Title page

INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME

Contract for:

Shared-cost RTD

Annex 1 - "Description of Work"

Project acronym: EGSO
Project full title: European Grid of Solar Observations
Proposal/Contract no.: IST-2001-32409
Related to other Contract no.: *(to be completed by Commission)*

Date of preparation of Annex 1:
19 October 2001
Operative commencement date of contract: *(to be completed by Commission)*

Shared Cost RTD CPF Form – Form A2

EUROPEAN COMMISSION
RESEARCH DIRECTORATES GENERAL
SHARED COST RTD CPF FORMS

EN C 1 FP5RTD

FOR COMMISSION USE ONLY

Project Acronym [2]		Proposal No [3]	

A2. Project Summary [20]

Objectives (maximum 1000 characters)

The ESGO aims create a virtual archive by federating solar data centres scattered across Europe into a grid in order to enhance access for both the solar and non-solar communities. EGSO will create a catalogue of solar features at different wavelengths to provide an innovative way of conducting data searches based on phenomena. It will create a unified catalogue of space- and ground-based solar observations to facilitate the identification of matching data sets based on date, time and location on the Sun, and provide search and visualisation tools using web-based interfaces to assist the user when examining the catalogues. Other tools will locate data identified in the search and retrieve them for more detailed analysis, extracting and calibrating them where necessary. EGSO will also provide the facility for users to upload their own software to analyse extended time intervals of large, complex data at source data centres, where their volume makes it impractical to retrieve them.

Description of the work (maximum 2000 characters)

The EGSO will be constructed from several component parts, catalogues, search and visualisation tools, and federation middleware that combine many solar data centres into a virtual archive.
The observational catalogues will be built from existing catalogues where possible, but will use a standard format designed to facilitate searching for matching observations at many wavelengths. Dependencies on ancillary data will be removed to produce catalogue files that are independent and self-describing; if required, observatories will be consulted to help resolve inadequacies in existing logs. The solar feature catalogue will be created using image recognition tools developed for the project. Full disk ground- and space-based imagery will be processed to build a catalogue of the occurrence and location of features such as filaments, active regions, flares, coronal holes, etc. Tools to search the catalogues using web-based user interfaces will be developed with visualisation techniques designed to aid the user in their selections of the data; full-disk synoptic images and quick-look products will be used to assist the process.
The federation middleware provides the underlying fabric of the EGSO. It must provide controlled access to selected data and other resources at the sites, while ensuring the security of the grid as a whole. While standard data reduction techniques based on SolarSoft will be used for the extraction and calibration of data, the capability will be created to execute code provided by the user for more complex analysis of large volumes of data at the data source.
Where possible, tools used in the project will be based on open source code - these will be augmented and refined as necessary and any such additions will also be open source. Extensive consultations with users will ensure EGSO provides the capabilities the community requires. Issues related to user authentication and security for the diverse user community will be resolved during the project.

Milestones and expected results (maximum 500 characters)

1) Architectural design of the EGSO based on scientific requirements, user community input, and surveys of appropriate technologies.
2) Demonstration federation of solar data centres, providing limited data access and a testbed for full development.
3) Creation of observing catalogues and the catalogue of solar features.
4) Full federation of data centres, with tools for integrated data extraction, processing and retrieval through interactive catalogue searches, providing full user access.

2. Project Objectives

Our knowledge of the complex behaviour of the atmosphere of the Sun and the associated outflowing of the solar wind are in the process of being revolutionized by the wealth of data that have been obtained in recent years. As the closest star that can be resolved in fine detail, and one that has direct influence on various aspects of life on Earth, the Sun is continuously observed by advanced space satellites, and dozens of ground-based observatories. The ceaseless and dramatic changes occurring at all heights in the solar atmosphere are now being studied by cutting edge instruments throughtout the entire range of wavelengths in the electromagnetic spectrum. Each wavelength generally isolates one specific level or physical process in the solar atmosphere. Utilizing a time sequence of observations at multiple wavelengths therefore allows us to now undertake ground-breaking four-dimensional studies of numerous intriguing solar phenomena. This form of research, combining multiple datasets from a broad collection of sources, will be essential, in understanding and solving the difficult questions that remain unanswered in solar physics.

However, there remain numerous hurdles in efficiently exploiting the broad mix of heterogeneous datasets that are dispersed throughout Europe and around the globe. The problems encountered by the scientist or other end users include the large number of different observatories and instruments operating, the lack of any standard interface or mechanism for searching and identifying relevant data. Also, the logical isolation of individual datasets from one another, and the enormous volume of data that is available. In order to rectify many of these problems, we are proposing to construct a grid to federate individual data centers and create a unified method for searching for and analyzing solar data sources. This grid will be unique because of the large number of individual data sources to be included into the grid, the inhomegeneity of the data cataloguing and storage among the various sites, and the broad range of wavelengths and temporal scales, from seconds to decades, of interest to solar physics.

In order to realize this grid, we will develop several innovative tools to allow the data to be made available to a wide audience in a simplified manner. These tools include: middleware and brokers that will allow multiple heterogeneous data resources to be virtually combined through a common metadata standard; a reusable set of modular components that will allow a straightforward integration of new data sources into the grid with a minimum of effort on the part of both the data centres. and the grid maintainers; and methods for augmenting the overall reliability of a grid made up of data sources of variable availability by providing redundancy through techniques such as caching, mirroring, and load-balancing. While these grid technologies will be developed in particular for the field of solar astrophysics, there use will be applicable to many other fields, especially those in which there are numerous data resources of widely varying size, reliability, or data management system to be combined in a data grid.

Specific Objectives.

The main objective of this project is to federate several digital solar archives into a data Grid, and develop search and retrieval tools that will make the solar data more accessible to the solar physics community, as well as to other scientific and non-scientific communities interested in solar data. The proposed activity will be known as the "European Grid for Solar Observations" (EGSO). The data included in this test bed include tens of terabytes of data generated at dozens of different observatories. In order to achieve this goal, the project will undertake the following major objectives:

- The development of a system architecture and associated search tools that will allow the user to interact with multiple existing solar archives as a single virtual entity, irregardless of the location of the metadata or data within the grid. . The tools and architecture should be flexible and scalable enough to seamlessly expand to include existing and future archives. As part of the initial study phase of the project, issues related to user authorization and limitations in access to data and other resources will be considered. The realization of this objective is covered by work packages 2 (WP2). ***The creation of the Trial Federation involving four sites can be used to demonstrate the success of this objective.***

- The implementation of tools that will allow the user to search data throughout the whole complex of archives, on the basis of essential parameters (such as time, heliocentric coordinates and spectral band) without any knowledge of details of the different instruments and observing activities that produced the data. In addition, more advanced visualization tools will be developed to assist the

user in making the data selection - the tools should be able to utilize the summary and synoptic full-disk images that are available at some sites. In response to initial data selection criteria, generate quick-look data (e.g. images reduced size and resolution) to help the user decide if particular datasets are of interest. This objective is covered in WP3. ***The ability to formulate requests through the interface and retrieve the selected data from the federated archives would demonstrate the successful completion of this objective.***

- The creation of a new format of observation catalogue which is derived from existing heterogeneous observing logs but is self-describing, complete in each fragment and with no dependencies on any other files. This will require the definition of a new metadata standard that is acceptable to several different communities within the field of solar physics. The realization of this goal is primarily described in work package 4 (WP4), but will begin with the consultation of the user community in work package 1 (WP1). ***The creation of at least a year's worth of fragments of the unified observing catalogues for space and ground-based observations would demonstrate the successful completion of this objective.***

- The investigation of methods to ensure an optimal reliability and functionality of the data grid as a whole, even when the individual nodes may not have the desired availability or network connections. The methods that will be developed include mirroring of data to more reliable nodes, caching algorithms that retain data that is most popular or least available, and schemes for network and load-balancing among various data centres. ***The development of the necessary tools and their use in the federated archive would demonstrate the successful completion of this objective.***

- The realization of a mechanism to extract and calibrate data identified in the searches, and return them to the user in an intelligent manner for more detailed analysis. This capability will also allow users to upload their own analysis code to source data centres to perform the detailed analysis of large data sets covering long time spans. The security measures needed to ensure that this activity does not compromise the system or Grid will be built into the middleware developed for the project. This is part of work packages 1 and 2 (WP1 & WP2). ***The ability for users to formulate and execute complex studies that are evaluated at the source node would demonstrate the successful completion of this objective.***

- The development of feature recognition algorithms in order to produce a catalogue based on solar features, e.g. the occurrence and location of flares, filament disruptions, etc. derived from the data stored within the grid itself. This will allow the user to make advanced searches, which have been up to now impossible, for observations related to special features or phenomena. This objective will be carried out in work package 5 (WP5). ***The creation of the Solar Feature Catalogue would demonstrate the successful completion of this objective.***

- The examination of the social issues necessary in linking multiple, sometime proprietary datasets generated by related but separate communities into a single common entity with certain enforced standards. This experience will be result of the work of the entire project, but most clearly the system definition of work package 1 (WP1). ***The successful completion of this objective will be contained as a section of the Final Report.***

The overall intention is to implement this project in a way that causes as little impact as possible to the existing archives so as to promote maximum uptake of the proposed Grid. At the same time, the system will be robust and simple enough to continue to operate beyond the duration of the project. Although only a few archives will be included in this Grid test bed, the architecture will be scalable so that it will be easy to include other data centres and new data sets in the future. The successful project will create a data Grid and tools that are widely accepted by the solar community, and, because of the attractiveness and ease of being involved, eventually expands to include the majority of solar datasets available globally.

3. Participant list

List of Participants

Partic. Role*	Partic. no.	Participant name	Participant short name	Country	Date enter project**	Date exit project**
C	1	University College London	UCL	UK	Start of Project	End of Project
P	2	Council for the Central Laboratory of the Research Councils	RAL-CCLRC	UK	Start of Project	End of Project
P	3	Osservatorio Astronomico di Torino	OATO	Italy	Start of Project	End of Project
P	4	Politecnico di Torino	Polito	Italy	Start of Project	End of Project
P	5	Observatory de Paris	BASS2000 Meudon	France	Start of Project	End of Project
P	6	Universite Paris Sud, Institute d'Astrophysique Spatiale	UPS-IAS	France	Start of Project	End of Project
P	7	Astrium plc	Astrium	UK	Start of Project	End of Project
P	8	University of Bradford	UNIBRAD	UK	Start of Project	End of Project
P	9	Solar Data Analysis Center	SDAC-NASA	US	Start of Project	End of Project
P	10	National Solar Observatory	NSO-NOAO	US	Start of Project	End of Project

*C = Coordinator (or use C-F and C-S if financial and scientific coordinator roles are separate)
P - Principal contractor
A - Assistant contractor

** Normally insert "Start of project" and "End of project". These columns are need for possible later contract revisions caused by joining/leaving participants.

4. Contribution to programme/Key action objectives

This proposal is submitted under the FP5 Action Line IST-2001-5.1.9 ("CPA9: Grid test beds, deployment and technologies"), with additional relevance to Action Lines IST-2001-3.1.3 ("Next Generation Digital Collections") and IST-2001-7.1.2 ("RN2: Technology and Applications Experiments"). In particular, under IST-2001-5.1.9, in order to establish the EGSO, the project addresses *"Test beds for progressive end-to-end service deployment and utilization of the Grid through integration and implementation of relevant components, such as network technologies, middleware and applications toolkits"*. To ensure that the EGSO will be widely accepted and able to expand, it addresses the *"development of Grid-specific toolkits and software and system architectures that ensure performance and scalability solutions together with usability and functionality"*. Also, *the "interoperability between different Grids"*, and the *"promotion of standards … to reduce the barriers of deployment"*. The enhanced access that the EGSO provides to existing and future collection of solar digital addresses the main objective of the Action Line IST-2001-3.1.3 *"To improve substantially, both qualitatively and quantitatively, access for citizens and professions to Europe's expanding repositories of cultural and scientific knowledge*

The project that is proposed will create a virtual archive by federating solar data centres across Europe. It will combine centres scattered across many countries into a unified data Grid and provide tools to locate and retrieve data, greatly enhancing access to these data. The project is designed to use the Grid to solve many of the problems that exist in the optimal utilization of solar observations. These comprise: finding out whether observations of a particular event or target exist, determining whether they are useful and then retrieving them in a form that can be used for detailed analysis. The Grid will be of immediate benefit to the European solar physics community, and will provide a standard entry point into solar data for other parts of the scientific community, including astrophysics, solar-terrestrial physics, space weather and climate physics. This collection of solar data will also serve as a resource for the general public by providing them a window on the solar research activity.

Although the Grid test bed created in the EGSO will involve only a limited number of data centres, this is determined by practical limitations in the initial phases of establishing this type of Grid rather than any intention to constrain the number of datasets included. Once established, it is intended that the Grid should provide mechanisms to easily expand to include as many additional institutes as wish to participate. The inclusion of groups from the USA in the consortium will ensure that standards and protocols developed within the project represent a global approach to a solar Grid and will be widely acceptable in the community. Consultation with solar-terrestrial groups will ensure maximum compatibility with that important community.

In order to establish the EGSO, many problems related to the Grid middleware will have to be confronted, and the architecture of the proposed system must be designed to address these issues. The solar community is distributed over many institutions throughout Europe in groups of varying sizes. Most solar datasets are managed by the teams that have been directly involved in the construction (and maintenance) of the space- or ground-based instruments that produce these data. Up to now, the data centres have acted mostly independently, providing basic access to their data through Web or other interfaces. The EGSO will provide access to data and other resources that is far beyond the current level of service. User authentication and site security concerns must be managed by the middleware if the impact on the individual sites of being part of the Grid is to be minimized. In order to provide different levels of access to the data, considering the diversity of the members of the community, it will be necessary to develop authentication methods for all potential users. The security model will be implemented in the middleware and will include a low-privilege access mode for unidentified users.

Although it is expected that the EGSO will encompass most of the solar data centres in Europe, it is inevitable that at some point it will be necessary to also interact with other data Grids, especially in other, related fields. The project will consult heavily with the solar-terrestrial community to try to ensure good interoperability with that discipline, independently of the different middleware software that may be used. The transfer of authentication and security information between grids will be one of the biggest problems, but since this problem has similarities to the problems described above of a providing access for a diverse and partly unknown user community, it is hoped that a common solution can be found to both problems - an extension to existing middleware that could have applications elsewhere.

The development of catalogue search and visualization toolkits are a key part of the project and will have applications in other projects.

5. Innovation

This project, the European Grid of Solar Observations, will provide a capability that currently does not exist anywhere in the world – the ability to search a large number of geographically scattered and heterogeneous archives in order to easily locate and retrieve the pieces of solar data that match the user's desired criteria. At the moment, the data from each space mission or ground-based observatory are dealt with essentially separately. Some data centres have gathered the data from a number of space missions or observatories together in a single location, but they are then treated independently and the tools to match observations based on criteria such as date, time, location on the Sun and type of observation do not exist. This limitation applies to all solar data, those in Europe as well as those in the rest of the globe.

The project therefore provides a unique service for an active and well-defined user community, and in order to do so confronts a number of general technical issues involved in establishing a data Grid.

The EGSO provides a novel problem in the creation of a data Grid because the data to be federated are of a very heterogeneous and distributed nature. There is no central institute that will serve as the primary source for the data. Instead, numerous institutes will contribute datasets, in a range of formats, to create an overreaching collection of solar data. Dealing with the differences in resources among different data centres will provide interesting insight into the social and technological problems in configurations of this type. This more "chaotic" situation, caused by the normal differences in participation by the numerous research groups in various space missions and ground-based observatories, may well common in other fields as well. The EGSO will build on the experiences and of other data grids to extend the state of the art in their application to more scattered organizations of data.

Solar observations, covering a broad range of wavelengths, atmospheric heights, and observational techniques, are by their very nature quite heterogeneous. For organizational, practical, and historical reasons, each instrument or observatory team has chosen its own way of reformatting and storing the data, with the FITS (Flexible Image Transport System) format (defined by the International Astronomical Union) providing only a loose standard. Each group has then chosen its own style of producing catalogues of the observations, some with severe dependencies on ancillary data. One innovative aspect this project will be to try to mitigate these problems by defining a common metadata standard applicable to all solar data, and creating a new unified catalogue of observations from existing data sources: online catalogues, observing logs, or directly from FITS file headers. The format of the common catalogue will be less extensive than some of the more detailed existing catalogues, but will contain all the information needed to facilitate and optimize searches for matching data sets. It will be self-describing, have no dependencies on other data sources, and will be partitioned so as to optimize user queries. This standardized format will allow common search procedures to be applied to all types of observations, even though the data themselves are inherently very different. Creating this unified observing catalogue removes the need to provide customized search capabilities for each catalogue at various data centres.

In addition, another catalogue will also be created, quite different in nature to the observing catalogues, and totally unprecedented in this field - the Solar Feature Catalogue. Many observations of the Sun only have general information about the coordinates being observed, and often only cover a small part of the solar disk; whether an interesting event occurs in the field of view during the observations is generally a matter of chance. The Solar Feature Catalogue (SFC) will be created using advanced feature recognition techniques (developed for the project) on full disk images. These techniques will utilize intelligent algorithms that cross-reference the different characteristics observed at several wavelengths to definitively identify the various features. The result will be a list of different features and their location on the solar disk as well as other derived characteristics, such as size, age, or complexity. This will allow users to formulate an entirely new set of queries based on the particular type of object or event, e.g. flares, filaments, coronal mass ejections, etc. The results of a search for a certain class of phenomena will then be fed back into the search engine of the unified catalogue of solar

observations to rapidly locate all data pertinent to the objects or events under study. This catalogue will be a unique application of advanced feature recognition techniques in solar physics to produce a uniform compilation of an expanded range of solar features. In addition, the algorithms developed to utilize the spectral or temporal domain (in addition to the normal spatial domain) will be on the cutting edge of this discipline and can presumably be transferred to other applications, such as remote earth sensing.

The format of the two catalogues, the unified observing catalogue and the Solar Feature Catalogue, will be finalized during the project. Currently there are a number of proposal for ways of specifying astronomical catalogues in the astrophysics community, e.g. from CDS, Strasbourg, and several other places. This project will apply the experience gained by these groups and extend upon their ideas in formulating the design of catalogues for the EGSO. Languages like XML (eXtensible Markup Language) are becoming increasingly popular for this type of application and it is considered to be a good candidate for the EGSO metadata catalogues. XML is self-describing and tools already exist to read XML, or convert it to HTML for use with Web browsers - it is anticipated that future Web enabling software will be able to use XML directly.

Tools to search the catalogues will be developed specifically for the project, but will draw on capabilities that have been pioneered in other areas of astrophysics. Catalogue search capabilities, tailored to nighttime observations, can be found in the Browser for Large ASTronomy Archives (BLASTA; http://ledas-www.star.le.ac.uk/usno/USNO/) at the Leicester Database and Archive Service (LEDAS), the VizieR Catalogue Service (http://vizier.u-strasbg.fr/) at CDS, Strasbourg, and the NASA/IPAC Extragalactic Database (NED; http://nedwww.ipac.caltech.edu/) at Caltech. In contrast, solar observations generally include numerous simultaneous observations over a larger range of wavelengths and deal with a greater number of individual images. In addition, users have a broad range of needs, covering time scales from minutes to decades, and layers of the solar atmosphere from the visible surface out to the extended heliosphere. Therefore, a specialized search interface, or rather interfaces, that is capable of intuitively presenting the whole variety of solar datasets will be developed. This work will rely on established methods for interface design and extensive user feedback to create an intuitive portal into the rather complex web of solar observations. This interface will also rely on user personalization and intelligent agents to tailor the response of the grid to each individual's needs.

Visualization techniques that could be of particular interest to this project are used at the CDS in Strasbourg in the Aladin Interactive Sky Atlas (http://aladin.u-strasbg.fr/aladin.gml) - this site allows interactive selection of a region of sky through a Web interface. The EGSO will expand these techniques to allow a user to: a) select a region of interest on a synoptic full-disk image (in a choice of wavelengths, close to a time of interest), b) determine what types of observations match defined search criteria, and c) select desired observations for processing and retrieval (using quick-look imagery to help with the final choice of data). The new range of requests that will be made possible by having numerous datasets federated into the grid will also necessitate the development of novel methods of visualising these extended and overlapping datasets.

The optimal location for the distributed catalogues developed for this project, and search engines that will use them to select and retrieve data, will be resolved during the project. Search engines tend to use two approaches: the search engines and catalogues wholly located on a remote server accessed via the Web, or a local (down-loaded) search engine using imported catalogue information, again with a web or web-like interface. The design of catalogue for the EGSO is intended to allow either technique. The fragments of the catalogue can either be stored where they are created (at the data centres), in centralized locations (two or three selected nodes), or can be down-loaded and searched locally. When network congestion is a problem, being able to download those portions of the catalogues covering the time interval of interest means that repeated searches in the same temporal range do not require continuous network access. Some centralization, or at least the caching of popular intervals, might be beneficial when gathering the catalogue fragments and this option will be studied. Application of advanced methods for load and bandwidth balancing will be implemented to allow for automatic optimization of the grid's efficiency and uniform utilization of grid resources. With a distributed catalogue, it will be much easier to add additional catalogues since there is no requirement for them to be ingested into a centralized location. New data sources can simply be added by lodging a *"locator block"* in a central repository declaring what new data and catalogues are available. The EGSO project will provide the capability for multiple routes to search the catalogue, leaving users to tailor their own

work environment. The EGSO will utilize, where applicable, the techniques for caching and automatic data replication, developed by such state of the art projects such as Akamai and FreeNet.

Tracking the usage of different portions of the EGSO datasets as they are mirrored and cached throughout the grid will be an intriguing problem. Keeping track of data usage, computing resource utilisation, and other parameters will be important for monitoring the behaviour of the grid and defining usage metrics. It will also be important, however, to provide reports back to the individual data providers about the use made of their data. This is important to show the usefulness of the grid itself and the interest from the community in the observations being obtained.

The final crucial innovation of the EGSO project that will provide innovation will be in the middleware used to federate the data centres. The project needs to provide access to a wide range of users, some of whom it will be difficult to authenticate with any degree of confidence. The middleware will control access to both public and proprietary data, and provide the mechanisms to process complex data studies at the data source while ensuring the security of the node and integrity of the Grid as a whole. The Globus middleware toolkit (produced by the Argonne National Laboratory, near Chicago, see URL http://www.globus.org/) is one example of a toolkit that could be used for this project. It will support many of the activities needed for the EGSO, while incorporating strong authentication through public-private dual key system. To accommodate the diverse solar user community, a minimal level of access for unknown or anonymous users will be implemented - this will afford the necessary access to the freely-available data, but provide complete control over the use of other resources. The processing of data using standard programs can easily be accommodated, but the ability to process large blocks of data from particular instruments on remote nodes will require careful control over user access in order to ensure the integrity of the system and judicious use of computing resources. The addition of these capabilities will provide extremely useful extensions to existing middleware, with applications in other diverse and scattered communities.

Tools for accounting, resource allocation, and creating usage metrics will be implemented, drawing from existing advanced tools developed by other data grids and applying their use to the case of a more diverse collection of nodes in the grid. These tools will have to take into account the possibility that the same data will be replicated at multiple sites around the node. It is expected, in addition, that many producers of data who release their observations into the grid will require usage statistics for their data in order to justify their participation and demonstrate to funding agencies the utility of their work. The required information will be made available using an interactive, near real-time application to be used both by monitors of the grid and individual partners.

Solar physics has in recent years developed a capability beyond that of most other fields of astronomy. This is the SolarSoftWare (SSW) system - a data analysis environment, and tree of analysis software that the solar physics community around the world uses and contributes to (see Web URL http://www.lmsal.com/solarsoft). The SSW system provides all the programs necessary for processing a large number of space and ground-based instruments. It will form the basis of the software used both for automatic reduction of the data before delivery to the user and for the additional processing done by the user after receiving the data. These programs are widely accepted and freely available, but have been limited by the difficulties in finding, identifying, and retrieving the data of interest. *It is this latter problem that will be addressed by the innovations of the EGSO.*

Relationship to other Grids

There are other grid projects currently underway that are complementary to this project. One intriguing example is the DataGRID, which is developing grid technologies for large and homogenous projects such as the Large Hadron Collider at CERN. While EGSO will have to deal with a larger and much more diverse range of data sources, we will, as necessary, study and implement the solutions developed for the DataGRID. Other grid projects, such as EuroGRID are more concerned with the linking of the High-Performance Computing centers, but it will be worthwhile to consult with this project to have further information about the lessons they have learned, especially in the are of data resource management..

In the United Kingdom, through funding provided by their research council (PPARC), the AstroGrid consortium is trying to create a Grid of astrophysical data centres within the country. The AstroGrid project should start in mid-2001, ensuring that study of many facets related to the choice of middleware and other required products will be well in hand by the time the EGSO team starts their work.

The AstroGrid consortium are part of the team that has submitted a proposal to the EC for the Astrophysical Virtual Observatory (AVO) under the IST-2000 Framework V programme. This project is oriented towards astrophysics but has many infrastructure requirements that are the same as the solar physics community. There is a similar initiative in the United States to create the National Virtual Observatory (NVO) to link together North American nighttime datasets.

In Italy, the nationally funded SOLAR project is underway with the goal of linking together all the Italian solar data resources, both from ground-based observatories and space instrumentation. This project has begun to confront problems such as metadata standards and creating virtual links among distributed databases that will be more rigorously addressed in the framework of the EGSO.

In the United States, spurred on by the recommendation of an National Research Council blue-ribbon panel to produce a "collaborative ... distributed data archive", several groups submitted a proposal for the development of the Virtual Solar Observatory (VSO). The EGSO project will ensure that the European solar physics community plays a key role in this work and that the different efforts proceed in tandem.

There are other projects that are focused on creating the definitions and structures necessary for creating digital libraries of scientific information, for example ARION, An Advanced Lightweight Architecture for Accessing Scientific Collections. The EGSO project will rely on the experience gained from these groups in the definition of metadata standards and innovative access methods.

6. Community added value and contribution to EU policies

The EGSO project aims to federate solar data archives together and provide the tools to select and exploit these data. In order to answer the most difficult problems of solar physics, scientists will need to perform studies that utilize many datasets covering a large range of wavelengths or times combined into a single overreaching view of the dynamic solar atmosphere. Since no single country possesses data covering the entire wavelength or temporal range, to guarantee access to the necessary observations this federation needs to be made at the European level.

Europe has access to a wide range of space-based data because its research teams have competed successfully at an international level and have helped build instruments on most of the spaced-based observatories in recent years. The UK has archives of most of these space-based observations, but has no ground based observatories of its own; France and Italy have their own ground based observations, but more restricted access to the space-based observations; other countries are even less fortunate. Many ground-based observatories regularly obtain valuable data, but because these observations are not visible to the global community, the data remain greatly underutilized. By linking these observatories' data together with high profile archives of ground-based or satellite archives, these data resources will be made much more valuable and in demand by solar physicists around the globe. By combining together these solar data in a pan-European data federation, scientists will have easier access to these data and can more rapidly and efficiently perform their research.

Ideally, from the standpoint of having continuous time coverage of ground-based observations, the EGSO will be part of a global federation of solar observations. Weather, the seasons and the rotation of the Earth mean that at best Europe has solar observations covering half of the day - to cover the other half requires access to ground-based data from countries on other continents. In the United States, a pre-proposal for a Virtual Solar Observatory (VSO) was recently submitted to the National Science Foundation – although some of its objectives differ, in many respects this is the US counterpart of the EGSO. Many European groups (including those involved in EGSO) have long established links with the institutions involved in the VSO proposal, and two institutions, the Solar Data Analysis Center (SDAC) and National Solar Observatory (NSO), are included in this proposal as non-funded members of the consortium.

Through the Yohkoh (Solar-A; launched 1991) and Solar-B (launch 2004) missions, Europe also has strong ties with the Japanese solar physics community. By discussing the details of the project with the US and Japan, the EGSO hopes to ensure that groups around the world are working in parallel and that even better access to the data will be possible for everyone in the future. Europe is in a position to take

a strong lead in this area since it is not evident that funding will be available in the US or elsewhere at this time. The Grid initiatives in the UK and Italy will help underpin some of the activities of the EGSO and ensure that the project can be completed within the projected time-scales.

Building on existing Expertise

The EGSO combines groups from across Europe involved in science and information technologies with the objective of creating an accessible federated solar data archive.

The data centres in the UK, France and Italy provide access to a large fraction of solar data sets available today. In almost all cases, the institutions involved not only host the data, but are also centres of expertise in interpreting the data. Many of the groups involved in building or operating instruments the many space-based observatories are located at, or have close links with the centres and are able to provide in-depth assistance to any users. Researchers at the centres in France and Italy hosting the ground-based data are frequent users of the facilities that provide the data and are expert in interpreting the observations. Also, several of the centres (e.g. MSSL and UPS-IAS) have a detailed knowledge of how to access and accumulate data from other sites using middleware such as mirror (an FTP package run under Perl). The centres in Italy and France have used relational database management systems (RDBMS) to manage their archives and are in a position to leverage that experience in the data grid where applicable.

Two of the data centres in the EGSO consortium, in Torino in Italy and at UPS-IAS in Orsay, France, were established using IT expertise. The Observatory of Torino hosts of the Italian copy of the SOHO data, but the Department of Automation and Informatics of the Politecnico di Torino working closely with the Observatory to establish the archive. Similarly, the UPS-IAS archive was established by a team mainly composed of computer scientists. This IT expertise is augmented within the consortium by the inclusion of the Department of Computer Science of University College London, and the Department of Cybernetics of the University of Bradford. In order to create the Solar Feature Catalogue, the expertise of a group involved in making and interpreting observations (Observatory of Paris-Meudon) is combined with that of a group expert in image analysis (University of Bradford) to produce a reliable tool that will assist researchers.

The Need for a Unified Catalogue of Solar Observations

In order to enable comprehensive searches over reliable metadata, a more unified catalogue of available observations is needed. This ideally would represent all data in a common self-describing format, but at a minimum must remove all dependencies on ancillary files and present a more consistent set of keywords to the catalogue search engines.

To understand the scale of the problem, the SOHO mission provides a useful case study. Whereas the data from the instruments on the Yohkoh mission were handled as one dataset and processed by a common reformatter, SOHO is a PI (Principal Investigator) led mission and the data for the various instruments were handled independently. Although the FITS format was adopted throughout the SOHO mission, the way in which files were created, and the keywords included, depended on each instrument teams. The CDS instrument probably has the most detailed catalogue of observations, but it is also one of the more difficult to use because of its complexity. In comparison the LASCO catalogue is very simple, consisting of an ASCII formatted table in month long blocks; the EIT catalogue is also in ASCII, but is mission long. SUMER and MDI are only catalogued within a general SOHO database.

The catalogues for the Yohkoh instruments are similar to each other, but are a totally different format to the SOHO catalogues; the TRACE catalogue is different again. The Yohkoh observing logs are in weekly files of a Yohkoh specific format, while the TRACE catalogue consists of daily files in a generic SolarSoft format (see URL http://surfwww.mssl.ucl.ac.uk/sswdoc/). An added complication of the Yohkoh SXT observing logs arises because the image coordinates are expressed in terms of CCD pixels - this requires access to the mission spacecraft attitude files before the heliocentric coordinates of the images can be determined. Complicating the issue is the fact that there is no single metadata standard to which all these missions adhere, resulting in conflicts and inconsistencies among the catalogues. Without the definition of a basic metadata standard, this problem will continue to manifest itself in future missions.

When ground-based observatories are considered, the picture is even more confused. Several ground-based observatories have made their data available on-line, but complete catalogues describing exactly what observations were obtained are not always readily available. Many ground-based observatories, less constrained than space missions, have not maintained a full record of operations, have not used consistent file formats for storing the data, and have not followed any recommended set of standard keywords. All of which further complicates access to the data and comparison with other datasets..

In reality, for the EGSO project to work, a unified observing catalogue would require a much smaller set metadata than that stored in many catalogues. Experiment teams often include information that is of interest to them, but redundant for the searches used to identify suitable datasets. In order to make the observing catalogue transportable and accessible, it is essential to try to remove all dependencies on other files, express it in a way that is self-describing (e.g. XML), and quantize it so that fragments covering relatively sort time intervals and individual instruments are easily available. Currently, many of the catalogues require proprietary software to interrogate them (e.g Oracle or Interactive Data Language). The unified catalogue can build on the experience in constructing catalogues in the past, but should not be tied to any particular product. This can be accomplished through the use of free-software alternatives (e.g. PostgreSQL or Ada,). Of course, the catalogue will be made accessible to searches from any type of client software.

EGSO will develop standards for both observing catalogues and data file headers that will simplify data access for new data if adopted by future space and ground-based observatories. By confronting the differences between existing observing catalogues at the catalogue level, we hope to minimize the effort needed to include data archives in the EGSO and thus maximize the data available to the solar physics and related communities

7. Contribution to Community Social Objectives

Improving Quality of Life

There are a number of areas where the creation of the EGSO will improve the quality of life of people in Europe. Clearly, for existing users of solar data, this project will dramatically improve the their access to the vast range of data available. A consequence of this is that researchers should become more effective in their work. Less time is wasted trying to find data and more spent using them to solve some of the problems of solar physics. This improvement is not just a question of efficiency, but also of scientific quality. Having a broad number of observations available through the EGSO will mean that the data found by the scientist will be the best data available that is suited to the scientists research topic. Even more exciting is the possibilities that will be afforded by this data grid to allow scientists to pose an entirely new set of questions that until now were almost impossible to reasonably address. These groundbreaking questions will be those relying on large numbers of overlapping datasets treated in a statistical manner. As the data grid is adopted by the solar physics community, we expect the productivity of the scientists to increase (as measured, for example by the number of publications), which will ensure that research funded through national science programmes becomes more cost effective.

Another benefit from enhanced access to solar data affects the quality of life of a much larger fraction of society. In recent years, we have begun to realize the potential threat that space weather poses to industrialized societies. We are becoming increasingly dependent on technology in our every day lives, with computers, communications and transportation becoming ever more important to us. The effects of space weather on power distribution grids are well established, but the influence on sophisticated electronic circuitry could become more important as their usage and level of integration increase. Most space weather phenomena originate from different types of activity in the Sun's atmosphere. Large solar flares can produce energetic particles that can (within minutes) affect spacecraft in near-Earth orbit. Coronal mass ejections (CME's) - clouds of plasma with an embedded magnetic field that are ejected from the Sun with velocities that sometimes exceed a 1000 km per second - can affect the Earth's environment tens of hours after the event causing geomagnetic storms and other effects. These CME's are known to result in communications (including GPS) blackouts, power outages, as well as again affecting spacecraft and astronauts in near-Earth orbit - at least one multi-million dollar communications spacecraft has already been lost due to a solar eruption.

Space weather forecasting is currently at a primarily reactionary stage, relying on either now-casting (a geomagnetic storm is in progress) or near-casting (a mass ejection has just passed a monitoring spacecraft an hours travel time from the Earth). In order to gain more that the one hour warning that near-casting might give (dependent on the velocity of the plasma cloud), we need to understand why these events occur back at their origins, in the solar atmosphere. The problem is that until the orientation of the magnetic field in the plasma cloud is known, the effects on the Earth cannot be reliably predicted. Currently the first time this orientation can be measured is when the cloud passes a spacecraft located a 1.5 million kilometres closer to the Sun than the Earth. If solar physics can provide the clues that will allow true forecasting tens of hours before the Earth is likely to be affected, many of the potential problems could be mitigated. Improved access to solar data will help in the search for this crucial information and would also help the forecasting of events in the future, reducing their effect on society.

Contribution to Improving Employment

As discussed in the previous section, enhanced access to solar data will make analysis of these data less onerous. This could make a considerable difference to smaller groups that have problems supporting the infrastructure needed for research. If solar data were available "on-tap", since the effort required for infrastructure would be reduced, the number of people that could be devoted to analysis could be increased. More opportunities could therefore be created for young scientists to continue their research within Europe, encouraging them to stay rather than seek positions abroad.

The framework that EGSO provides will also be important for storage of data from new instruments in the future. With an established and accepted set of standards and guidelines to work to, project effort that would have otherwise been devoted to data archiving will be available to deal with the intricacies of building the complex instruments needed for modern research (particularly on a space-borne platform). In essence, an established product (the EGSO) removes the need to repeatedly devote manpower to solve the same problem again and again. If the standards of EGSO become accepted worldwide, this represents a significant amount of effort that can be devoted to more demanding activities.

The simplified access afforded by the EGSO also provides the platform on which groups might build more sophisticated ways to present solar data to the general public. Today, public outreach is an increasingly important part of all science and technology projects. Research is expensive and it is essential that efforts are made at every opportunity to demonstrate how important and exciting it can be. The Sun is much more dynamic than most people realize and it has a greater effect on their lives than they imagine. The EGSO project will produce some outreach material itself, but the opportunities it presents for others (e.g. schools) to work on this topic could produce some exciting public resources in the future.

8. Economic development and scientific and technological prospects

The European Grid of Solar Observations (EGSO) will design, create and deploy a grid that federates solar data archives across Europe, and provides the tools to identify and retrieve observations for more detailed analysis. Although a few of the problems that must be solved to create the EGSO could be unique to this particular application, many are common to any grid that attempts combine diverse, widely scattered scientific archives into a data grid, and some solutions will have wider applications in the developing e-science and e-commerce communities.

The EGSO will federate a set of heterogeneous solar data archives located in sites across Europe. It will initially involve institutes in four countries that have shown a concerted interest in solar data archiving, but will be designed to easily include any other groups who wish to make their data available, in Europe, or overseas. The final federation realized by this project will initially include around 10 TBytes of data, in several tens of thousands of files, but this will increase rapidly as instrumental data rates increase and other institutes are added.

A requested piece of data could be located anywhere in the EGSO and some datasets will be duplicated on multiple nodes. This will increase the resilience of the whole archive, reducing the Quality of

Service (QoS) that must be provide by individual nodes, and allowing user sites to choose different sources depending on their location within the Internet. Tools under development for other network applications that determine the best choice of alternate sources, depending on network traffic and availability, will be adapted for use in the ESGO - this will allow the source selection to be adjusted dynamically as the network changes. Since some smaller institutes that are unable to support the elevated access the grid might imply, the mechanisms to source their data through larger centres need to be developed and will have applications in other data grids. Popular datasets will be cached at a few central location to optimize the flow of data - this requires a federation-wide awareness of data usage and location.

The users of the EGSO will be widespread and diverse. Access must be designed so that it is simple but effective for users with little understanding of the instruments, while being scalable for more advanced users. The large user base presents problems of both identification and security. Some users will be widely recognized within the European solar community, while others will not be immediately known to the partners of the grid. All users need to be provided with good access, but controls must be enabled such that the integrity of the grid and its nodes is guaranteed.

Data that are in the public domain could be mixed with proprietary data and, in order to respect the intellectual property of these data, access control must only permit the use of these file by authorized individuals. The system must be able to dynamically define access lists for protected files in order to properly cope with datasets that are obtained and made available in near real-time (a capability needed for the solar forecasting and space weather communities). Although there has been a move towards an open data policy on recent small space missions, provision of this type of controlled access will provide greater flexibility for future expansion, and encourage institutions that practise a regime of more restricted access to include their datasets in the federated archive. Degraded quick-look images will be available as part of the catalogue search in order to stimulate collaborative use of the data. This type of highly flexible security system, which is able to follow the data as they are replicated at multiple nodes, will be important for all scientific fields, but also for the protection of intellectual property of digital data in numerous other applications.

Some datasets should optimally be reduced at the data source to minimize the volume of data to be transmitted - included in these are datasets that require computer intensive image reconstruction, and where selected datasets are embedded in much large data files. This implies the need for users to execute software on remote nodes to produce metadata products. Standard extraction or image reconstruction software will be made available and, drawing on the consortium's experience in data reduction and the expertise from other data Grids, is expected to present few problems. However, when users wish to examine some unique aspect of the data over an extended time interval (perhaps years), or large samples, the ability for users up upload un-screened software to a remote node must be provided. This type of statistical solar physics could yield important information replated to long term trends in the evolution of solar activity, but is much more demanding on the EGSO - it must be done in a way that does not compromise the security of the data source, or any other node in the federation. The problem of working the federated grid through site firewalls in a way that is secure and acceptable to existing site security regimes will also have to be solved. Correctly handling user supplied programs, even in such simplified languages as JavaScript, is an ongoing problem in system design. It will continue to be an issue, be and even more important to solve in applications running on an always-on grid connecting a large number of systems. This project will begin to examine some of these issues in a real-world application of a Grid architecture.

To optimize the search for data within the EGSO, a new format of catalogue of solar observations must be developed. This will simplify the task of creating the EGSO and will provide metadata standards and guidelines for future instrument and space missions. The metadata standard, by removing one of the main hurdles to integrating multiple datasets, will be essential in the future for creating a whole new range of powerful applications in solar physics. It will also define a standard entry point into the solar data archive for applications outside of the field of solar physics, for example astrophysics, solar-terrestrial physics, climatology and space weather. The catalogue search and visualization tools will scale for different user, and involve the interactive use of imagery and catalogue information in order to initially select the data, and the production of quick-look imagery to confirm the selection. The solar feature catalogue will require new techniques of image recognition at many wavelengths that could be used in other applications (e.g. remote sensing).

The creation of the EGSO will significantly improve access to the wealth of solar observations available in Europe. This will be a boon to the European solar physics and solar-terrestrial physics communities and will greatly enhance the quality of research supported by money provided by funding agencies in the individual countries of the European Union. By combining large, high-profile datasets (often from satellite missions), with the numerous less well known datasets obtained from the dozens of solar observatories in Europe, the visibility of the latter data will dramatically increase. By virtue of being combined with observations in other wavelengths, either from space-based instruments or other terrestrial observatories, their usefulness and scientific value will also be augmented. In some cases this may provide the impetus to continue operations of an observatory in times of restricted budgets. This reasoning is especially true for observatories from associated states or the former soviet union, who obtain world class data but have limited accessibility and visibility to much of the community. There is much interest from many of these small groups to participate in a federation of solar archives.

By raising the visibility of European solar observations, the expertise developed in Europe in these fields will increase. This will allow these groups to play a greater role on the international stage as well as be involved in additional collaborations with other scientists interested in using their data. European scientists will become knowledgeable in the new work involved in statistical solar physics and will be in an excellent position to provide some deeper insights into the sometimes puzzling workings of the Sun.

As part of the structure of the EGSO, a User Group will be created that will provide feedback to the consortium developing the grid - this will allow us to make changes in response user experience and concerns. We plan to hold sessions to consult users at some of the regular scientific conferences that involve the solar physics community, e.g. the European Physical Society, the European Geophysical Society, and the meetings organized by JOSO (Joint Organization of Solar Observations) and CESRA (Community of European Solar Radio Astronomers). In addition, members of the consortium coming from a computer science background will present the lessons learned in establishing the grid to the IT community.

The EGSO will provide the basis for a grid that could be extended across the Atlantic, and possibly globally. It will not be static, but will be extended with each new project both inside the European Union and elsewhere. The software constructed for creating the EGSO will be made of easily reusable components to allow other institutes to quickly add their datasets to the grid with a minimum of effort on both sides. It will be possible to integrate space missions planned by ESA (e.g. Solar Orbiter) and other agencies (Solar-B, STEREO, etc.) into an existing framework of data archiving and analysis with all the benefits this brings.

As far as possible, it is planned to utilize open source software to implement the EGSO. All software tools developed for the EGSO will also be open source to extend the existing software and encourage further development of the grid. This will encourage involvement in, and acceptance of the EGSO and be one of the routes by which the results of the project are disseminated. Having the software developed by the EGSO widely distributed and adopted by other groups will establish the partners as key participants in this field. In the later phases of the project, demonstrations of its capabilities will also be made at the scientific conferences mentioned above.

The computer science and other IT groups involved in the EGSO project will be well placed to use the experience gained in creating the federated archive for other other scientific and technical projects. The provision of widely used standards would allow the archiving of future space missions to be out-sourced to industry, with these groups participating in such contracts. Some of the expertise gained related to network access and security will have direct applications in many other areas.

9. Workplan:

9.1 General description

9.2 Workpackage list

9.3 Workpackage descriptions

9.4 Deliverables list

9.5 Project planning and timetable

9.6 Graphical presentation of project components

9.7 Project management

10. Clustering

The European Scene

EGSO is a partner in the Accompanying Measure proposal GRIDSTART (IST-2001-34808), a cluster designed to serve as a focus of European Grid activities.

GRIDSTART (Grid Dissemination, Standards, Applications, Roadmap and Training) was submitted in October 2001, and is led by the Edinburgh Parallel Computing Centre (EPCC). The GRIDSTART partners comprise the co-ordinating partners of the following IST projects: AVO, EUROGRID, DATAGRID, DAMIEN, GRIDLAB, GRIA, GRIP, DATATAG, EGSO and CROSSGRID. These projects comprise the major part of IST-funded GRID activity. In GRIDSTART, the partners will collaborate to identify synergies and common actions across all these projects and other EU and nationally funded initiatives. They will work together in the consolidation of technical developments from these projects and the establishment of international standards. GRIDSTART will provide the framework supporting this collaboration and exchange of ideas through the establishment of a technical committee, a strong coherent standards activity and effective dissemination involving all projects, not only to the technical community, but also to potential early adopters.

Relations to other Grid activities

Europe has an emerging and active set of GRID projects. These are concerned with the development of the necessary software infrastructure to support industrial and scientific applications. At the European level, the major projects have already been identified in this proposal and comprise the partners of the GRIDSTART consortium. These application-based activities are complemented by the GEANT project that is primarily concerned with the provision of high-bandwidth interconnectivity. The GRIDSTART consortium will look to establish an early and appropriate dialogue with the GEANT project. The consortium is aware of the following national activities that will be able to benefit from and contribute to the activities of the GRIDSTART project:

AstroGrid and AVO, GridPP, SpaceGrid

11. Other contractual conditions

Some travel to the US because of US groups – is US an associate state??.
Attending GGF could be US or Japan

Appendix A - Consortium description

C6 Description of the Consortium

The EGSO consortium consists of the following partners:

1. University College London (UCL), **Coordinator,**
   Dept. Space and Climate Physics (MSSL) and Dept. Computer Science (UCL-CS).
2. Rutherford Appleton Laboratory (RAL)
3. Observatory of Torino (OATO)
4. Politecnico di Torino (Polito)
5. Observatory of Paris-Meudon (BASS2000-Meudon)
6. Institut d'Astrophysique Spatiale (UPS-IAS)
7. Astrium
8. University of Bradford (UNIBRAD)
9. Solar Data Analysis Center (SDAC-NASA)
10. National Solar Observatory (NSO-NOAO)

Many of the groups have worked together before. MSSL and RAL have been collaborating on project for more than 20 years and both have been involved with OATO and UPS-IAS since the start of SOHO (although they have work with Ester Antonucci since SMM days). Yohkoh and SOHO have brought several groups together in joint observing and data analysis projects, also involving the SDAC and NSO. The links with the IT groups have their origins in the need that the data groups discovered several years ago for that type of expertise.

The EGSO project aims to federate solar data archives across Europe together and provide the tools to select and exploit these data. The definition of the system architecture of EGSO will involve all the groups, including the archive groups (MSSL, RAL, OATO, BAS2000, UPS-IAS, SDAC and NSO), the IT groups (Polito, UCL-CS, UPS-IAS and UNIBRAD), and assisted by Astrium (who bring an industrial perspective to the consortium). SDAC and NSO will provide the international input to this design. This activity will include the selection of the middleware toolkit (e.g. Globus) needed to provide the access and resource control at the data centres when federating the data.

The initial EGSO federation will include data at four sites across Europe, providing a wide range of space and ground-based observations. The sites are MSSL (Yohkoh, TRACE and HESSI datasets), RAL (SOHO, TRACE dataset), UPS-IAS (SOHO and TRACE), and the OATO (SOHO and Italian ground-based optical observations). MSSL and RAL are both involved in the UK's AstroGrid project, with OATO and UPS-IAS involved in similar projects in Italy and France.

The IT groups will also work with the archive groups to define a self-describing format for the unified observing catalogue that is completely transportable and extensible, and will formulate the requirements of the search engine and visualization tools needed to interrogate the unified observation catalogue. UPS-IAS and Polito have already developed expertise in this when deploying their SOHO archives. While the data centres are generating of the fragment of the unified observing, the IT groups will create the search engine and visualization tools – the programme calls for these components to come together with the federation activity during the third year of the project. Meudon will work with NSO and the Observatory of Naples (working through OATO) to resolve of the problems associated with cataloguing the ground-based observations. Similarly, the SDAC will collaborate with MSSL, RAL and IAS on cataloguing the space-based instruments.

An innovation of the EGSO will be the creation of a new type of catalogue, the Solar Features Catalogue – this will provide a new way of searching the data based on particular types or classes of phenomena. Paris-Meudon has an established expertise in feature recognition at optical wavelengths and currently creates synoptic maps and tables of filaments, sunspots, etc. They will work with UNIBRAD to extend the existing software to include EUV, X-ray and radio wavelengths and create a reliable tool that can easily be used by researchers. These will then be used to create the catalogue with the space and ground based observers assessing the performance of the algorithms.

As the many components of EGSO are brought together during the system integration phase, MSSL and Plolito will work with Astrium to ensure the components are produced to standards necessary for deployment in other applications outside the EGSO project.

C7 Description of Participants

University College London (UCL):

University College London (UCL) is the largest of over 50 colleges and institutes which make up the federal University of London. UCL is one of the pre-eminent universities in the United Kingdom with a strong emphasis on research. In 1998, its income was £185M, with and expenditure £179M.It has more than 5000 staff and 9000 undergraduates. There are more than 3000 postgraduate students covering all areas of science, engineering and medicine – in recent years a set of five hospitals have become linked to UCL.

UCL has been actively participating in international programmes, in particular EU funded projects. The University has not only contributed with expert advice in the writing of work programmes for the Framework, but has been very successful in bids. In recent years, UCL has in fact outstripped the Universities of Oxford and Cambridge in securing EU funding for collaborative research - at present the University has more than 350 EU funded research programmes under FP4 and FP5. While UCL participates in all aspect of FP5 (with 180 funded projects) the University has been most successful in the IST area with more than 80% of submitted proposals leading to contractual negotiations.

The ***Department of Space and Climate Physics*** (also know as the Mullard Space Science Laboratory, MSSL) is located away from the main campus in the Surrey Hills, approximately 40 miles to the south-west of London. The Department undertakes frontline research in Climate Physics, High Energy Astrophysics, Solar Physics, Space Plasma Physics and Photon and Particle Detection Systems. As the UK's largest university space physics institute, it includes professionally-staffed electronic, mechanical, computing and software engineering groups and has designed and built instruments for more than 30 orbiting spacecraft and 200 sounding rockets. There is a total staff of 140 and 20 PhD students; the Department offers teaching to PhD level in each of its research disciplines and also offers two Masters degree programmes. . MSSL has long been recognized a centre of excellence in the provision of space instrumentation, and in exploiting the data returned by the missions it has participated in.

Within the Solar Physics Group, MSSL has been involved in missions dating back to the 1970's with OSO-5. Since then it has been involved in the NASA Solar Maximum Mission (1980-1989), NASA's Spacelab-2 (1985), the ISAS Yohkoh mission (1991 to date), and ESA/NASA Solar and Heliospheric Observatory (SOHO; 1995 to date). Future missions include Russia's KORONAS-F mission (2001) and Japan's Solar-B missions (2004). Of these, MSSL was the PI group on SMM and Yohkoh. In 1994, the Yohkoh Data Archive Centre (YDAC) was established by the Laboratory – this was, and still is the only archive of Yohkoh data in Europe. In 2000, the YDAC was renamed the Solar UK Research Facility (SURF), a mark of its expanding role as a major solar archive.

MSSL will contribute the Yohkoh archive and the TRACE, and HESSI datasets to the EGSO and will be one of the four sites used to initially establish the data Grid test bed. MSSL is the *coordinating partner* for the EGSO project, and will be involved in all aspects of the project.

The ***Department of Computer Science*** evolved in 1980 out of the Department of Statistics and Computer Science which was formed in 1975. Since then, the Department has enjoyed a period of continuous growth and currently has 30 academic staff (including 9 professors) and some 45 research staff and around 45 PhD students. In addition to its data communications and distributed systems strengths, research in the department now also includes artificial intelligence, neural networks, software systems engineering, intelligent systems, multimedia, communications, imaging, computer vision and virtual reality.

The Computer Science has almost two decade's worth of experience in Framework funding, having participated with projects in RACE, ACTS, ESPRIT and the IST programmes. Under the FP4 ACTS Programme: Propsect, MISA and VITAL; under ESPRIT: ITEX and HIPPARCH. Under the FP5 IST Programme: IST-1999-10357, FORM; IST-1999-10299, ANDROID; IST-2000-25131, UWA; IST-2000-26084, ET Cluster.

Rutherford Appleton Laboratory (RAL-CCLRC)

The Space Science and Technology Department (SSTD) of the Rutherford Appleton Laboratory (RAL) has 200 staff dedicated to supporting the programmes of the UK's Particle Physics and Astronomy Research Council and the Natural Environment Research Council. It has also undertaken a large number of contracts for agencies, industry and other commercial customers. RAL provides world-leading research and technology development, space test facilities, instrument and mission design, and studies of science and technology requirements for new space missions.

Much of the Department's work is in collaboration with UK university research groups and a range of institutes around the world. Most of these collaborations have been set up to support the European Space Agency (ESA) and NASA missions, though RAL is also working on projects with other countries and organisations including Australia, Japan, Morocco, Pakistan, Russia, Spain and the European Union.

RAL is part of the Central Laboratory for the Research Councils (CLRC) which was formed in 1995 to support the programmes of the other research councils. SSTD is a partner member of the British National Space Centre (BNSC).

RAL has been particularly active in solar physics, providing space hardware, software and data handling and archive facilities as well as scientific research for many missions. These comprise the NASA Solar Maximum Mission (1980-1989), NASA's Spacelab 2 (1985), the ISAS Yohkoh mission (1991 to date), and ESA/NASA Solar and Heliospheric Observatory (SOHO; 1995 to date). Future missions include the USAF Solar Mass Ejection Imager (SMEI, to be launched late 2001) and also has involvement in developing missions such as Japan's Solar-B and the NASA STEREO mission. In many of these missions, RAL has taken the Principal Investigator role.

RAL will contribute the SOHO archive and TRACE dataset to the federated archive and will be the one of the four sites used to establish the data Grid test bed.

Osservatorio Astronomico di Torino (OATO)

The Astronomical Observatory of Torino is one of the 12 astronomical and astrophysical observatories operating in Italy – all are government research organizations. Within the present year the twelve Italian observatories will be combined into a single research institution, the Istituto Nazionale di Astrofisica (INAF). The observatory personnel consists of 28 researchers, 21 technical staff involved in the research projects of the observatory, 11 administrative employees and 9 employees in the service area (library, etc.).

The research is carried on in 5 disciplinary groups: Solar Physics, Planetology, Stellar physics, Extragalactic physics, and Astrometry. The Solar Physics group, consisting of 8 persons, is primarily involved in space solar physics and in particular is part of the experimental team of the Ultraviolet Coronagraph Spectrometer (UVCS), successfully operating onboard of the Solar and Heliospheric Observatory (SOHO) since 1995. The UVCS is a US-Italian joint programme, supported by NASA and the Italian Space Agency (ASI). The instrument is detecting for the first time the ultraviolet emission of the extended corona of the Sun and is providing new information on the regions where the solar wind is accelerated.

The European Space Agency (ESA) has granted the right to host one of the three European SOHO archives in Torino (letter of the ESA Directorate of Science, DSci/RMB/SO/VD/mr/val/2533, 20 March 1995). The Torino SOHO archive, SOLAR (SOHO Long-term ARchive), is currently part of an Italian test bed of solar archive federation. This includes the ARTHEMIS archive at the Astronomical Observatory of Naples, and SOLRA, the radio data archive of the Astronomical Observatory of Trieste (including data from the Kanzelhohe Observatory in Austria). The 3-node network, SOLARNET, will make available to the national and international community from a single interface the SOHO data, the data of the French-Italian telescope, THEMIS, and the radio and optical data obtained by the various Italian solar observatories (Rome, Naples, Catania, and Trieste).

Within this proposal, OATO also represents the contributions of the observatories of Napoli and Trieste - the three institutes (combined into a single institute under the INAF) will contribute the combined datasets of SOLARNET, and the expertise developed in creating this federation to the EGSO. The

Observatory of Naples, in particular, will be one of the four sites used to establish the data Grid test bed and will be heavily involved in the definition of the architecture of the EGSO.

Politecnico di Torino (Polito)

The Department of Automation and Informatics (DAUIN) of Polito worked closely with the Observatory of Torino to establish one of the European copies of the SOHO archive (SOHO is an ESA/NASA joint space mission for solar observation still in progress). It has also been involved in the design of the visualization software for UVCS data (UVCS is an instrument part of the scientific payload of SOHO, co-funded by NASA and Italian Space Agency).

The contribution of Polito will be in the areas of design and implementation of the search and visualization tools, and in the design and implementation of system management mechanisms (e.g. security, caching, middleware). The Department of Automation and Informatics of Polito is very active in the 5th Framework Programme, as it is involved in more than 10 contracts currently in execution, in different areas. Research activities currently carried out within DAUIN and relevant for this project include: security protocol verification, techniques for the visualization of scientific data, computer networks and protocols.

Observatory of Paris-Meudon (BAS2000-Meudon)

The Observatory of Paris is the biggest astrophysical institute of France. Nearly the third of the French astrophysicists work there. All astrophysical fields are present in Observatory of Paris. Observatory of Paris is made up of 3 sites: One in Paris, one in Meudon, near Paris (known as Observatory of Meudon) and one in the center of France: the radio astronomy observatory of Nancay. Before being included in Paris Observatory, the Meudon site was devoted to the observation of the Sun. The first photo plate of the Sun was taken there by Jules Janssen, in 1870. Photos of the Sun have been taken in Meudon regularly since 1909. A very complete archive of nearly 70,000 photographic plates is archived in Meudon.

Every day, weather permitting, the spectroheliograph of Meudon takes two plates at chromospheric wavelengths (H alpha and CaII K3), one at photospheric wavelength (CaII K1v), and one at CaII K3 with the disk of the Sun overexposed in order to see prominences at the limb. The site has approximatly 300 sunny days per year. When it's not possible to take pictures of the Sun, the archive is completed with spectroheliograms from Coimbra (Portugal); so it's possible to build synoptic maps of the Solar activity. Spectroheliograms are scanned every day and all data are on-line since 1998 (in free access). The Meudon spectroheliograph will be soon updated with a CCD camera so the data will be directly on-line, without human intervention. And it will be possible to have more wavelengths (specially both wings of H alpha, giving the opportunity to calculate dopplergrams of the surface of the Sun). This update should be ended at the beginning or spring 2002.

The Solar Department of Paris Observatory, DASOP, is in charge of this archive and is responsible of the French full Sun database BASS2000 Meudon, which includes Meudon spectroheliograms, Nancay radioheliograph daily observations, Nancay decametric network observations of the Sun and Pic du Midi Observatory H alpha coronograms. DASOP publishes also synoptic maps of solar activity and was the main contributor in the design and construction of the Franco-Italian Solar telescope THEMIS (Canaries Islands, Spain).

In January 2002, DASOP and the space department (planetary and plasma physics, involved in many space missions, such as CLUSTER satellite) of Paris Observatory will join together and become LESIA (Laboratoire d'Experimentation Spatiale et d'Instrumentation Astrophysique). There will be in LESIA a total staff of nearly 200 and 20-30 PhD students. This new department will host specialists of all fields from solar photosphere to Earth magnetosphere (and even planetary magnetospheres).

Institut d'Astrophysique Spatiale (UPS-IAS)

The Institut d'Astrophysique Spatiale is part of the CNRS University Paris-Sud. The laboratory specializes in space astrophysics, grouping about 150 researchers, technical and administrative staff and 25 PhD students and post-docs. It covers 3 major scientific themes solar physics, planetology and physics of galaxies.

The Institute has been involved in many major space projects. Among these we can mention major missions from the European Space Agency:

- the Infrared Space Observatory (IAS have been in charge of the calibration of the camera)
- three experiments on the solar observatory SOHO (EIT, GOLF, SUMER were designed, built and calibrated at IAS),
- the XMM-Newton X-ray observatory (IAS was also in charge of the calibration of the cameras), - two instruments on the Cometary mission ROSETTA (CIVA VIRTIS calibration) and on the MARS-EXPRESS mission the spectral imager OMEGA,
- Planck surveyor (IAS is in charge of the High Frequency Instrument and the corresponding data processing) and ground-based and balloon borne experiments (PRONAOS, DIABOLO and ARCHEOPS) preparing Planck.

IAS plays a major role in the data reduction and archiving for several of these missions: SOHO, ISO (French data center), and in the near future the MARS missions. The development of these activities is a major priority of the Institute.

The espertise of the Institute for Data Centers is built on its major role in the SOHO data center and archive. IAS has provided facilities for several hundred of visitors who came to work on the solar data. The archive of the SOHO data (which is now operational) serves a significant fraction of the European Scientific Community. For ISO, The Institute contributed to the definition and building of the ESA archive especially for the ISOCAM instrument. IAS will contribute the SOHO and TRACE data to the EGSO, and will be one of the four sites used to establish the Grid test bed.

University of Bradford (UNIBRAD)

The **Department of Cybernetics** of the University of Bradford is a new department founded in 1997 and focussed on the vast opportunities for research and wealth creation in our era of computer networks. The critical challenge is to educate graduates to fill the thousands of job opportunities in small companies striving to respond to the global market of the Internet, the new automotive components required by vehicles that use networks of computers, and the new systems required to optimize health care delivery. The Department teaches degree in the following programs:

1. The Cybernetics and Virtual Systems (including human-machine interface, Internet and Communications, Robotics and Virtual Systems);
2. Cybernetics with Transportation Systems (allowing networks or computers to 'drive by wire');
3. Medical Cybernetics;
4. The Internet law and Society degree (a world first) giving understanding of both the technical and the legal complexities of Internet commerce.

The Department has a good record of holding research grants using image recognition for various industrial applications. There are the following research contracts: Flat light technology, funded by Sidefact (1987–1989; PI – Dr. S. Ipson); VME-based image restoration, funded by DERA (1992–1993; PI: Dr. S. Ipson); Automatic die bending machine funded by BUSM (1992–1993; PI: Dr. S. Ipson); and Controls for high speed machinery, funded by EPSRC/BUSM (1994–1996; PI: Dr .S. Ipson). Also there have been a few research contracts on the robotic telescope and its archive (1988–1990, PI: Dr. J. Baruch), public understanding of science (1993–1994, PI: Dr. J. Baruch). Currently, there is an EPSRC grant on the communication systems in transportation (PI: Dr. J. Baruch) and several other grants associated with cybernetic applications in industry. There have been several consultancy agreements held by the members of the department funded by BUSM (1995 and 1997, Dr. S. Ipson), and by Eureka (1998-1999, Dr. S.Ipson).

Astrium

Astrium is the new company formed in April this 2000 by the merger of Matra Marconi Space and the space divisions of Daimler Chrysler Aerospace. It combines the vast experience, resources and facilities of the major space companies in three of Europe's principal space nations, France, Germany and the United Kingdom. The company has interests in all space applications, Science and Earth Observation, Global Navigation, Telecommunications, Ground Systems, Military Programmes, Launch Vehicles and Orbital Infrastructure. Astrium has a long and successful association with universities and user communities interested in space science, and welcomes the opportunity to support them where appropriate.

For the EGSO, Astrium's areas of interest are in the definition and design of the project, and in the realisation of EGSO during the integration and testing of the components. Their involvement in related studies and projects such as ERSIS and Infoterra will be of considerable value in the design stages of the project, and they believe that some of the innovations that the EGSO project is proposing will be relevant to applications they have planned for the future. The EGSO project will be managed within the Earth Observation and Science Division of the Company.

Solar Data Analysis Center (SDAC-NASA)

The Solar Data Analysis Center (SDAC) is located at NASA Goddard Space Flight Center (GSFC) in Greenbelt, Maryland USA. The SDAC has it origins in the SMM Data Analysis Center that was established in 1981. In 1991, the SMM-DAC was reconstituted as the SDAC to signify its expending role. The SDAC and the National Space Science Data Center (NSSDC, also at GSFC) fulfill complementary, not overlapping roles in assisting the solar physics community. The SDAC is an analysis center: while it archives data, the emphasis is on analysis, with scientific expertise and facilities unavailable to many users of the NSSDC. It is part of a science branch at NASA-GSFC, and funded with MO&DA resources.

The SDAC has unique professional ties to other members of the solar physics community that allow it to provide such services as daily solar imagery, eclipse bulletins, the SOHO Web and anonymous ftp services, and the active Yohkoh, SOHO, and TRACE archives. Scientists and programmers working at the SDAC have contributed a significant part of the SolarSoftWare (SSW) tree of analysis software, and the master copy of the SSW tree is hosted at GSFC

Solar data dating back to OSO-7 are held at the SDAC, and currently more than 3 GBytes are available in on-line storage. Data sets include several instruments from Solar Maximum Mission (SMM), Compton Gamma Ray Observatory BATSE flare data, Yohkoh, TRACE and GOES soft X-ray photometer data.

The Solar Data Analysis Center is a non-funded member of the consortium. It will contribute its expertise in the archiving of space-based data, and will work with the other consortium members on defining and creating the space-based part of the unified observing catalogue. SDACS will also work with MSSL and RAL on Grid federation issues.

National Solar Observatory (NSO-NOAO)

The National Solar Observatory (NSO) is part of the National Optical Astronomy Observatory (NOAO) and has its primary headquarters in Tucson. NSO telescopes on Kitt Peak include the McMath-Pierce Solar Telescope Facility containing the world's three largest solar telescopes (1.6-meter main and two 0.9-meter auxiliaries), along with the Vacuum Telescope and the Razdow small solar patrol telescope. The National Solar Observatory also operates telescopes at Sacramento Peak, New Mexico, that include the Vacuum Tower Telescope, the Evans Solar Facility, and the Hilltop Dome Facility.

The National Optical Astronomy Observatory is funded by the National Science Foundation and operated by the Association of Universities for Research in Astronomy (AURA), Inc. NOAO was formed in 1982 to consolidate all AURA-managed ground-based astronomical observatories under a single Director. The observatories comprise Kitt Peak National Observatory, Cerro Tololo Inter-American Observatory, and the National Solar Observatory. Today, NOAO also represents the US astronomical community in the International Gemini Project.

NOAO's purpose is to provide the best ground-based astronomical telescopes to the US astronomers, to promote public understanding and support of science, and to advance all aspects of US ground based astronomical research. As a national facility, NOAO telescopes are open to all astronomers regardless of institutional affiliation.

The National Solar Observatory is a non-funded member of the consortium. It will contribute its expertise in the archiving of ground-based optical observations of the Sun and will work with the Observatory of Paris-Meudon and the Observatory of Naples to define the ground part of the unified observing catalogue.

CV's of Key people.

- **Dr. Robert Bentley, UCL (Project Manager)**

Dr. Bentley has been involved in solar physics since the NASA Solar Maximum Mission (SMM; 1980-1989) and. gained his PhD in solar physics in 1986. During the latter years of SMM, he was the PI for the MSSL part of the XRP instrument. From 1987, he was the overall Project Manager of the Bragg Crystal Spectrometer (BCS), an instrument carried on the Japanese Yohkoh mission (1991-present) - the BCS was built by an international consortium led by MSSL, and including RAL, the US Naval Research Laboratory, and the National Astronomical Observatory of Japan. In 1994, he established the Yohkoh Data Archive Centre (YDAC) and has been its Project Manager since. He has been involved in the SOHO, TRACE and HESSI missions and is the author of several analysis guides. One of the architects of the SolarSoftWare analysis environment, in 1996 he helped conceive the Whole Sun Catalogue. He also wrote the on-board code for the microprocessor of the ROSAT Wide Field Camera and the Yohkoh BCS.

- **Prof. Luigi Ciminiera, Polito (Technical Coordinator)**

The group at Politecnico di Torino (Polito) will be led by Luigi Ciminiera. Prof. Ciminiera is professor of computer engineering and Director of the Department of Automation and Informatics. He was responsible for the Polito contribution to the SCARAB project (within the 4th FP) that was successfully completed in 2000. Prof. Ciminiera served also as a reviewer for CEC-sponsored projects.

- **Dr. David Pike, RAL**

BSc 1972 St Andrews PhD 1976 UC Santa Cruz/St Andrews Involvement in solar physics began in 1991 when he spent a year at ISAS, Japan as the UK representative for the BCS instrument on the Yohkoh satellite. On return to the UK, he was appointed as software leader for the SOHO-CDS instrument. He was responsible for coordinating the design and implementation of the ground based planning and data analysis software by personnel from the UK, Norway and US.

Since the launch of SOHO Dr Pike has, in addition, been responsible for the scientific operation of CDS and for liaising with the user community to ensure the optimal use of CDS in meeting user requirements.

- **Isabelle Scholl, UPS-IAS**

Isabelle Scholl is the Systems Manager of MEDOC (Multi-Experiment Operati ons and Data Centre for SOHO at IAS). She has been involved in this project since the beginning, initially as the system engineer. She designed, specified and implemented the overall system architecture including network communications, database administration and user computing facilities. She designed the official SOHO archive access system and managed its development. She is now working to extend this software in order to send catalog queries on distributed archive sites. She is the scientific/technical coordinator for all the MEDOC activities and with CNES and NASA interfaces, operations, dedicated links and the SOHO Archive. She served as an expert member in several project reviews for other French Scientific Archive Centers. Previously, she held various positions as Associate Professor in Computer Sciences at Strasbourg University and Software Engineer for a private company for 10 years.

- **Dr. Jean Abouderham, Paris-Meudon**

Dr. Aboudarham gained his PhD at Universite Pierre et Marie Curie, Paris in 1986 and has been with at Paris-Meudon Observatory (France) since 1989. He worked at Institut d'Astrophysique Spatiale (Orsay, France) from 1992 to 1998 for SOHO/SUMER instrument (associated scientist) and for MEDOC (Multi Experiment Data and Operation Center) He is the scientist responsible of BASS2000 Meudon, and has been a Member of Paris Observatory Scientific Council since 1999. His main main fields of research are: Radiative transfer in solar flares; Effects of particle beams during stellar flares; Determination of EUV Solar flux, using SOHO observations and effects on Earth ionosphere

- **Prof. Ester Antonucci, OATO**

Ester Antonucci has been Professore Incaricato of Physics from 1976 to 1983 and Professore Associato of Experimental Physics from 1983 to 1995 at the University of Torino. In 1995 she became Senior Astronomer at the Observatory of Torino. She visited for extended periods Stanford University (US), Rutherford Laboratory (UK) and Goddard Space Flight Center (NASA-US). Her main field of research is solar physics and solar space physics. She has been Deputy Principal Investigator of the Soft X-ray Polychromator Experiment flown on SMM (NASA) in 1980 and she is at present co-Investigator of the Ultraviolet Coronagraph Spectrometer (UVCS) of the SOHO mission (ESA-NASA). She is part of the

editorial boards of the European scientific journals Solar Physics and Annales Geophysicae, vice-chairman of the Commission on Astrophysics of the Committee on Space Research (COSPAR) and member of the International Academy of Aeronautics. She is author of more than 140 publications and editor of three volumes of Advances in Space Research.

- **Dr. Valentina V. Zharkova, UNIBRAD**

Dr. Zharkova obtained an MSc first class with distinction in Applied Maths, Kiev University (1975), PhD in Astrophysics (1983), the Main Astronomical Observatory of the Ukrainian Academy of Sciences, Advanced Certificate in Computing Sciences, Kiev University (1989). Worked at Kiev University, Physics and Applied Maths Department as a Junior Research Scientist (78-82); Senior Research Scientist (82-84); Lecturer (85-91); Senior Lecturer (91-94). Then she worked at Glasgow University, Physics and Astronomy Department as Senior Research Fellow (92-93), then Research Fellow (94-95). Currently she is a Lecturer at Cybernetics Department, Bradford University, UK. She held 1 Rolling and 2 Research grants at Kiev University on the CORONAS space project, also she was a PI of 2 SOHO Guest Investigator programs on investigation of solar flares with MDI data. She supervised 7 MSc, 2 PhD in Astrophysics and more then 100 final year projects in Computing, all successfully defended.

- **Dr S S Ipson, UNIBRAD**

Dr S S Ipson graduated in 1969 with a BSc in Applied Physics following periods spent the Rutherford Appleton Laboratory, BP Research laboratory at Sunbury and the AEA at Winfrith. After carrying out theoretical and experimental work at A.E.R.E. Harwell, and the Universities of Bradford, Oxford and Heidelberg he was awarded a PhD in theoretical nuclear physics in 1975 from the University of Bradford. He joined the Physics department at Bradford University in 1972 teaching a variety of Science and Engineering based subjects at all levels from first year to postgraduate. In 1986, he transferred to the department of EE, became Tutor for the MSc Course in Real-Time Electronic Systems and started his own company. From 1986 his main research interest have related to imaging and he has supervised seven PhD students, held several research contracts and co-authored numerous publications. Transferred to CIVS in 1999.

Appendix B - Contract Preparation forms