Belgian Science Policy Office
belspo

BRAIN-be

BELGIAN RESEARCH ACTION THROUGH INTERDISCIPLINARY NETWORKS

Proposal's title in English (maximum 12 words):
Valorization of 70 years Solar Observations from the Royal Observatory of Belgium

Proposal's title in Dutch:
Valorisatie van 70 jaar zonnewaarnemingen van de Koninklijke Sterrenwacht van België

Proposal's title in French:
Valorisation de 70 ans d'observations solaires de l'Observatoire Royal de Belgique

Proposal's acronym (max 15 characters):
VAL-U-SUN

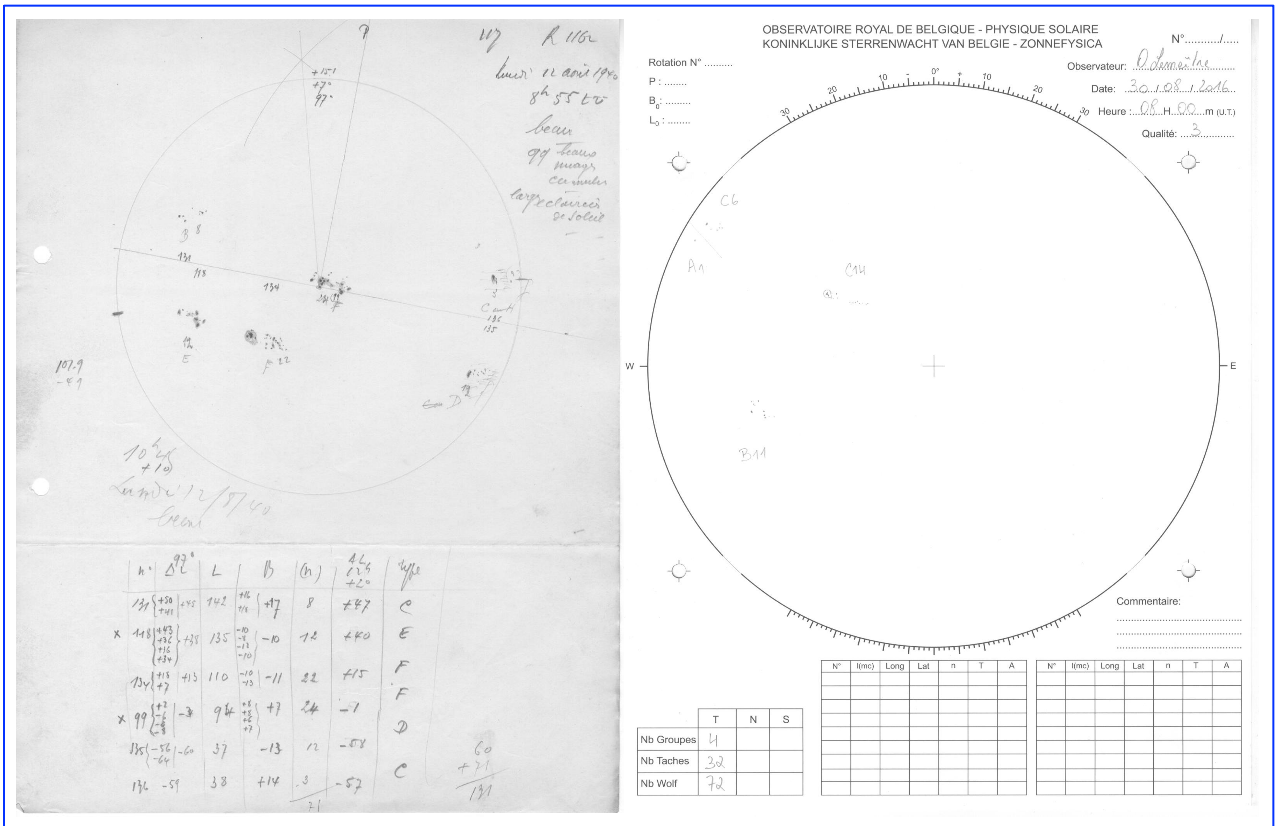

Figure 1 Two samples (left 1940, right 2016) from the Sunspot Drawings Collection obtained at the USET (Uccle Solar Equatorial Table) infrastructure at the Royal Observatory of Belgium. This unexplored heritage collection spans more than 70 years of observations with 23021 handmade drawings showing 87193 sunspot groups (numbers of 2016-08-30)

Research area (tick where appropriate):

2.1 Thematic axis 1&2: "Ecosystems, biodiversity, evolution" and "Geosystems, universe and climate"	
2.1.1 Research at the Princess Elisabeth Base in Antarctica	☐
2.1.2 Universe	☐
2.1.3 Biogeochemical cycles	☐
2.1.4 Biodiversity and Ecosystem functioning	☐
2.1.5 Scientific support for developing climate services	☐
2.1.6 Dynamic Earth	☐
2.2 Thematic axis 3&6: "Cultural, historical and scientific heritage" and "Management of collections"	
2.2.1 Crowdsourcing for federal heritage	☐
2.2.2 Management of digital data/collections: interface and interoperability	☐
2.2.3 Conservation and valorization of 3D federal digital heritage	☐
2.2.4 Management and valorization of the federal audio, photographic and audiovisual heritage	☐
2.2.5 Geolocation for the valorization of the historical, cultural and scientific heritage	☐
2.2.6 Health	☐
2.2.7 Unexplored heritage	**x**
2.2.8 Conflict and Memories in Belgium	☐
2.2.9 The federal heritage of the Southern Netherlands (15th-18th century)	
2.3 Thematic axis 4&5: "Federal public strategies" and "Major societal challenges"	
2.3.1. Violence and discrimination	☐
2.3.2. Migration	☐
2.3.3. Security	☐

The proposal "VAL-U-SUN" is submitted in response to the Call 2.2.7 "Unexplored heritage".

We want to highlight however that the proposal also partially overlaps with the Call 2.1.2 "Universe" ("The evolution of solar activity over all time-scales as a contribution to space climate") and that the results of the project can also contribute to Call 2.1.5 "Scientific Support for the Developing Climate Services" (by providing the service of a long-term data source of solar activity).

TABLE OF CONTENTS

PROPOSAL'S SUMMARY SHEET

PROPOSAL'S TITLE AND ACRONYM:

Valorization of 70 years Solar Observations from the Royal Observatory of Belgium: VAL-U-SUN

NETWORK, BUDGET AND DURATION:

	First and last name	Institution	Budget (€)
Coordinator (partner 1):	Laure Lefèvre	Royal Observatory of Belgium	258400
Partner 2:	Rainer von Sachs	Université Catholique de Louvain	191840
International Partner 1:	Thierry Dudok de Wit	Université d'Orléans	0
		TOTAL:	**450240**

Duration: **4 years**

PROPOSAL SUMMARY (MAX 1-2PG):

RESEARCH AREA: **ASTROPHYSICS, SOLAR PHYSICS**

SUMMARY

Sunspots are dark spots appearing in groups on the solar surface and are manifestations of solar magnetism. Sunspots have been counted since the invention of the telescope in the early 17th century. In 1844 (Schwabe) it was discovered that time series of the sunspot count show a periodic pattern of approximately 11 years. This became known as the solar cycle, which deeply affects the entire solar system. The solar magnetic field embedded in sunspots is the driving force behind the solar variability that on a day-to-day basis influences the space environment of the Earth (“space weather”). In the mid 19th century, Dr. Wolf of the Zürich Observatory created a solar activity proxy by summing up the total number of sunspots with 10 times the total number of sunspot groups. With its 400 years, this “Sunspot Number” is the longest standing observational series of solar activity and has proven to be remarkably useful, not just for solar physics but in a large variety of physical sciences. Cross-correlation with other indices allows reconstructing the solar driving of Earth's Climate over the past few centuries. The sunspot number and the solar cycle are also a “Rosetta stone” for understanding solar-like variable stars. The digital library “The NASA Astrophysics Data System” collects more than 70 refereed papers per year containing the exact phrase “sunspot number” in the abstract or in the title.

Over the centuries, the Royal Observatory of Belgium (ROB) has played an essential role in the study of the Sunspot Number. Sunspot observations were regularly recorded since the early days of ROB (mid 19th century). Around 1908, Mgr. Eugène Spée (1843-1924), a Jesuit priest, started photographic observations of the photosphere at the ROB solar dome using a Grubb heliograph, which is still in use for the ROB visual sunspot drawings (see Figure 1). In 1939 and during World War II, ROB started up a solar observing program in collaboration with the successors of Dr. Wolf at the Zürich Observatory. During 40 years, the counts derived from the ROB Sunspot Drawings were communicated to the World Data Center for the Sunspot Number at the Zürich Observatory. The original solar drawings, obtained at the ROB Uccle Solar Equatorial Table (USET) instrument, have been preserved and scanned and form a homogeneous series of uninterrupted drawings since March 1940. It is one of the longest stable observation collections of the Sun still active. In 1981, the Zürich Observatory ceased the World Data Center activity and in agreement with the international astronomical organization “COSPAR”, the World Data Center was transferred to ROB. As such, the ROB World Data Center “SILSO”[1] is responsible for collecting observations and counts of sunspots from an international and worldwide network of observing stations and for producing from these data the International Sunspot Number.

Figure 2 The building and dome of the Uccle Solar Equatorial Table (USET).

Today, USET is one of the last remaining operational telescope facilities at the Brussels/Uccle site of the ROB and is **a key-Belgian research infrastructure** dedicated to solar physics. The collection of USET sunspot drawings and the SILSO collection of international observations, are unique datasets of scientific heritage. Despite their undisputed and long-standing value, these Belgian **heritage collections of astronomical observations** have only been used to extract in-house the International Sunspot Number by a procedure defined back in the Zürich era. The original

1 Sunspot Index and Long-term Solar Observations World Data Center: SILSO-WDC, http://sidc.be/silso/

observations have never been made available for external research, nor have they been explored with modern scientific methods and remain therefore below their maximum scientific potential.

The first heritage collection, the ROB/USET sunspot drawings, consists of more than 23000 sunspot drawings acquired since March 1940. These drawings were originally made on paper but have been scanned and their scientific content has been largely digitized. This endeavor was paused due to funding limitations. As of today, the digitized drawings lack contextualization and metadata, for example about the observation conditions. Moreover, modern quality control is missing. As a consequence, the USET drawing collection could not be made accessible to the scientific community yet nor used in any scientific publication.

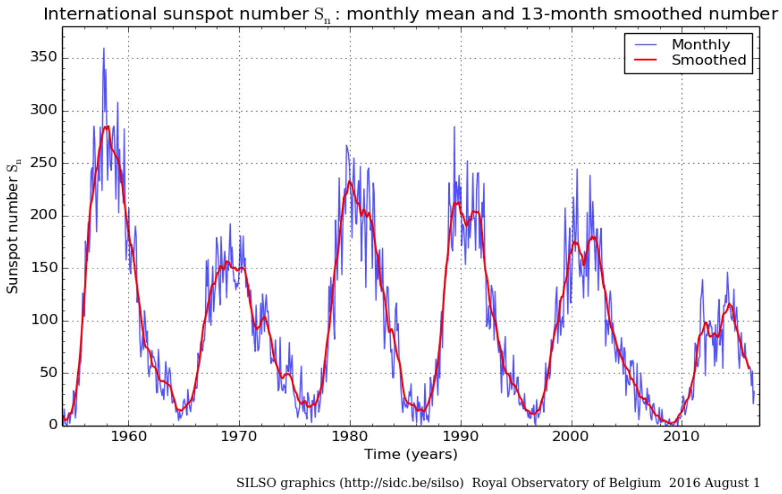

Figure 3 The International Sunspot Number derived from the SILSO database.

The second heritage collection is the SILSO database that contains the numbers of sunspots and sunspot groups since 1981. These numbers are provided by individual stations of the worldwide SILSO network and are used to compute the International Sunspot Number (Figure 3). Despite its success, the statistical stability of the International Sunspot Number needs constant attention. The 'averaging' process that combines the individual observations into one official International Sunspot Number was defined in the 1850s. The process depends on the choice of a reference to which an observation is compared. To this day, the reference is provided by a single station, called the 'pilot station', and as result the International Sunspot Number is sensitive to any problem that could arise with that station. The performance of a particular station in reference with the pilot station is checked only once a year. A quality control of the observations of a station is done on an irregular basis. Therefore, the results of both these controls are not shared with the observers or others. Even worse, the full content of the database itself is available neither to the scientific community, nor to interested people.

This project will *not* use the ROB sunspot data collections for studies of long term solar activity monitoring nor will it produce an alternative or improved sunspot number. The central focus of this project is to open up the ROB Sunspot Data Collections to the international research communities and to contemporary scientific methods by

(1) contextualizing of the collections with standardized documentation and metadata,
(2) updating them with new quality control mechanisms and
(3) making them accessible through an Open Access portal.

A multidisciplinary consortium has been set-up to achieve these goals, in which solar physicists at the Royal Observatory of Belgium collaborate with statistical experts at the Université de Louvain and with an international partner from the Université d'Orléans (France). The present proposal is particularly cost-effective as only the young professionals in this consortium are charged to the project while the senior researchers at all 3 institutes work in co-funding.

KEYWORDS (MAX 5):

DIGITAL HERITAGE, ASTRONOMICAL DATA BASES, OPEN ACCESS
SOLAR ACTIVITY, SUNSPOTS

SECTION I: NETWORK COMPOSITION

NETWORK COORDINATOR (PARTNER 1):

First and last name:
Laure LEFÈVRE

Gender: Male ☐ Female ☒ Language: French ☒ Dutch ☐

Institution and unit
Royal Observatory of Belgium
Solar Influences Data analysis Center

Tel: **+32 2 790 39 23** Email: **laure.Lefèvre@oma.be**

Website: **http://sidc.be**, **http://sidc.be/silso**

First and last name of the Director of the research Institute
Ronald Van der Linden

Type of institution:
☐ University/ University College ☒ Public scientific institution ☐ Non-profit research centre

Address: **Ringlaan -3- Avenue Circulaire, B-1180 Brussels (Belgium)**

Date: 08/09/2016

Signature of Director-general

PARTNER 2 (DUPLICATE IF NECESSARY, WITH NECESSARY PARTNER NUMBER CHANGES)

First and last name:
Rainer Von Sachs

Gender: Male ☒ Female ☐ Language: French ☒ Dutch ☐

Institution and unit :
Université Catholique de Louvain
Institut de statistique, biostatistique et sciences actuarielles (ISBA)

Tel: **+32 10 478 806** Email: **rvs@uclouvain.be**

Website:
http://www.uclouvain.be/rainer.vonsachs

First and last name of the Rector of the university:
Vincent Blondel

Type of institution:
☒ University/ University College ☐ Public scientific institution ☐ Non-profit research centre

Address: **Voie du Roman Pays 20, boite L1.04.01, B-1348 Louvain-la-Neuve (Belgium)**

Date: 5 September 2016

Signature of the partner

INTERNATIONAL PARTNER 1 (IF APPLICABLE, DUPLICATE IF NECESSARY, WITH NECESSARY PARTNER NUMBER CHANGES)

First and Last name:
Thierry DUDOK DE WIT

Gender: Male [X] Female [ ]

Institution and unit:
Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E), Université d'Orléans et CNRS

Tel: **+33 238 25 52 77** Email: **ddwit@cnrs-orleans.fr**

Website: **http://lpc2e.cnrs-orleans.fr/~ddwit/**

First and last name of the Director of the research Institute or Rector of the university:
Michel TAGGER

Type of institution: [X] University [ ] Public scientific institution

Address: **LPC2E, 3A avenue de la Recherche Scientifique, 45071 Orléans, France**

Please specify the type, content and source(s) of the co-financing:
Local funding

We the undersigned, hereby declare our interest in participating in the proposed project as international member of the network, and to co-fund at least 50 % of our participation.

Date: **24 Aug 2016**

Signature of the partner *Signature of the contractor (for agreement)*

1. COMPLIANCE OF THE PROPOSAL WITH THE SCOPE OF THE CALL

The present proposal is 100% compliant with the Scope of the Call “Unexplored Heritage”, as confirmed by the Table below.

“Unexplored Heritage” Call Requirements (section 2.2.7 of the Brain.be call document)	**VAL-U-SUN compliance confirmation**
The scientific institutions and the federal departments store collections that have never been the subject of scientific study according to contemporary scientific standards.	Section 2.3 below describes the ROB Sunspot Data Collections and why they have never been the subject of scientific study according to contemporary scientific standards.
They are therefore often difficult to access either physically or digitally and suffer from a lack of metadata, which significantly limits their relevance for current research.	Because of limitations of the available metadata and the lack of systematic quality monitoring, the ROB Sunspot Data Collections have been kept private up till now. At the end of this project, the ROB Sunspot Data Collections will be suitable and accessible for Open Access research.
The research project will concern the updating, contextualization and accessibility for research of a specific collection or a set of collections.	Section 2.4 below lists specific work packages for the contextualization, the updating and the accessibility of the ROB Sunspot Data Collections.
The scientific importance - on the national and/or international level - of stimulating the scientific valorization and accessibility of the collection or collections […] must be explicitly and sufficiently demonstrated.	The collections of historical observations of the World Data Center of the Sunspot index are indisputably of the highest scientific importance (e.g. http://www.nature.com/news/spotty-sunspot-record-gets-a-makeover-1.18145)
The research projects will be based on an appropriate multidisciplinary approach which enables a specific scientific need to be fulfilled and/or the heritage in question to be made more accessible for a wider audience.	The specific scientific need that VAL-U-SUN will fulfill is to guarantee the long term stability and quality, and accessibility to external users of the ROB Sunspot Data Collection. This long term work requires a multidisciplinary team of solar physicists at one hand and statisticians at the other hand.
This type of approach will also enhance the potential of the heritage in question to function as the primary object of research or as a reference collection for a contemporary scientific issue.	VAL-U-SUN will help contextualize two ROB Sunspot Data Collections (see section 2.3) and make them widely available for research on solar activity on the mid and long term.

2. SUBJECT OF THE PROPOSAL

2.1 STATE OF THE ART AND OBJECTIVES

The techniques behind sunspot drawings and counting have been static for many decades. Despite the advance of CCD-equipped telescopes, sunspot drawings have been deliberately kept unchanged to continue historical time-series in similar circumstances as centuries ago. In this sense, the ROB Sunspot Data Collections truly are a historical heritage.

Nevertheless, increasing concerns on the stability of the International Sunspot Number have sparked a strong increase in the research activity on the process from raw sunspot observations to an internationally agreed, stable solar activity index. A series of international workshops (Cliver et al. 2015) have been organized to correct historical discrepancies in the calibration of the International Sunspot Number. Such discrepancies could be identified by studying changes in the ratio of the sunspot number with similar indices, such as the Group Sunspot Number (Hoyt et al. 1994, Hoyt & Schatten 1998a,b). By relating these discrepancies to known transitions in the circumstances in which the sunspot number was produced, it has been possible to recalibrate specific incoherencies in the sunspot number (Clette et al 2015, 2016).

To avoid a multiplicative accumulation of historical errors, a “back-bone” method (Svalgaard 2013b) was constructed based on overlapping, “high quality” sunspot observers with particular long time spans. Obviously, in this process it became clear that proper contextualization and documentation of the data are crucial.

The international effort to recalibrate the Sunspot Number is not finished yet. In the last few years, new questions came up. Observations point out that the magnetic field strength in sunspots has had a secular decrease from 1998-2012 (Livingston et al. 2012, and references therein). In addition, after the solar minimum of 1996, the strong correlation between the International Sunspot Number and the solar radio flux at 10.7 cm (e.g. Kundu 1965) appeared to break down (Svalgaard & Hudson 2010). Also, the long-term ratio of the number of spots to groups has decreased from its nominal value of 10 (Clette et al. 2014). These issues also lead to new questions on the confidence level of the International Sunspot Number and the “quality” of each contributing station in the SILSO network. The SILSO WDC considers it as a must that all the research it produces should be transparent and reproducible by independent researchers. The whole science community should be able to participate in this upcoming research, not only the SILSO World Data Center.

Therefore, the over-arching goal of this project is thus to make the ROB Sunspot Data Collections (see section 2.3) accessible for external researchers. In particular we want to

a) Make the USET sunspot drawings available as a “high quality” observation station from 1940 till present, with proper contextualization and documentation to make it a primary reference (ref. back-bone method).
b) Open up the SILSO database of individual sunspot count observations to investigations of external researchers and provide state-of-the-art quality monitoring of the stations.

To achieve this goal, we will use the following objectives as guiding principles:

1. The shared ROB Sunspot Data Collections must be as complete as possible, i.e. they must retain a maximum of the scientific content that was available in the original data.
2. The quality of the long-term data sets must be monitored continuously and the results of the quality checks must be accessible to both the data providers and the user. In case of quality loss, the affected observation station must be alerted.
3. The ROB Sunspot Data Collections, together with the metadata, full context and documentation must be readily (online) available following an Open Access philosophy.
4. The science communities, for which the ROB Sunspot Data Collections are relevant, must be made aware of the updated ROB Sunspot Data Collections.

2.2 METHODS

A number of methods will be needed to achieve the goals and objectives (section 2.1) for both ROB Sunspot Data Collections (2.3). Whereas the Work Plan in section 2.4 gives a complete overview of all the tasks and methods employed, we focus here on the details of the more technical methods.

2.2.1 SUNSPOT DRAWING DIGITIZATION METHOD

The USET station uses a software application for sunspot drawing digitization called DIGISUN. The application allows an operator to extract from a drawing the number of groups on the Sun, their positions, the number of spots inside each group, their Zürich and McIntosh classification, and their dipole tilt. A first processing of the USET sunspot drawings with DIGISUN was done during the FP7-SOTERIA project between Nov 2008 and Oct 2011 but was never completed: 10% of the drawings have not been digitized.

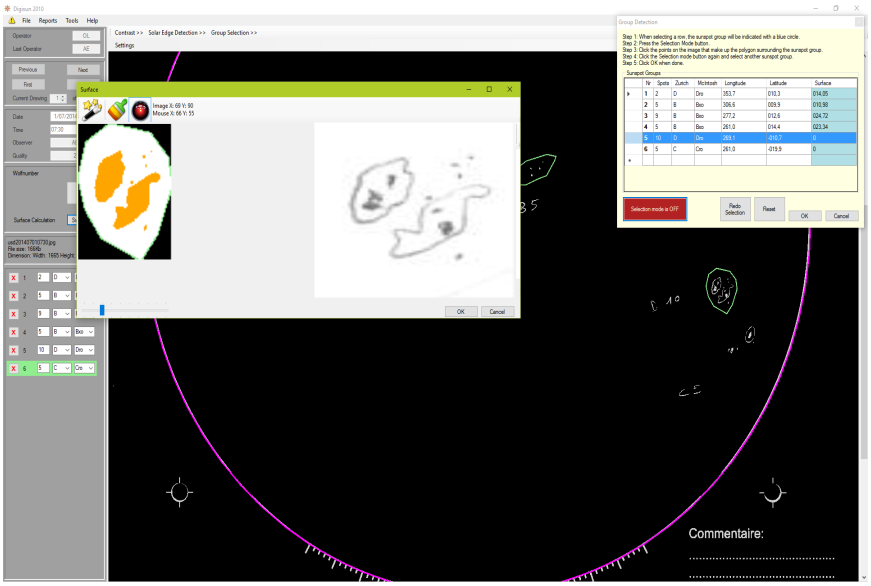

Figure 4 The DIGISUN application to digitize solar drawings with the new module for area determination.

Meanwhile, an improved DIGISUN application has been prototyped that includes an area-measurement module as shown in Figure 4. Using a semi-automatic technique, including masking and region growing within an operator-defined region, pixels are counted that correspond to the sunspot area. With the location of each pixel known, the spherical projection effect is corrected to provide a more consistent time series of the area during the solar rotation. Within the present project, the DIGISUN prototype remains to be commissioned after which it can be massively applied to the USET sunspot. By using this same technique consistently over the full range of the USET sunspot drawings collections since 1940, we will guarantee the consistency of our time series across the past and current solar cycles for long-term analyses.

The resulting database of sunspot group characterization will undergo critical quality assessment. Techniques developed for the analysis in Lefèvre et al. (2011, 2014) will be used to assess the accuracy of the extracted parameters.

2.2.2 METHODS RELATED TO QUALITY MEASURES AND REFERENCES STATIONS FOR THE SILSO NETWORK

The SILSO database contains sunspot counts (see section 2.3.2) from 285 observing stations that contributed during some time between 1940 and the present day, but at any given time, only a subset of stations contribute data. Figure 5 presents the raw sunspot numbers (monthly averaged) given by 11 different stations. The (upper) red curve shows the data of the Locarno station, which is used nowadays as the pilot or reference station. The (second) blue curve corresponds to the sunspot numbers from the USET station in Uccle. This figure highlights one of the main challenges we are facing, which is the high inhomogeneity of the database. A significant effort will be dedicated in this project to research the quality of the contributions of each station and the implementation of quality measures. By quality, we mean long-term and overall stability of the observations within the network.

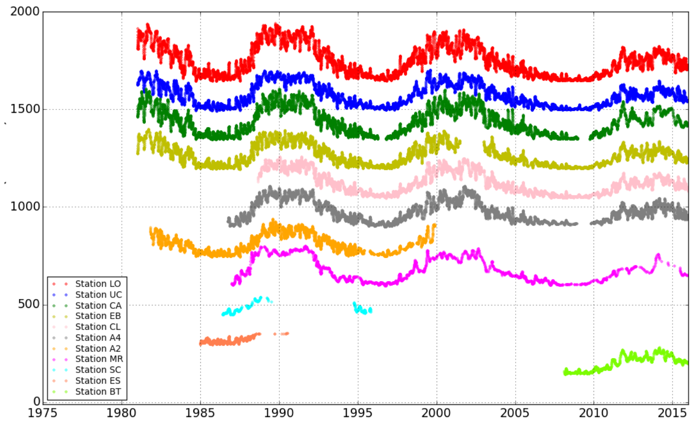

Figure 5 The sunspot number (vertical axis) as derived from the data available for single stations, as a function of time (horizontal axis).

Most sunspot records lack confidence estimates, which are essential for performing statistical tests and for doing quality assessment. To make up this shortfall, we recently developed a series of methods for estimating the precision, out of which confidence estimates can be derived (Dudok de Wit et al. 2016). For example, one method consists in fitting a model of an inhomogeneous Poisson point process to the data per station and subsequently calculating a standard deviation for each station at every day. These methods will be adapted to the databases and applied in a systematic way to all records.

We will perform an extensive check for the stability and quality of past and present observers of the network, with the aim to define a robust subset of the network's observation stations to be used as a reliable reference multi-station pool that will replace the current single pilot station. For this, modern methods for statistical quality control must be used and adapted.

The main criteria for including a station in the multi-station reference, is its ***long-term stability***. As a second priority the variance (or rather the standard deviation) from the network should be approximately stable over time and if possible low (stability being more important than level). A basic approach of this reference was proposed with the average of a subsample of stations in Clette et al. (2016). In this project, it will be the first approximation leading to the use of more advanced methods taking the full database into account.

We plan on using a series of complementary nonparametric methods that should allow to detect, on the one hand, sufficiently high level shifts (jumps or break points), and on the other hand, changes from a temporal regime of constant level into a period of smoothly in- or decreasing level (drift). Segmentation-based methods (Fryzlewicz 2007, 2014, 2016; Davies and Kovac, 2001) can be used to detect break points, while longer-term drifts are expected to be detectable by "integrated difference" or CUSUM-statistics methods (Page 1954). It is clear that the latter ones will need longer periods of adaptation of the detection method before being able to give alert, and it will be part of the research program to investigate which method will be suited to perform best with respect to these two objectives.

A complementary method to determine the sample of reference stations would be by separating the solar signal from the purely human/observational variations by using methods such as Principal Component Analysis (PCA) or SVD. Alternatively ICA can be used, a method which aims at splitting the data series into its “most independent” components. ICA is a member of a group of algorithms developed in signal processing and commonly referred to as ``blind source separation'' (Kruskal 1969, Coman 1994). Alternatively, a mixed effects approach (Chau and von Sachs 2016, for curves) might be appropriate, too, to model this separation into a common (solar) source and a station-specific variation.

2.2.3 METHODS CONCERNING ACCESSIBILITY

The ultimate goal is to make the ROB Sunspot Data Collections fully accessible for modern scientific exploitation. All the original data, as well as all results of the project will be accessible online following a full Open Access/Open Data policy. One-way accessibility in itself is however not sufficient, we aim at a more interactive engagement with our target user communities, hence registration will be asked before accessing the data distribution section of the website. In particular we foresee to actively communicate the results of the project through

- the International Astronomical Union's (IAU) Inter-division B and E working group on Coordination of Synoptic Observations of the Sun (see section 3.3. Follow-up Committee)
- the series of International Workshops on the Sunspot Number (http://ssnworkshop.wikia.com/wiki/Home, Cliver et al 2015)
- the annual European Space Weather Week (see http://www.stce.be/esww13/ for this year's edition)
- the ROB communication channels (e.g. http://www.stce.be, http://www.astro.oma.be/en/)

As deliverables of the project, we will also issue direct email-messages to specific target groups

- "*Quality issue notifications*" will be sent by email to the SILSO stations of which the quality has noticeable changed. ROB will monitor the quality once per month and keep a digital log book (ticket system) of all communication with the data providers.
- *"Sunspot group configuration bulletins"* will be sent by email to registered space weather networks on a daily basis on the distribution and characteristics of the sunspots presently on the solar disc as an input for e.g. solar flare forecasts. For example, the International Space Environment Service (ISES, http://www.spaceweather.org) has a pre-defined for this (called "USSPS").

Finally, we foresee an online consultation platform. This platform (essentially a part of the website with access on invitation) will show a number of sunspot drawings and the participants are asked to count the number of spots and groups. The idea is to assess how people with different levels of expertise interpret a sunspot drawing in terms of the sunspot number. As the sunspot number is the sum of the number of sunspots and 10 times the number of sunspot groups, the identification of sunspot groups is critical. We will invite a number of participants from three groups with different expertise level: expert solar physicists, SILSO network observers, and the general public. We will study how the interpretation varies from one participant group to the other, e.g. is there a difference between the participants groups and if so, are the SILSO observers (many of whom are amateurs without training) interpreting drawings like expert solar physicists or more like the general public? This investigation is motivated by the realization that early sunspot observers did not have any prior expertise and thus probably resembled the general public of today. Also, as a SILSO observer gains more expertise, (s)he might shift from one category to the other and thereby increase the quality of his/her interpretations.

2.3 DATA

We focus on the two ROB Sunspot Data Collections:
(1) the USET sunspot drawings and (2) the SILSO network sunspot counts

2.3.1 USET SUNSPOT DRAWINGS

Visual sunspot drawings (see Figure 1) have been collected at ROB since March 1940 and are continuously extended every sunny day at the Uccle Solar Equatorial Table (http://sidc.be/uset), one of the last active telescope platforms at the historical ROB Uccle site. As of 2016-Aug-30, 23021 drawings have been collected showing in total 87193 sunspot groups.

In the last years, newly acquired drawings have been (and are) routinely scanned and digitized through a PC application called 'DIGISUN' to extract the following information:

(1) General information about the drawings: date and time, seeing quality, observer, Carrington rotation, Julian date, group count, spot count, sunspot number, P, B_0 and L_0 angles and operator identity.
(2) Calibration of the digital scans of the drawings: pixel coordinates of the Sun's north and center, scan angle and physical radius of the solar disc.
(3) Sunspot groups parameters: latitude, longitude, McIntosh type, Zürich type, number of spots inside the group, and latitude and longitude of the two dipole elements.

Over the full data set, all drawings have been scanned but only 90% has undergone the above digitization process thereby storing the extracted parameters in a (currently private) database. The current DIGISUN process does not extract all scientific valuable information present in the drawings. In particular, the total area of all sunspots on the disk is a missing feature that will be addressed in this project.

Besides the above parameters, more contextualization information is in principle retrievable from the ROB archives (historical context, changing set-up/instrumentation, team composition) and from other log-books but is not openly accessibly to external users. Systematic quality tracking of the different USET personnel and the extracted parameters is currently missing, which is why the digitization database could not be released to open access research yet.

2.3.2 SILSO NETWORK SUNSPOT COUNTS

As part of its task as SILSO World Data Center for the Sunspot Number, ROB collects sunspot counts from a worldwide network of contributing observers. Since 1981, SILSO has accumulated 586247 sunspot counts from over 285 solar observing stations (amateurs and professionals, ROB/USET being one of them). Through an online web-interface connected to the SILSO database, the contributing observing stations specify for each observation:

(1) the observation time and seeing quality,
(2) the number of spots and the number of sunspot groups in total for the entire solar disc, and
(3) optionally: separated these numbers for each solar hemisphere and the central zone of the solar disc.

Up till now, the resulting SILSO database has been uniquely used to produce the so-called International Sunspot Number through a reduction technique, based on a simple comparison with a trusted reference station, that has not evolved significantly over the last 150 years. Concerns about the stability of the resulting International Sunspot Number on long timescales are at the heart of the present project. At the present time, the SILSO database does not contain any contextualization metadata on the observing station such as the geographical location and instrumentation, the coverage over time or the quality/stability of the observing station and possibly changes therein. Some of this information is available off-line in ROB archives or can be retrieved in dialogue with the observing stations.

2.4 WORKPLAN AND DETAILED DESCRIPTION OF THE TASKS

The work breakdown structure of this project is aligned with the main aspects of the call, namely "updating, contextualization and accessibility for research of [...] a set of collections".
We thus foresee the following work packages:

- WP1 "Management"
- WP2 "Contextualization of the ROB Sunspot Data Collections"
- WP3 "Updating of the ROB Sunspot Data Collections"
- WP4 "Accessibility of the ROB Sunspot Data Collections".

These work packages consist of a number of tasks, each ending with at least 1 deliverable. The tasks typically start and end at the project milestones when formal consortium meetings are held:

- Milestone Meeting 1 (month 6): End of start-up phase
- Milestone Meeting 2 (month 24): Mid-term Review of the project
- Milestone Meeting 3 (month 42): Start of the last half-year finalisation.

In the following text we present an overview of each of the work packages and the tasks therein. An overview table ofl the 18 deliverables is presented at the end of this section.

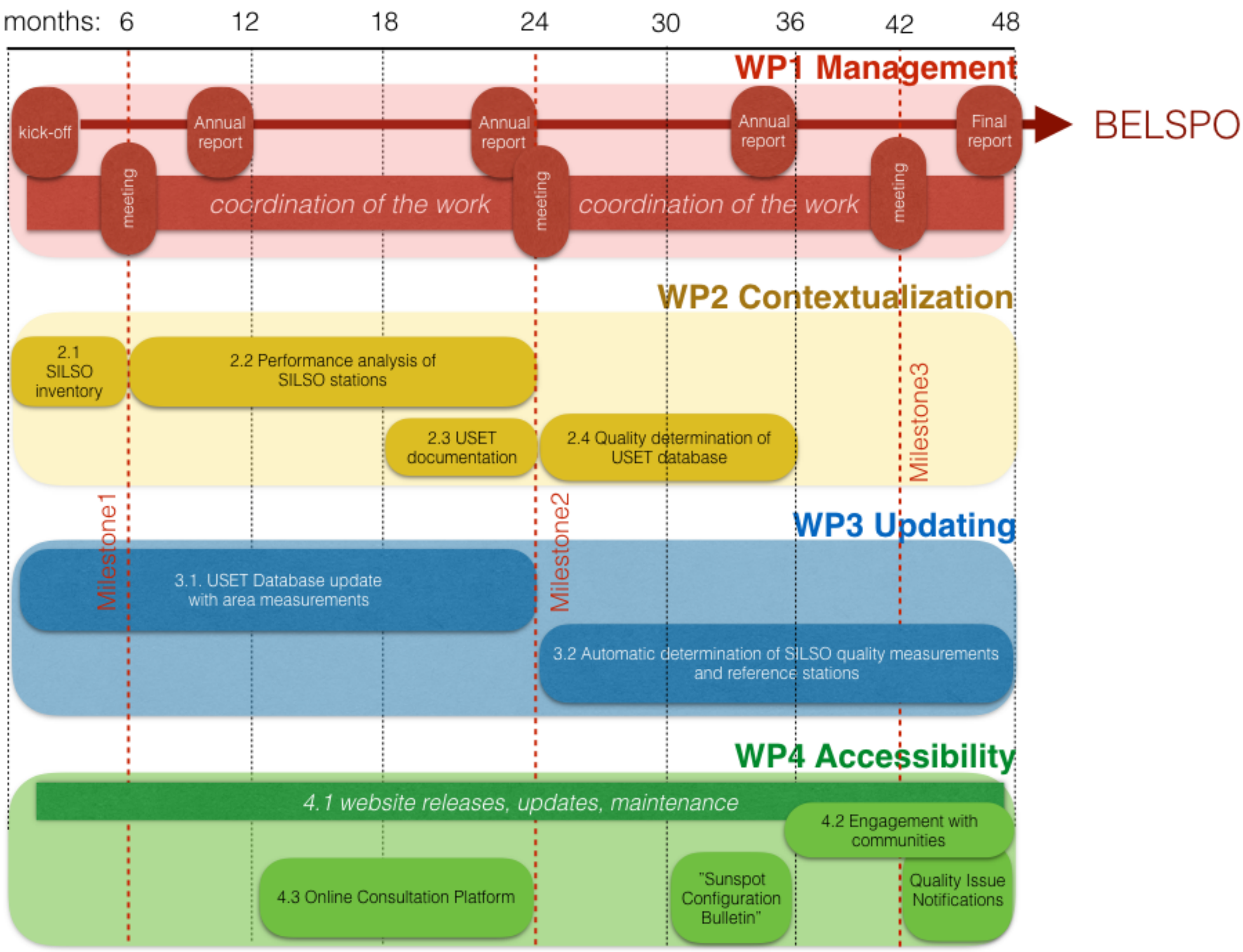

Figure 6 Work breakdown structure and overall planning of the VAL-U-SUN project

WP1: Management

(WP manager: ROB/Lefèvre)

The project coordinator Dr. Laure Lefèvre will be responsible for the overall management of the project and take up all the roles foreseen in the BRAIN-BE-Call2016 document. For clarity we split these roles here in two tasks:

Task 1.1 Interfacing with BELSPO

Scheduling: Month 1 – Month 48, duration: 48 months
Participants: ROB/ Lefèvre (coordinator)

Work items:
The project coordinator will be the point of contact between BELSPO and the consortium:

- To coordinate the production of the interim and final project reports intended for BELSPO;
- To inform BELSPO of any problems that might interfere with the correct implementation of the project;
- To coordinate the synthesis and translation of the research results, with a view to applications and support for decision-making;
- To coordinate the publication and dissemination of the research results; meetings related to the project's progress between the network and BELSPO.

Task 1.1 Deliverables:
We foresee annual reports and a final report as the formal deliverables of this task.

Task 1.2 Coordination of the work

Scheduling: Month 1 – Month 48, duration 48 months
Participants: ROB/ Lefèvre (coordinator)

Work items:
Assisted by the respective WP-managers, the project coordinator will:

- coordinate all activities to be carried out in the framework of the project;
- coordinate the internal meetings between the network members;
- coordinate the meetings with the Follow-up Committee and write the reports of these meetings.

Task 1.2 Deliverables:
We foresee 3 milestones throughout the project (Month 6, 24 and 42) when meetings will be organized with the consortium members and follow-up committee to streamline the coordination between finishing and starting work packages. Minutes of the meetings will be recorded and are deliverables of the project.

WP2: Contextualization of the ROB Sunspot Data Collections
(WP manager: UCL/von Sachs)

For datasets covering decades, it is essential to record and monitor the circumstances in which the observations were made and to follow up the evolving quality of the observations. Understanding and analysis is greatly facilitated if this context is readily available and accessible to the researcher. In this work package, we work on the contextualization of the ROB Sunspot Data Collections by standardizing and extending the available documentation and metadata by carefully recording the quality of the time series and discrepancies therein.

Task 2.1 SILSO network documentation and metadata inventory

Scheduling: Month 1 – Month 6, Duration: 6 months
Participants: ROB/Clette (coordinator), ROB/Lefèvre, UCL/von Sachs, UCL/student, CNRS/Dudok de Wit

Work items:
The work consists of 4 subtasks:
1. To retrieve the available documentation for each SILSO observing station from internal ROB archives and notebooks
2. To interact with the observing stations on additional information they can share about their observational set-up and procedures, and to ask them permission of sharing their data and documentation through an open access portal
3. Produce single-station confidence levels, following the techniques described in Dudok de Wit et al. (2016). Based on fitting a model of an inhomogeneous Poisson point process to the data per station, a standard deviation is attributed to each station at every day
4. To structure all the above in a coherent, structured data-pack that can be made available to third parties

Task 2.1 Deliverables:
An online report will contain the available documentation of each of the SILSO network stations. The online report will be available on the Open Access website (see Task 4.1).

Task 2.2 Performance analysis of SILSO stations

Scheduling: Month 6 – Month 24, Duration: 18 months
Participants: UCL/von Sachs (coordinator), ROB/Lefèvre, UCL/student, UCL/Ritter, ROB/Clette

Work items:
This corner stone research task consists of the following 2 subtasks:
1. To determine which criteria can be used successfully to assess the quality of the data.
2. To implement methods/techniques to assess the long-term stability of the stations from the database once the stability/quality criteria have been determined

By quality, we mean long-term and overall stability of the observations within the network. That is, on the one hand, not only do we want to clean the data base from apparently outlying values but also detect those observations which – according to the criteria to decide upon - are not compatible with the distribution of the main bulk of the reported data within the network. On the other hand, we need to develop criteria which lead to stability not only of the mean behavior of the data pool over time but also to stability of the employed (not too complex) reporting methods that each observer is supposed to follow without difficulties. Three different criteria to be looked upon first should be: mean stability (i.e. long-term stability of the mean value), variance stability, and variance size where “variance” stands for variability. The size of this “threshold” value of variability is of course important to be determined in a stable way over time.

Part of the research will be centered on determining the specific time-range that defines "long-term", i.e. how often do the samples of the stations have to be re-examined to make them of high quality (i.e. less variable).

A prominent approach to determine the long-term stability is to first have a reference determined by the average network, and then to look at the time series for that initial reference, thus removing the solar signal. This is the method that was proposed in Clette et al. (2016) using the ratio between the stations time series and the average network (k-factors method). The aim is to see whether and along which periods in time these k-factor time series can be considered to be approximately constant over time.
One possibility is to look for break points in time of these series, with the aim to exclude periods where there is instability. Two different scenarios of instability have to be distinguished (and detected): jump like breakpoints clearly indicating a level shift, and changes into a regime with drift (indicated by a sequence of not too high, and hence individually not significant, level changes). Since these k-factor series are quite noisy, we suggest to work on the smoothed k-values.
There are a series of complementary modern nonparametric methods that would allow detecting the aforementioned two scenarios of break points (jumps and drifts) – (see methods described in Section 2.2.2). It will be part of the research program to investigate which method will be suited to perform best with respect to these two objectives, and to subsequently develop the correct calibration of the alert threshold for the presence of a break point.

Task 2.2 Deliverables:
The results of this research task will be made available in an online report on the Open Access website (see Task 4.1) and be published in peer reviewed journals when applicable.

Task 2.3 USET Observatory documentation

Scheduling: Month 18 – Month 24, Duration: 6 months
Participants: ROB/Clette (coordinator), ROB/Lemaitre

Work items:
The USET sunspot drawings form an homogenous, systematic observational collection but obviously the details of the procedures, instrumentation and personnel evolved over the more than 70 years of observations. In this task we will make an inventory of all the historical information available in ROB logbooks, databases and annual reports and extract all context information that can be of eventual use for the end data user.

Task 2.3 Deliverables:
The contextualization report of the USET Observatory will be made available in an online report on the Open Access website (see Task 4.1).

Task 2.4 Quality determination USET database

Scheduling: Month 24 – Month 36, Duration: 12 months
Participants: ROB/Lefèvre (coordinator), ROB/Clette

Work items:
The work consists of the following 3 subtasks:

1. The USET database will be contextualized by critical comparison with other sunspot catalogs overlapping in time such as the sunspot catalog of the Royal Greenwhich Observatory catalog (1874-1976, http://www.ngdc.noaa.gov), the Debrecen Photoheliographic Data (DPD) sunspot catalogue (http://fenyi.solarobs.unideb.hu/DPD/index.html), the USAF catalog, and Kodaikanal and Mount Wilson Sunspot data. Techniques developed for the analysis in Lefèvre et al. (2011,

2014) will be used to assess the accuracy of the extracted parameters.

2. In particular, the quality of the newly created sunspot area parameter (task 3.1) will be assessed through comparison with areas measured at the Catania Astrophysical Observatory. Their method uses a fully manual approach, with which the area is measured using overlaying transparent sheets that give the area of sunspots in mm^2. Because the Catania Astrophysical Observatory does not correct for spherical projection, this sanity check will restrict to sunspots that lie within [0-30º] of the disk center, where the areas are not significantly affected by the projection effect. From Lefèvre et al. (2011, 2014), it appears evident that the areas will not be identical, what is to be assessed is the coherency of the differences.

Task 2.4 Deliverables:
The results of this research task will be made available in an online report on the Open Access website (see Task 4.1) and be published in peer reviewed journals when applicable.

WP3: Updating the ROB Sunspot Data Collections

(WP manager: ROB/Clette)

The ROB Sunspot Observations Collections will be enriched with new information derived from state-of-the-art statistical and software methods, as well as from the results of the previous WP (contextualization). This update will prepare these heritage data collection for modern exploitation through Open Access by the broad science community.

Task 3.1 USET database update with area measurements

Scheduling: Month 1 – Month 24, Duration: 24 months
Participants: ROB/Lemaitre (coordinator), ROB/junior, UCL/student

Work items:
The scope of this task is to reprocess all drawings through the enhanced DIGISUN tool to expand the database with area measurements per sunspot group. In the process, a small fraction of the drawings currently missing from the DIGISUN database will also be added. The enhanced DIGISUN tool is available and following a short commissioning, it will be ready for use.

There are presently just over 23000 drawings to be addressed. We have assumed a DIGISUN processing time of 5 min per drawing. Some drawings will require less than this (e.g. days without sunspots), while for other drawings (e.g. very old ones), more time might be needed. Using the ROB standard working time assumptions (7.6 hours/day, 220 working days/year), this leads to 27 person*months to complete this task (see table in section 2.5). As this a highly repetitive task, we will aim at parallelizing the effort over several people.

Task 3.1 Deliverables:
The updated USET database of digitized drawings will for the first time be made available for external research through the Open Access website (see Task 4.1).

Task 3.2 Automatic determinations of quality and reference stations for the SILSO network

Scheduling: Month 25 – Month 48, 24 months duration
Participants: UCL/von Sachs (coordinator), ROB/Lefèvre, UCL/student, UCL/Ritter, ROB/Clette

Work items:
The work consists of the following 2 subtasks:

1. Choice of the best technique(s) to determine the reference stations for the SILSO network. This task is directly linked to the results of the research from task 2.2 on finding the stability criteria for assessing the quality of a given station. Quite naturally those should also serve for determining the pool of reference stations by including those stations into the pool which are most stable according to the chosen criteria. In addition, one needs to compare these techniques with the alternative approach (as described in Section 2.2) of modeling the sources of instability by separation into either orthogonal (or independent) components or via an additive model (factor model, mixed effects modeling). Having a more complete probabilistic model at hand for what should be considered as “too variable” will improve the quality of the pool of reference stations. Finally, all developed methods should be compared with respect to their capacity to be used in an “online” version (with horizons of one to three months), for which some modifications for their calibration will probably be necessary.
2. Software implementation within the SILSO infrastructure.
   As most of the programs will be developed in R (programming language for statistical computing) and the SILSO infrastructure is mostly in Fortran and Python, an adaptation of the research software to operational software will be necessary.

Task 3.3 Deliverables:
The updated SILSO network observations database will for the first time be made available for external research through the Open Access website (see Task 4.1).

WP4: Accessibility of the ROB Sunspot Data Collections
(WP manager: ROB/Lemaitre)

The ultimate goal is to make the ROB Sunspot Data Collections openly accessible for modern research and hence all work culminates in the present work package.

All project results will be published online under an Open Access policy (Task 4.1). One-way accessibility in itself is however not sufficient, we aim at a more interactive engagement with our target user communities (Task 4.2). Finally, a specific task is foreseen for the creation of an online consultation platform and the analysis of the resulting feedback (Task 4.3).

Task 4.1 Open Access data and documentation website

Scheduling: Month 1 – Month 48, 48 months duration
Participants: ROB/Lemaitre (coordinator)

Work items:
All the results of the project, starting from the enriched data set collections, the associated metadata, and the rebuilt contextualization will be available on the website, including also:
- full historical documentation of USET and SILSO
- project reports and papers
- accessibility of the database and added metadata to the SILSO network
- definition of the open data policy.

Task 4.1 Deliverables:
The SILSO website will be expanded to an open access data and documentation website containing all the project deliverables, readily accessible for registered users.

Task 4.2 Engagement with Scientific Communities

Scheduling: Month 25 – Month 48, 24 months duration
Participants: ROB/Clette (coordinator)

Work items:
The work consists of the following 3 subtasks:
1. Dissemination sessions at relevant community meetings.
2. Creation of "Sunspot Configuration Bulletin" e-mail messages to relevant space weather consumers.
3. Creation of "Quality Issue Notification" e-mail messages to SILSO observers

Task 4.2 Deliverables:
The results of this project will be presented at relevant scientific meeting through dedicated sessions, contributed talks and/or posters.

Task 4.3 Online Consultation Platform

Scheduling: Month 13 – Month 24, 12 months duration
Participants: ROB/Lefèvre (coordinator), ROB/Lemaitre

Work items:
As part of the SILSO website, we will create an online consultation platform were the opinions of different classes of contributors are collected (see section 2.2.3 for more detail). The subtasks are:
- identify critical items that need consultations

- identify participants groups from whom we want feedback
- set-up online platform
- analyse the outcome, write report

Task 4.3 Deliverables:
The results of this online consultation will be made available through a dedicated report.

Nr.	Tasks	Deliverables	Due date (months)
1.1	Interfacing with BELSPO	D1 Annual Report Month 12 D2 Annual Report Month 24 D3 Annual Report Month 36 D4 Final Report	12 24 36 48
1.2	Coordination of the work	D5 Minutes Milestone Meeting 1 D6 Minutes Milestone Meeting 2 D7 Minutes Milestone Meeting 3	6 24 42
2.1	SILSO network documentation and metadata inventory	D8 Online report	6
2.2	Performance analysis of SILSO network station	D9 Research report/Scientific publications	24
2.3	USET Observatory documentation	D10 Research report/Scientific publication	24
2.4	Quality determination USET database	D11 Research report/Scientific publication	36
3.1	USET database update with area measurements	D12 Updated USET database	24
3.2	Automatic determination of quality and reference stations for the SILSO network	D13 Report on developed procedures and software, updated SILSO database	48
4.1	Open access data & documentation website	D14 Open Access Website serving all results of this project	6 (first version)
4.2	Engagement with Scientific Communities	D15 Presentations at relevant scientific meetings. D16 Sunspot Configuration Bulletin D17 Quality Issue Notifications	TBD <48 36 <48
4.3	Online Consultation Platform	D18 Report on results of online consultation	24

Table 1 Overview of Tasks and Deliverables

2.5 TIMETABLE

		Year 1		Year 2		Year 3		Year 4		Funded Man-months	
	Semester	1	2	1	2	1	2	1	2	BELSPO	other sources
Work Package 1: Management											
Task 1.1 Interfacing with BELSPO	**C (ROB)**	half a day every month								1	0
	P2 (UCL)	0	0	0	0	0	0	0	0	0	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**1**	**0**
Task 1.2 Coordination of the work	**C (ROB)**	a day every month								2	0
	P2 (UCL)	0	0	0	0	0	0	0	0	0	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**2**	**0**

		Year 1		Year 2		Year 3		Year 4		Funded Man-months	
	Semester	1	2	1	2	1	2	1	2	BELSPO	other sources
Work Package 2: Contextualization of the ROB Sunspot Data Collections											
Task 2.1 SILSO network documentation and station inventory	**C (ROB)**	2	0	0	0	0	0	0	0	1	1
	P2 (UCL)	4	0	0	0	0	0	0	0	4	0
	IP (UO)	2	0	0	0	0	0	0	0	0	2
									Sub total:	**5**	**3**
Task 2.2 Performance analysis of SILSO stations	**C (ROB)**	0	1	2	2	0	0	0	0	4	1
	P2 (UCL)	0	7	7	6	0	0	0	0	16	4
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**20**	**5**
Task 2.3 USET Observatory documentation	**C (ROB)**	0	0	0	3	0	0	0	0	2	1
	P2 (UCL)	0	0	0	2	0	0	0		2	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**4**	**1**
Task 2.4 Quality determination USET database	**C (ROB)**	0	0	0	0	4	3	0	0	6	1
	P2 (UCL)	0	0	0	0	1	1	0	0	2	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**8**	**1**

		Year 1		Year 2		Year 3		Year 4		Funded Man-months	
	Semester	1	2	1	2	1	2	1	2	BELSPO	other sources
Work Package 3: Updating of the ROB Sunspot Data Collections											
Task 3.1 USET database update with area measurements	**C (ROB)**	6	6	6	7	0	0	0	0	25	0
	P2 (UCL)	2	0	0	0	0	0	0	0	2	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**27**	**0**
Task 3.2 Automatic determinations of quality & reference stations for SILSO	**C (ROB)**	0	0	0	0	1	1	2	1	4	1
	P2 (UCL)	0	0	0	0	6.5	6.5	6.5	6	20.5	5
	IP (UO)	0	0	0	0	0	0	0	0	0	0
									Sub total:	**24.5**	**6**

		Year 1		Year 2		Year 3		Year 4		Funded Man-months	
	Semester	1	2	1	2	1	2	1	2	BELSPO	other sources
Work Package 4: Accessibility of the ROB Sunspot Data Collections											
Task 4.1 Open Access data and documentation website	**C (ROB)**	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	3.5	1
	P2 (UCL)	0	0	0	0	0	0	0	1	1	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
	Sub total:									**4.5**	**1**
Task 4.2 Engagement with Scientific Communities	**C (ROB)**	0	0	0	0	0.5	0.5	1	1	2	1
	P2 (UCL)	0	0	0	0	0	0	0.5	0.5	1	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
	Sub total:									**3**	**1**
Task 4.3 Online Consultation Platform	**C (ROB)**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	2	0
	P2 (UCL)	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	2	0
	IP (UO)	0	0	0	0	0	0	0	0	0	0
	Sub total:									**4**	**0**

Total Man-months (1)	
C (ROB)	59.5
P2 (UCL)	59.5
IP1 (UO)	2

(1) BELSPO + other sources
(2) C = Coordinator, P = Partner, IP=International Partner.

3. EXPECTED RESEARCH RESULTS AND THEIR IMPACT FOR SCIENCE, SOCIETY AND DECISION MAKING

3.1 EXPECTED RESEARCH RESULTS

This project will *not* use the ROB sunspot data collections for studies of long term solar activity monitoring nor will it produce an alternative or improved sunspot number. Instead, the research will focus on new methodologies, and techniques for a better quality control, and a better access to the yet unexplored information buried in the ROB Sunspot Data Collections.

Among the highlights, we expect the following research results:

- Open Access publication of more than 70-year long coherent collection of sunspot observations from the USET facility, one of the last active telescope platforms at the ROB site in Uccle. The data set will be fully digitized, documented and contextualized.
- Besides the sunspot number that can be derived from the above dataset, we will extract a new characterization of solar activity from the USET drawings, namely the sunspot area. Also this new parameter will be fully validated and available for external researchers.
- Open Access publication of the SILSO database with a collection of more than 35 years of sunspot number observations contributed by 285 observing stations worldwide. The database will be fully documented and contextualized, including modern quality control techniques on individual stations.
- An automated procedure will be developed to identify a set of high-quality observing stations that together can form a new multi-pilot reference group.

The base research activities supporting the above developments will be published in peer reviewed journals;, we expect at least one research paper describing each dataset. The statistical work on quality control will be embedded in a PhD thesis. Finally we will work on a digital messaging system that will disseminate research results in near-real time (“Quality Issue Notifications”, “Sunspot configuration Bulletin”).

3.2 EXPECTED IMPACTS OF THE RESEARCH AND COMPLIANCE OF THE RESEARCH WITH THE EXPECTED IMPACTS

In the table below we adopt the proposed format and describe examples of the impact on science, policy and society and estimate their relative importance.

Impact on	Examples of topics	Description of the impact	Impact
Science	Studies of solar activity	The digitized USET drawing collection will provide an invaluable dataset with the position, configuration and area of every sunspot group over the last 8 solar cycles.	**75%**
	Recalibration efforts of the Sunspot Number	The USET drawing collection will be a prime candidate to extend the back-bone method used to recalibrate the sunspot number. The opening of the SILSO database will allow external researchers to participate in investigations and to engage in reproducible research.	
	Usage of the Sunspot Number in other domains	The USET and SILSO sunspot data collections will contribute to a better established International Sunspot Number which is a reference data-set to cross-calibrate and connect short ranging indices like the Total Solar Irradiance and long ranging indices such as the decays of cosmogenic radionuclides (^{10}Be and ^{14}C).	
Policy	Anticipating the impact of solar activity and space weather on human society	Solar activity and space weather have an increasing influence on our society as it becomes more dependent on space infrastructure for communication, position and other services. The solar cycle as shown by the International Sunspot Number is the only available forecast on a time-scale (years) that is relevant for political decision-making and major engineering innovation.	**20%**
Society	Differentiating correctly between man-made and natural sources of climate change	Awareness of solar variability and space weather has significantly increased in the last few years, reaching a risk for over-interpretation by the media and general public. Specifically in the climate debate, it is very important to communicate very precisely on a possible influence (or not) of solar activity on climate change. This project will not directly address such issues but contribute to the factual observations supporting the debate.	**5%**
Total			**100%**

3.3 FOLLOW-UP COMMITTEE

The appropriate committee to follow up the progress of the present project is the working group on “Coordination of Synoptic Observations of the Sun” of the International Astronomical Union (IAU, http://www.iau.org/about/). Below we present a letter from Dr. Pevtsov who co-chairs this working group and who agrees to put the results and follow-up of this project on the agenda of the IAU working group.

DECLARATION OF INTEREST IN PARTICIPATING IN THE FOLLOW-UP COMMITTEE (NOT MANDATORY - DUPLICATE AS APPROPRIATE)

First and Last name: Alexei Pevtsov
(pevtsov@email.noao.edu, http://www.nso.edu/staff/apevtsov)

Institution:
(primary) - US National Solar Observatory,
(secondary) - University of Oulu, Finland.

Describe what aspect(s) of the project is/are of interest to you and why:

I am a co-Chair of the international Astronomical Union's (IAU) Inter-division B and E working group on Coordination of Synoptic Observations of the Sun (http://www.iau.org/science/scientific_bodies/working_groups/255/). This working group is comprised of more than 50 international scientists, whose interests include preservation and a broad dissemination of historical solar observations, developing new statistical approaches to analyse historical data, and using such data in their personal research. These are the key topics that included in this proposal, and thus, the outcome of the proposed project is likely to be of interest for all scientists in this working group.

Describe with what information/data/support (money or in kind) you could contribute to the project:

The results of this project will be posted for discussion at the IAU working group, and the members will provide their feedback on intermediate results and future developments. In addition, some scientists/members of IAU WG may offer their expert advice on parts of project (e.g., preservation, data analysis) relevant to their own areas of research. These discussions will be taking part online (e.g., via internet/Google forum). As a co-Chair of the IAU WG, I will initiate and moderate these discussions upon a request from the project scientists. In addition, the group may arrange for informal face-to-face meetings with the project researchers during scientific conferences when there are a sufficient number of group's members attending a particular conference. Since these face-to-face meetings will be conducted in relation to regular conferences, no funding from the project will be required.

3.4 SUSTAINABILITY, GENDER DIMENSION, ETHICS

The present project deals with the contextualization, updating and accessibility of astrophysical data (sunspot observations) and hence has no direct relation with sustainability, gender dimensions or ethics.

As the project focuses on archival research and analysis, no animal experiments or populations are involved.

For the sunspot count observations contributed from the worldwide network of SILSO observing stations, we will take into account the data and privacy rights of each observer by asking permission to openly distribute their data.

Most of the studies will take place in Belgium and/or at the Space Pole itself, which keeps the travel limited. A maximum of the tasks will be performed with existing material and equipment, keeping thus the 'hardware' footprint of the under control.

For more details on gender issues, please see section 4.3.

4. NETWORK

4.1 PARTNERS' EXPERTISE

EXPERTISE OF THE COORDINATOR: LAURE LEFÈVRE, ROYAL OBSERVATORY OF BELGIUM

Laure Lefèvre has a multidisciplinary background. She majored in Mathematics (Classes préparatoires aux grandes écoles: 1998) and then as an engineer (Telecom Strasbourg) with a major in Physics and a Master in Astrophysics (2001). She realized her PhD in 2006 in Astrophysics in data analysis applied to massive stars, in parallel between Montréal (Université de Montréal, Canada) and Strasbourg (Université de Strasbourg, UdS). Her first post-doc was in the COROT satellite team, where she continued to apply advanced time-series and statistical analyses to the study of stellar pulsations. At the end of 2009, when the flux of data from COROT abated, she started working at the Royal Observatory of Belgium on solar data in the context of the SOTERIA[2] FP7 project.

Relevant publications (all accepted by international peer reviewed journals):

1. Clette, F., Lefèvre, L., Cagnotti, M., Cortesi, S., Bulling, A.: 2016, *The revised Brussels Locarno Sunspot Number (1981-2015)*, Solar Physics, in press.
2. Dudok de Wit, T., Lefèvre, L., and Clette, F., 2016, *Uncertainties in the sunspot numbers: estimation and implications*. Solar Physics, in press.
3. Lefèvre L., Clette F., *Survey and Merging of Sunspot Catalogs*, 2014, Solar Physics, 289, 545
4. Lefèvre, L., Clette, F.: 2011, *A global small sunspot deficit at the base of the index anomalies of solar cycle 23. Astron. Astrophys.* 536, L11. doi:10.1051/0004-6361/201118034.
5. Lefèvre, L., Clette, F., & Baranyi, T., *In-depth survey of sunspot and active region catalogs*, 2011, IAU Symposium, 273, 221

Recent research projects:
L. Lefèvre was involved in the EU FP7 SOTERIA project (http://soteria-space.eu) that helped digitize the drawings that will be used in this study, and in the EU FP7 COMESEP project (http://comesep.aeronomy.be . She was also involved in the FP7 SOLID (http://projects.pmodwrc.ch/solid/) project as well as the COST TOSCA project (http://lpc2e.cnrs-orleans.fr/~ddwit/TOSCA/Home.html). She also participated in the organization of an international series of workshops to reevaluate the International Sunspot Number (http://ssnworkshop.wikia.com/wiki/Home) and is now guest editor on a dedicated Special edition of Solar Physics.

International contacts:
The international network first developed in the context of the SOTERIA project has extended and now reaches the critical mass to support a project like this one. It includes (non exhaustively) collaborators at the Konkoly Observatory (Hungary), at the DTU (Denmark), Observatoire de Paris (France), NSO (US), Stanford (US), NASA (US), University of Tromso (Norway), University of Oulu (Finland), PMOD WRC (Switzerland) and Astrophysical Institute Potsdam (Germany).

2 SOTERIA (Nov 2008-Oct 2011) was the first space weather project funded by the European Commission (FP7).

Management skills:
The coordinator has an initial formation on management from her engineering school. She managed students of different levels (college, master) at Université de Montréal (UdM), and at ROB and taught Physics at UdM. She managed a work package dedicated to dissemination from the TOSCA project (production of movies, leaflet…).

EXPERTISE OF PARTNER 2: RAINER VON SACHS, UNIVERSITÉ CATHOLIQUE DE LOUVAIN

RAINER von SACHS is a professor at the Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain (UCL). He is a statistician with long-year expertise in time series analysis, especially locally stationary time series, on the one hand, and nonparametric methods based on wavelet analysis, recently in the modeling for multi- and high dimensional data, on the other hand. He has experience in modeling and analysis of, among others, (time-varying) biomedical time series (such as EEG), in image analysis and recently also in classification problems for astronomical data (supernovae). He has supervised both theoretical and applied PhD and master theses, has been the principal investigator of an ARC project on time-varying analysis of multivariate data, and has participated in various network projects (ARC, PAI). He has been chairman of the interdisciplinary research institute IMMAQ at UCL 2010-2012.
His international contacts are vast, with collaborations with institutions in England (Bristol, Cambridge, LSE, Lancaster), France (Paris TEC), Germany (RWTH Aachen, HU Berlin, TU Braunschweig, TU Kaiserslautern), the USA (UC Irvine), to name but a few. He is a Fellow of the IMS and an Elected Member of the ISI.

Relevant publications (all accepted by international peer reviewed journals):

1. Varughese, M., von Sachs, R., Stephanou, M, and Basset, B. (2015). Nonparametric transient classification using adaptive wavelets. MNRAS 453, 2848-2861.
2. Timmermans, C. and von Sachs, R. (2015). A novel semi-distance for measuring dissimilarities of curves with sharp local patterns. J. Statist. Planning Inference 160, 35-50.
3. Timmermans, C., Delsol, L. and von Sachs, R. (2013). Using BAGIDIS in nonparametric functional data analysis: predicting from curves with sharp local features. J. Multivariate Analysis, 115, 421-444.
4. Freyermuth, J.-M. , Ombao, H. and von Sachs, R. (2010). Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. J. Amer. Statist. Assoc. 105, 634-646.
5. Chau, J. and von Sachs R. (2016). Functional mixed effects wavelet estimation for spectra of replicated time series. Electronic Journal of Statistics, to appear.

Recent research projects:

- ARC project 12/17-045
- ARC project 07/12-002 : coordinator
- PAI network participations for 2002-2007, 2007-2012, 2012-2017
- DFG grant 2015- (co-investigator with PI: A. Steland, RWTH Aachen)
- Various FNRS and FSR grants

International contacts

- Associate Editor for the Journal of the Royal Statistical Society, Series B [2003-2006]
- Member of IMS (Fellow), ISI (Elected), Bernoulli Society, BSS
- FNRS Contact Group on “Wavelets”

EXPERTISE OF THE INTERNATIONAL PARTNER 1: THIERRY DUDOK DE WIT, UNIVERSITY OF ORLÉANS

Thierry DUDOK DE WIT is a professor at the University of Orléans, and is affiliated to the Laboratoire de Physique et Chimie de l'Environnement et de l'Espace (LPC2E, CNRS, Orléans). He's a solar-terrestrial physicist with a keen interest in statistical techniques for understanding solar variability and its impact on the Earth's environment.

Relevant publications (all accepted by international peer reviewed journals):

1. Dudok de Wit, T., Lefèvre, L., and Clette, F. (2016). *Uncertainties in the sunspot numbers: estimation and implications*. Solar Physics, in press.
2. Lilensten, J., Dudok de Wit, T., and Matthes, K. (2015). *Earth's climate response to a changing Sun*, EDP Sciences, Paris.
3. Dudok de Wit, T. (2011). *A method for filling gaps in solar irradiance and solar proxy data*. Astronomy and Astrophysics, 533:A29.
4. Dudok de Wit, T., Moussaoui, S., Guennou, C., Auchère, F., Cessateur, G., Kretzschmar, M., Vieira, L. A., and Goryaev, F. F. (2012). *Coronal Temperature Maps from Solar EUV Images: A Blind Source Separation Approach*. Solar Physics, 283:31–47.
5. Schöll, M., Dudok de Wit, T., Kretzschmar, M., and Haberreiter, M. (2016). *Making of a solar spectral irradiance dataset I: observations, uncertainties, and methods*. Journal of Space Weather and Space Climate, 6(27):A14.

Recent research projects:

- COST Action 724 on space weather: national representative
- EC FP7 Projects SOTERIA, ATMOP, and SOLID on space weather: work package leader
- COST Action 1005 on Sun-climate connections: coordinator
- NASA Solar Probe Plus mission (launch 2018): lead CoI on the search-coil magnetometer
- ESA Solar Orbiter mission (launch 2018): CoI on the search-coil magnetometer and EUI ultra-violet imager
- ESA Cluster mission (2000-2018) : CoI on the Whisper sounder
- ISSI (International Space Science Institute, Switzerland): team leader for a team solar activity forecast

International contacts:

- Working group on the revision of the sunspot number: member [2013-]
- European Space Weather Week (ESWW) Meeting, Programme Committee member [2008-2013]
- Heliospheric physics advisory group at the French Space Agency: chair [2012-2015]
- Science Committee of the International Space Science Institute (Switzerland): member [2011-2015]
- Editor for the Journal of Space Weather and Space Climate [2012-]

4.2 NETWORK'S COMPOSITION

Given the interdisciplinary nature of the proposed work (solar physics versus statistics), the network consists of two fully complementary Belgian partners: ROB provides besides the original data, the solar physics and instrumental/observations expertise, while the Université de Louvain at the other hand provides the expertise in state-of-the-art statistical techniques. A third (international) partner from the Université d'Orleans contributes specific knowledge related to the uncertainty estimates on the sunspot number.

The two Belgian partners contribute an equal number of man*months (59.5). On both sides, we believe a very cost-effective approach is proposed with the bulk of the work being done by junior collaborators, including a PhD student, while the overview by senior scientists is co-funded from the institutes. The international partner works fully on own resources.

4.3 GENDER

The present project deals with the contextualization, updating and accessibility of astrophysical data (sunspot observations) and hence has no direct relation with gender studies.

Each of the participating partner institutes in this project follows a policy of equal opportunities, both when selecting new employees as well as during the rest of the research career.

Nevertheless, we are aware that women are underrepresented in science- and technology-related professions and that visible participation of woman scientist in leading research positions can play a game changing role towards next generations of students and researchers. In this respect, we underline the importance of having a female researcher as coordinator of the project.

In general, we will adhere to the framework on gender equality as outlined in http://ec.europa.eu/research/science-society/gendered-innovations/index_en.cfm

Particular attention will be paid to gender equality in dissemination/ communications that may include an immediate interface with girl students.

5. INTERNATIONAL DIMENSION OF THE PROJECT

As ROB hosts the SILSO Wor*ld* Data Center for the Sunspot Number, the international dimension is at the heart of the project: SILSO is a worldwide observing network across more than 30 different countries.

In recent years, Dr. F. Clette (senior participant in this project) has co-organized a series of international Sunspot Workshops (Cliver et al. 2015, http://ssnworkshop.wikia.com) to channel the booming activity around sunspot number related research. An upcoming Topical Issue in the *Solar Physics* journal has further intensified the international research activity on this topic, up to a level that the manpower resources of the SILSO WDC become very challenged. This project thus comes extremely timely to strengthen the SILSO team at the international level. By preparing the ROB sunspot data collections for external researchers, more of the base research can be outsourced to international partners. In the process, the “unexplored” Belgian heritage collections will become become a reference collection at the international level for a contemporary scientific issue.

The international impact of this project stretches however beyond Solar Physics. Climate change is a global concern. Although mankind drives recent climate change, natural factors such as solar activity can offset this anthropogenic effect, in particular on longer timescales. As the Sunspot Number is the longest running observational series of solar activity, it has a central role in this debate. The results of this project will help constraining the confidence levels on the sunspot number.

Finally, on much shorter timescales, the Sunspot Number becomes increasingly important since the solar cycle is the only forecasting reference on a time-scale beyond one year. International organizations (e.g. the European Space Agency ESA) are setting up *Space Situational Awareness* programs to predict the impact of solar variability and space weather. Space technology is designed more resilient during the maximum phase of the solar cycle. The results of this work will help better characterizing the solar cycle and hence its forecasting.

6. RESOURCES

6.1 RESEARCH BUDGET

BUDGET OVERVIEW:

EURO	Coord.	Partn.2	Int.Partn.1	TOTAL
Staff	236 190	172 800	0	**408 990**
General Operation	8 000	8 000	0	**16 000**
Specific Operation	0	0	0	**0**
Overheads	12 210	9 040	/	**21 250**
Equipment	2 000	2 000	/	**4 000**
Subcontracting	0	0	/	**0**
TOTAL	**258 400**	**191 840**	**0**	**450 240**

BUDGET OF THE COORDINATOR (P1):
LAURE LEFÈVRE, ROYAL OBSERVATORY OF BELGIUM

EURO	Year 1	Year 2	Year 3	Year 4	TOTAL
Staff	77 588	77 588	40 508	40 508	**236 190**
General Operation	2 000	2 000	2 000	2 000	**8 000**
Specific Operation	0	0	0	0	**0**
Overheads	3 979	3 979	2 125	2 125	**12 210**
Equipment	2 000		/	/	**2 000**
Subcontracting					**0**
TOTAL	**85 567**	**83 567**	**44 633**	**44 633**	**258 400**

BUDGET OF PARTNER 2:
RAINER VON SACHS, UNIVERSITE CATHOLIQUE DE LOUVAIN

EURO	Year 1	Year 2	Year 3	Year 4	TOTAL
Staff	42 800	43 500	43 500	43 000	**172 800**
General Operation	2 000	2 000	2 000	2 000	**8 000**
Specific Operation	0	0	0	0	**0**
Overheads	2 240	2 275	2 275	2 250	**9 040**
Equipment	2 000	0	/	/	**2 000**
Subcontracting	0	0	0	0	**0**
TOTAL	**49 040**	**47 775**	**47 775**	**47 250**	**191 840**

BUDGET OF THE INTERNATIONAL PARTNER 1 (IP1):
THIERRY DUDOK DE WIT, UNIVERSITY OF ORLÉANS

EURO	Year 1	Year 2	Year 3	Year 4	TOTAL
Staff	0	0	0	0	**0**
General Operation	0	0	0	0	**0**
Specific Operation	0	0	0	0	**0**
TOTAL	**0**	**0**	**0**	**0**	**0**

6.2 JUSTIFICATION OF THE REQUESTED BUDGET

BUDGET JUSTIFICATION FOR THE COORDINATOR:
LAURE LEFÈVRE, ROYAL OBSERVATORY OF BELGIUM

Staff

Personnel's name (*if known*) and profile (e.g. Name, PhD in economics, 2 years experience)	M/M to be financed	Estimated full time monthly cost	Total Cost	M/M not financed
L. Lefèvre, profile: see section 4.1	22	7365	162030	0
F. Clette, PhD in Solar Physics, Head of SILSO-WDC, >25 years experience	0	N/A	0	7
O. Lemaitre, SILSO observer & IT specialist, > 8 years experience	15	3000	45000	0
Junior assistant (e.g. summer job students), to be hired	12	2430	29160	0
Petra Vanlommel, PhD Solar Physics, expertise in communication and valorization, >20 years experience	0	N/A	0	3.5
TOTAL	**49**		**236190**	**10.5**

Specific Operation

Description:	Cost
(none outside the general operations budget)	0
TOTAL	**0**

Equipment

Description:	Cost
Personal computing	2000
TOTAL	**2000**

Subcontracting

(none)	
Description of tasks:	Cost
...	0

BUDGET JUSTIFICATION FOR PARTNER 2:
RAINER VON SACHS, UNIVERSITE CATHOLIQUE DE LOUVAIN

Staff

Personnel's name (*if known*) and profile (e.g. Name, PhD in economics, 2 years experience)	M/M to be financed	Estimated full time monthly cost	Total Cost	M/M not financed
Rainer von Sachs (see 4.1)	0			9
Christian Ritter, PhD in statistics, 26 years of experience	2.5	11280	28800	0
PhD Student (To be identified)	48	3000	144000	0
TOTAL	**50.5**		**172800**	**9**

Specific Operation

Description:	Cost
(none)	
TOTAL	**0**

Equipment

Description:	Cost
Personal computing	2000
TOTAL	2000

Subcontracting

(none)	
Description of tasks:	Cost [3]
TOTAL	**0**

BUDGET JUSTIFICATION FOR THE INTERNATIONAL PARTNER 1 (IP1):
THIERRY DUDOK DE WIT, UNIVERSITY OF ORLÉANS

Staff

Personnel's name (*if known*) and profile (e.g. Name, PhD in economics, 2 years experience)	M/M to be financed by BELSPO	Estimated full time monthly cost	Total Cost	M/M financed by other means
Thierry Dudok de Wit (see section 4.1)	0	N/A	0	2
TOTAL	**0**		**0**	**2**

Operation

Description:	Cost to be financed by BELSPO	Cost financed by other means
(none)		
TOTAL	**0**	**0**

6.3 PRINCESS ELISABETH STATION CAMPAIGN BUDGET - BELARE (IF APPLICABLE)

Not Applicable.

7. APPENDICES

7.1 LITERATURE REFERENCE LIST

Arlt, R. & Fröhlich, H.-E. (2012). The solar differential rotation in the 18th century. Astron. Astroph. 543, A7

Artl, R., Leussu, R. Giese, N. Mursula, K. & Usoskin I.G. (2013). Sunspot positions and sizes for 1825-1867 from the observations by Samuel Heinrich Schwabe. MNRAS, 433, 3165

Carrasco, V.M.S., Lefèvre, L., Vaquero, J.M., Gallego, M.C. (2015). Equivalence Relations Between the Cortie and Zürich Sunspot Group Morphological Classifications. Solar Phys. 290, 1445, DOI: 10.1007/s11207-015-0679-y

Chau, J. and von Sachs R. (2016). Functional mixed effects wavelet estimation for spectra of replicated time series. Electronic Journal of Statistics, to appear.

Clette, F., Berghmans, D., Vanlommel, P., et al. (2007). From the Wolf number to the International Sunspot Index: 25 years of SIDC. Advances in Space Research, 40, 919-928

Clette, F., Lefevre, L. (2012), Are the sunspots really vanishing? Anomalies in solar cycle 23 and implications for long-term models and proxies. J. Space Weather Space Climate 2(27), A6.

Clette, F., Lefevre, L., Cagnotti, M., Cortesi, S., & Bulling, A. (2016), The Revised Brussels-Locarno Sunspot Number (1981 - 2015), Sol. Phys., in press

Clette, F., Svalgaard, L., Vaquero, J.M., & Cliver, E.W.(2014). Revisiting the Sunspot Number. A 400-Year Perspective on the Solar Cycle. Space Sci. Rev. 186, 35.

Cliver, E.W., Clette, F., Svalgaard L., and Vaquero, J.M. (2015) , Recalibrating the Sunspot Number (SN): The 3rd and 4th SN Workshops, Cent. Eur. Astrophys. Bull. 39 (2015) 1, 1–19

Comon, Pierre (1994). Independent Component Analysis: a new concept?, Signal Processing, 36, 287–314

Davies, P. L. & Kovac, A. (2001). Local extremes, runs, strings, and multiresolution. Annals of Statistics, 29, 1-65.

Dudok de Wit, T., Lefevre, L., & Clette, F. (2016). Uncertainties in the sunspot numbers: estimation and implications, to appear in Solar physics

Fryzlewicz, P. (2007). Unbalanced Haar technique for nonparametric function estimation Journal of the American Statistical Association, 102, 1318-1327.

Fryzlewicz, P. (2014). Wild Binary Segmentation for multiple change-point detection. Annals of Statistics, 42, 2243 – 2281.

Fryzlewicz, P. (2016). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. Preprint LSE, under revision.

Hoyt, D. V., Schatten, K. H., & Nesmes-Ribes, E. (1994). The one hundredth year of Rudolf Wolf's death: Do we have the correct reconstruction of solar activity? Geophys. Res. Lett. 21, 2067.

Hoyt, D.V., & Schatten, K.H. (1998a). Group Sunspot Numbers: A New Solar Activity Reconstruction. Solar Phys. 179, 189.

Hoyt, D.V., & Schatten, K.H. (1998b). Group Sunspot Numbers: A New Solar Activity Reconstruction. Solar Phys. 181, 491.

Kruskal, J.B. (1969). Toward a practical method which helps uncover the structure of a set of observations by finding the line transformation which optimizes a new "index of condensation", Pages 427–440 of: Milton, R.C., Nelder, J.A. (eds), Statistical computation; New York, Academic Press

Kundu, M.R. (1965). Solar Radio Astronomy (Interscience: New York).

Lefèvre, L., Clette, F., (2011) A global small sunspot deficit at the base of the index anomalies of solar cycle 23, Astron & Astroph, Volume 536, id.L11, 4 pp

Lefèvre, L., Clette, F., (2015) Survey and Merging of Sunspot Catalogs, Sol. Phys., Volume 289, Issue 2, pp.545-561

Livingston, W., Penn, M.J., & Svalgaard, L. (2012). Decreasing Sunspot Magnetic Fields Explain Unique 10.7 cm Radio Flux. Astrophys. J., Lett. 757, L8.

Page, E. (1954). Continuous Inspection Schemes. Biometrika, 41, 100-115.

Svalgaard, L. (2013). Solar activity - past, present, future. J. Space Weather Space Clim. 3, A24.

Svalgaard, L., & Hudson, H.S.(2010), in Understanding a Peculiar Solar Minimum, ASP Conference Series 428, eds., R. Cranmer, J.T. Hoeksema, & J.L. Kohl (Astron. Society Pacific: San Francisco), p. 325.

Willis, D.M., Coffey, H.E., Henwood, R., Erwin, E.H., Hoyt, D.V., Wild, M.N., Denig, W.F. (2013a). The Greenwich Photo-heliographic Results (1874 - 1976): Summary of the Observations, Applications, Datasets, Definitions and Errors. Solar Phys. 288, 117, DOI: 10.1007/s11207-013-0311-y

Willis, D.M., Henwood, M.N., Wild, M.N., Coffey, H.E., Denig, W.F., Erwin, E.H., Hoyt, D.V. (2013b). The Greenwich Photo-heliographic Results (1874 - 1976): Procedures for Checking and Correcting the Sunspot Digital Datasets. Solar Phys. 288, 141, DOI: 10.1007/s11207-013-0312-x

7.2 LIST OF ACRONYMS

COST	(European) COoperation in Space & Technology
CUSUM	Cumulative SUM Control Chart (see Page (1954))
DIGISUN	A software application for digitization of scanned sunspot drawings
DPD	Debrecen Photoheliographic Data
FP7	Framework Program 7 (Research Funding program of the European Commission)
IAU	International Astronomical Union
ICA	Independent Component Analysis (blind source separation method)
ISES	International Space Environment Service
ISN	International Sunspot Number
PCA	Principal Component Analysis (blind source separation method)
RGO	Royal Greenwich Observatory
ROB	“Royal Observatory of Belgium”, a Belgian Federal Scientific Institute
SIDC	Solar Influences Data analysis Center
SILSO	“Sunspot Index and Long-term Solar Observations”, a World Data Center
SOTERIA	“Solar-Terrestrial Investigations and Archives” (EU FP7 project, 2008-2011) http://soteria-space.eu/meetings.php
SVD	Singular Value Decomposition (blind source separation method)
TBD	To Be Determined
TOSCA	“Towards a complete assessment of the impact of solar variability on the Earth’s climate” (COST Action ES 1005, 2011-2015), http://lpc2e.cnrs-orleans.fr/~ddwit/TOSCA/TOSCA/Home.html
UCL	Université Catholique de Louvain
UO	Université d’Orléans
USAF	United States Air Force
USET	“Uccle Solar Equatorial Table”, a platform for solar telescopes at ROB
USSPS	ISES message format to encode “SunSpot PoSitions”
WDC	“World Data Center”