesup.ufrgs.br
[li] http://www.ceta-ciemat.es
[lii] http://www.ciemat.es
[liii] http://www.innova-red.net
[liv] http://www.lncc.br
[lv] http://www.redclara.net
[lvi] http://sc3.uis.edu.co
[lvii] https://www.lncc.br/sinapad
[lviii] http://www.unrc.edu.ar
[lix] http://www.risc-project.eu
[lx] http://www.ibioba-conicet.gob.ar
[lxi] http://www.labinfo.lncc.br
[lxii] http://www.ccg.unam.mx and http://www.ibt.unam.mx
[lxiii] http://www.bios.co/
[lxiv] See Neshich, G. (2007). **Computational biology in Brazil.** *PLoS computational biology*, *3*(10), e185; Palacios, R., & Collado-Vides, J. (2007). **Development of genomic sciences in Mexico: a good start and a long way to go**. *PLoS computational biology*, *3*(9), e143; Moreno, E., Lomonte, B., & Gutiérrez, J. M. (2008). **Computational biology in Costa Rica: The role of a small country in the global context of bioinformatics**. *PLoS computational biology*, *4*(3), e1000040 and references therein
[lxv] http://www.embnet.org
[lxvi] http://sms.cbi.cnptia.embrapa.br/SMS/
[lxvii] http://www.tractor.lncc.br
[lxviii] http://www.cta.lncc.br

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA

[lxix] The European Molecular Biology Laboratory is an intergovernmental organisation specialising in basic research in the life sciences with 21 member states, one prospect and two associate member states http://www.embl.de
[lxx] http://www.cria.org.br
[lxxi] http://www.langebio.cinvestav.mx
[lxxii] http://www.inbio.ac.cr/
[lxxiii] http://cipotato.org
[lxxiv] http://ciat.cgiar.org Colombia is a CIAT host country and CIAT has a strategic alliance with the country's Ministry of Agriculture and Rural Development (MADR) and the Colombian Corporation of Agricultural Research (CORPOICA).
[lxxv] http://biogeodb.stri.si.edu/bioinformatics/en/
[lxxvi] See Gassner, A., Alvarez, L. M., Bamba, Z., Beare, D., Bernardo, M., Biradar, C., ... & Chukka, S. R. (2013). White Paper: Shifting the goal post-from high impact journals to high impact data and http://www.cgiar.org
[lxxvii] See Arcila, C., Piñuel, J. L., & Calderín, M. (2013). The e-Research on Media & Communications: Attitudes, Tools and Practices in Latin America Researchers. *Comunicar*, *20*(40), 111-118; Calderón, C. A. (2013). e-Investigación en Ciencias Sociales: Adopción y uso de TIC por investigadores sociales de América Latina; and references therein.
[lxxviii] Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: the challenge of digital devices. *Theory, culture & society*, *30*(4), 22-46.
[lxxix] http://www.scielo.org
[lxxx] http://lareferencia.redclara.net/
[lxxxi] http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Portada.asp?idioma=e

A project of the Seventh Framework Programme (FP7)
This project is funded by the European Commission
A project implemented by RedCLARA