Past solar data

Status, recovery and future preservation

Frédéric Clette

WDC – SILSO, Royal Observatory of Belgium

Outline

- Types of solar data + examples
- Specific weaknesses and risks
- Broader data problems and challenges: international science data organizations
- Business models for solar repositories
- Key questions

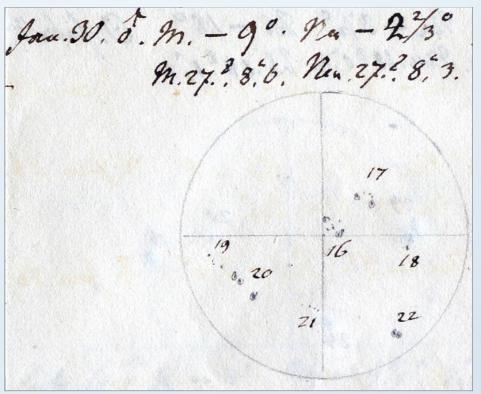
What kind of data ?

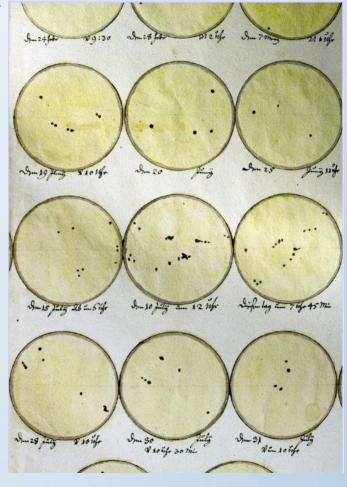
- Paper documents (printed or handwritten):
 - Drawings: sunspots, corona (eclipses)
 - Tables: sunspot counts, catalogues
 - Logbooks and reports
- Photographic images (1874 to ~2010):
 - White-light photosphere (sunspots, photometric)
 - Filtergrams/spectroscopic rasters: H-alpha (flare patrol), CaII-K
 - Magnetograms (Mount Wilson)
 - Spectra
 - Solar eclipses (corona)
- Electronic images (electron tube to CCD, after ~1970)
- Radio (paper-roll recordings, magnetic tapes):
 - Total flux (e.g. F10.7cm)
 - 1D-disc scans or 2D images
 - Spectrograms

Historical drawings

- Staudacher, Schwabe: digitization by R. Arlt (AI Postdam)
- Data not accessible: "ask the author"

(*Historical drawings resource page http://obs.astro.ucla.edu/resource2.html*)

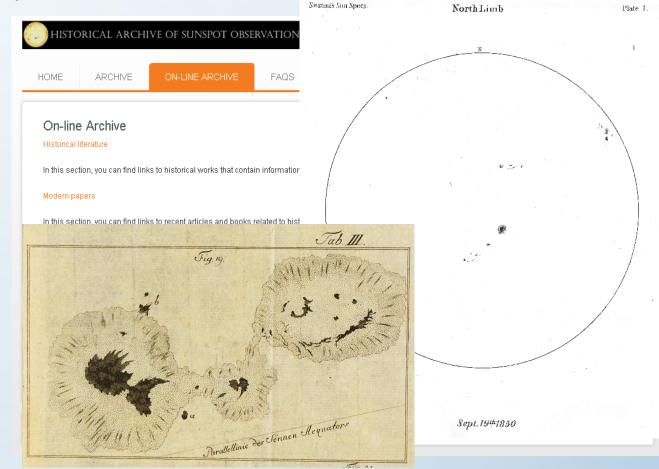




The Historical Archive of Solar Observations (HASO)

 Digital archive for sunspot drawings and related historical documents in digital form (University of Extremadura, J. Vaquero): http://haso.unex.es/

Revised group number database (Vaquero et al. 2016): catalog of raw sunspot counts by historical observers (1610-2015)



Zurich sunspot drawing collection

- Drawing collection (1882-1980):
 - Digitization started in 2016, ETH Library (Swiss federal program)



Digitization of sunspot drawings

Aim of the project

Internationally and in scientific circles, there is strong interest in the sunspot drawings stored in the ETH Zurich University Archives. By presenting detailed archive metadata and high-quality digital reproductions of original drawings online in the e-manuscripta.ch portal, access to this unique observation series will be optimized and simplified for solar research.

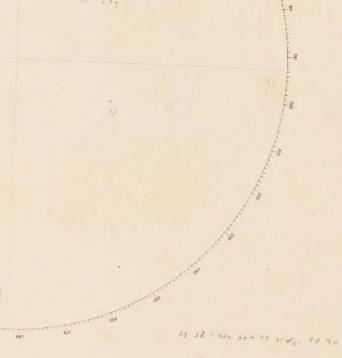
Description of the project

Observation of sunspots began in 1884 using projection image drawings at the Zurich observatory, and was passed on as an ongoing series from 1887 to 1995. In an initial phase, around 23,000 drawings of the observation staff under the respective direction of Rudolf Wolf, Alfred Wolfer and William Brunner will be indexed and digitized by the end of 2016.

From 1947, Helmut Müller assumed responsibility for recording sunspots. A subsequent project will see the indexing and digitization of the later 27,000 or so drawings from the series made at the main Zurich observatory – with additions particularly from Locarno-Monti.

Synergies and context

International solar research

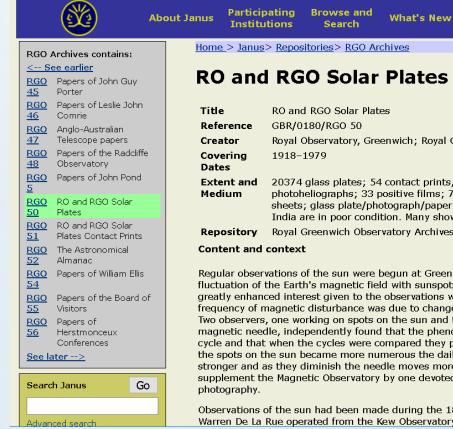


Synoptic Data, MPS, Goettingen

Nº 10

Greenwich photographic collection

- Cambridge Library (D.Willis et al. 2016) : discovery.nationalarchives.gov.uk/details/r/C15306 https://janus.lib.cam.ac.uk/db After 1976: Debrecen (A.Ludmany, T.Baranyi) •
- Loss: all plates before 1918 damaged beyond use during WWII (only contact prints remaining). •



Royal Observatory, Greenwich: Royal Greenwich O

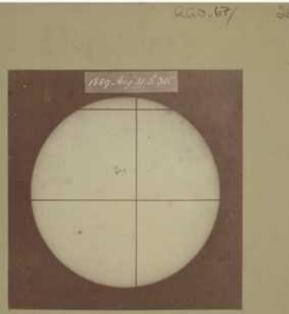
20374 glass plates; 54 contact prints, prints or p photoheliographs: 33 positive films: 7 cellulose it sheets; glass plate/photograph/paper; The plates India are in poor condition. Many show mould dar

Royal Greenwich Observatory Archives

Regular observations of the sun were begun at Greenwich because fluctuation of the Earth's magnetic field with sunspots and other greatly enhanced interest given to the observations when it was frequency of magnetic disturbance was due to changes taking pla Two observers, one working on spots on the sun and the other st magnetic needle, independently found that the phenomena they cycle and that when the cycles were compared they proved to be the spots on the sun became more numerous the daily swing of stronger and as they diminish the needle moves more feebly. It supplement the Magnetic Observatory by one devoted to the dire

Observations of the sun had been made during the 1860s on a p Warren De La Rue operated from the Kew Observatory. Early in th





Synoptic Data

Feedback

Mount Wilson photographic archive (CaII-K)

- www.astro.ucla.edu/~ulrich/MW_SPADP/ (site inaccessible!) re-digitization(L. Bertello et al. 2010):
 - Escaped full loss (single copy of raw data recovered after hard-disk crash)

/	署
2	MATA

Mt. Wilson Solar Photographic Archive Digitization Project

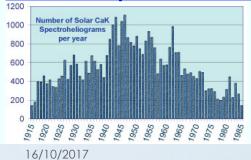
The Mount Wilson Archive

This material is based upon work supported by the National Science Foundation under Grant No.0236682 Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views (

This archive contains over 150,000 images of the sun which were acquired over a time span in excess of 100 ye stored and maintained at the Pasadena California offices of the Observatories of the Carnegie Institution of Was images called White Light Directs, ionized Ca K line spectroheliograms and Hydrogen Balmer alpha spectroheli

This project has digitized essentially all of the CaK. Solar images and step wedge images (available after 1962) original logbook parameters of observation time and scan format.

Browseable CaK images are available from 1915 to 1985.



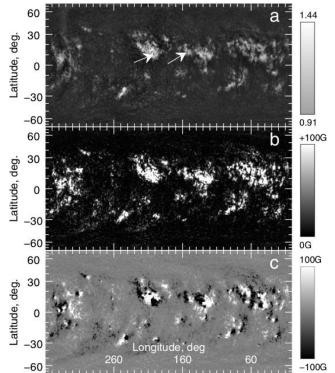
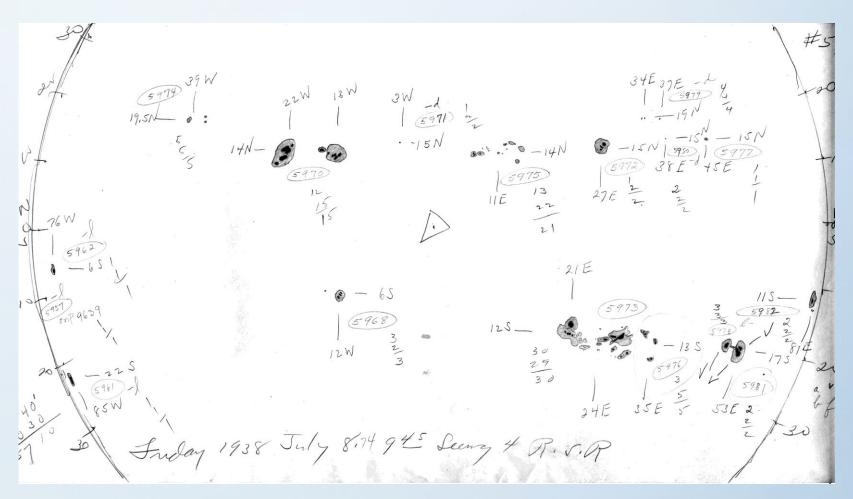


Fig. 1. Synoptic maps of the Ca K-line intensity (*upper panel*), unsigned (*middle*), and signed (*lower panel*) magnetic flux for CR1708. In the magnetogram (*lower panel*) the black and white halftones correspond to magnetic fields with negative and positive polarity. The Ca K-line observations are taken from the MWO, the magnetic field data from the NSO/KP. Two arrows point to dark circular voids (corresponding to sunspot umbrae) in the middle of bright plages. Panel c) is scaled between ± 100 G (see scale on the right side of the panel).

Mount Wilson sunspot Drawings

Simple FTP: ftp://howard.astro.ucla.edu/pub/obs/drawings/



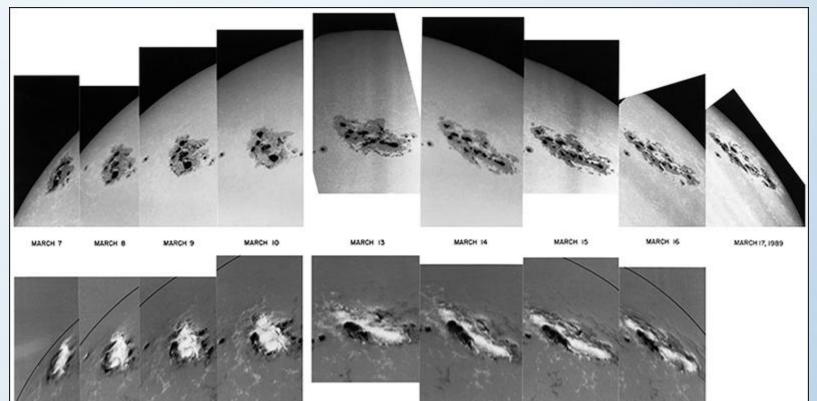
Mount Wilson catalogs and indices

• Distribution through a central platform: NOAA-NCEI

	IAL CENTERS FOR DIMENTAL INFORMATION	Search Q
NOAA > NESDIS > NCEI (formerly N		comments privacy policy
Geomagnetism Nighttime Earth Space Observations Weather Solar Data Home Online Publications	STP Home Solar Data Home Solar Calcium Pl 1.) DATA 1a.) Mt. Wilson Daily Calcium Plage Areas and Ca (Foukal 1b.) San Fernando K-line Facular Data 1988-1998 (D 1c.) McMath/Hale/Big Bear daily observations 194 NASA SPDS & NOAA ESD 2.) INDEX Daily Summary Data from M	 Ilcium Network Areas Apn Data 1915-1999) aily areas and irradiance indices Chapman) 2-1987 (positions, areas Dodson-Prince) MData Rescue
FTP Access	Daily Big Bear Calcium K-line Filt Daily Sacramento Peak Calc	-
	3.) SGD Table Regions of Solar Activity 1969-1 monthly report Solar-Geophysical Data white light plage regi	sunspot regions are tied to overlying Calcium
	Solar Calcium Data (All ftp)	
	For many years, solar observatories have monitored the Sur filter that allows only this wavelength's light through, calciun calcium K232 images are reduced to provide calcium plage p K index. Ca K active regions can be seen moving across the and higher in the solar atmosphere (chromosphere).	n filtergrams, or spectroheliograms. Daily full disk ositions and areas or summed to produce a daily Ca solar disk, above the white light sunspot regions,
	1a.) Daily Calcium Plage Areas and Network Areas from M Peter Foukal	4t. Wilson Observatory Apn Data 1915-1999

NOAO photographic archive

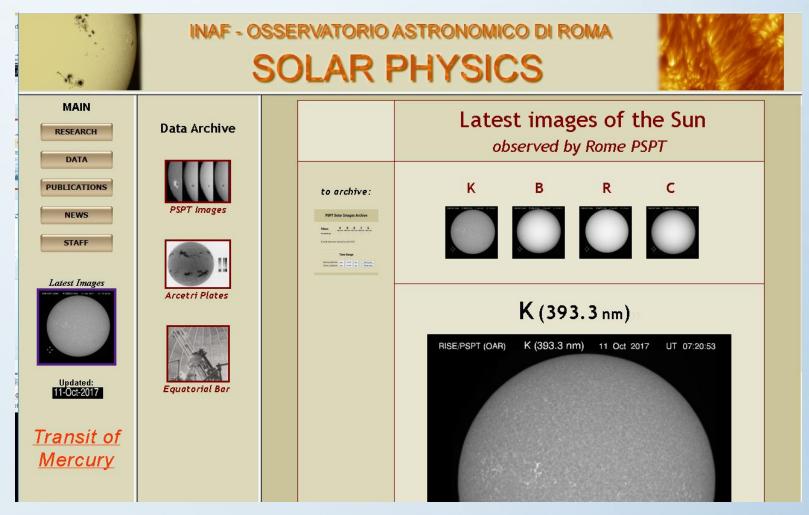
 http://archive.noao.edu/home/main: filtergrams, magnetograms



7-17 march 1989 (NOA0/AURO/NSF)

PSPT CaII K and photospheric archive

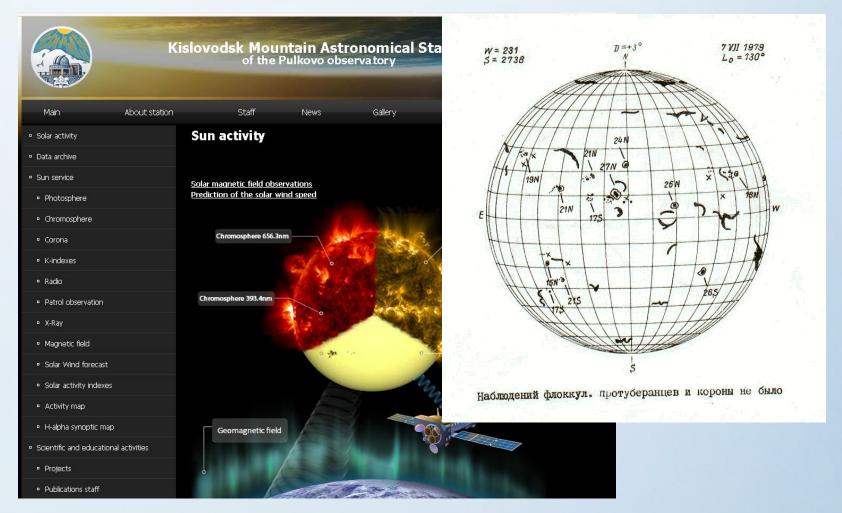
www.mporzio.astro.it/solare/



KANZELHOEHE SOLAR OBSERVATORY (Austria) Kanzelhöhe CESAR Ha 1980-09-17T10:41:58Z cesar.kso.ac.at/: drawings, Ha images Kanzelhöhe Observatory Synoptic Archives Navigation Archives Docs & Info October 12, 2017 Julian Date: 2458038.8012 » Data Policy 07:13 UT Dist. = 0.998 AU App. Diam. = 1923.9" Elevat Browse Data » Latest Images Sun - Photosphere for Today » Latest Hα (ESA SSA) » Hα Wing Test » Two Weeks Photosphere Kanzelhöhe Sunspot Drawing » Two Weeks Chromosphere No » Daily Overview Obs » Sunspot Numbers Side reversed! » Syn. Archive » Sunspot Drawings »Hα Kanzelhöhe Continuum » Whitelight 06:15 UT 06:32 UT » CallK Synoptic *.jpg *.fits.gz Synoptic *.jpg *.fits.gz » Elare Detection Normalized *.jpg *.fits.gz *.jpg *.fits.gz » Syn. Archive » Syn. Archive Normalized » Filaments » Complete Data Archive » Complete Data Archive Download Data » Archive / Ftp-Server / Local Kanzelhöhe Ha Prominence Images $R_{1} = 0$ » Fast Mirror Archive / Graz Monthly Summary 06:17 UT g = 0 Ha/GOES Intensity Plot Database Search f = 0 » Syn. Archive » Observation Database » Sunspot Numbers « 2017-10-11 « 1 Month 2017-10-12 » Observed Flares Panorama: 2017-10-12 07:03 (UT) » KSO observing logs Misc. » Solar Ephemeris 1 at 1 at 1 » Debrecen Photohel. Data » Data Policy » KSO Homepage » Institute of Geophysics, Astrophysics and Meteorology (IGAM) » University of Graz » Contact

Kislovodsk Mountain Astronomical Station

solarstation.ru: sunspot, Ha-Ca feature catalogues and maps



NAOJ drawing and CCD image archive

http://solarwww.mtk.nao.ac.jp/en/database.html

1917		Solar	国立天文台 三鷹太陽地上観測 r Science Observatory, NAOJ 2017
English / Japanese	Solar Activity Databa	ase	10 20 30 44 50 80 76 30 90 100 110 126 130 140 150 180 170 180 180 200 210 220 230 240 250 280 270 280 290 80
Home	Overview		
Dverview	National Astronomical Ob	· ·	
nstruments	Tokyo) have been conduct numerous sunspots and fro		
lesearch	example in the 17-th century there was a period of		1081 \$ "
)atabase	investigate such a long tim data are crucial.	e scale variation of the su	
Gallery	We have made the followi		
	publish an article based on	i these data, you are requi	
	Database Calendar Database		
			The IA I A I A I A I A I A I A I A I A I A
		Solar Fla	$P = -24.86$ β_{1}^{1} γ_{1}^{2} γ_{1}^{2} γ_{1}^{2} γ_{1}^{2} γ_{1}^{2} γ_{2}^{2} γ_{1}^{2} γ_{1}^{2} γ_{2}^{2}
		Observing	$B_{\mu} = -9.10$ $L_{\mu} = 99.00$
NAOJ News 2016,3	White-Light Full-Disk	White-Light,G-band Fi	$ \begin{array}{c} R \\ (196) 17 \cdot 132 \\ 9 \cdot 72 \\ C.N. \end{array} $
Special Poster	Ca-K Full-Disk	Ca-K Full-Disk Image	
(PDF,21MB,3.6MB)	Ha Full-Disk	Ha Full-Disk Images,	01413 11 00 1928-1554
	He 10830Å Full-Disk	Infrared Stokes-Polari	1 1 (16 m + 4 + 4 + 182 z.6
	Magnetograms	2010	196
0	[Infrared Stokes-Polari	

Synoptic Data, MPS, Goettingen

Multi-station data portals: BASS2000

 bass2000.obspm.fr: Paris-Meudon, Pic-du-Midi, Coimbra, USET-Brussels

BAse de donnée BASS Observations systém Jour précédent		
ACCUEIL	RECHERCHER DES OBSERVATIONS	
Dernières observations Derniers films Archive long terme Actualité	Ce formulaire permet d'interroger la base des Observations. Spécifier une ou un intervalle de dates. La recherche peut être limitée en choisissant un domaine de longueurs d'onde ou un ou plusieurs instruments. En changeant les Options de sortie , vous pouvez changer l'aspect de la page de résultats.	
RECHERCHER		
Des observations Des fichiers Des structures solaires Des cartes synoptiques HELO features cat.	Date selection 1 Jan derniers jours OU du: 01 Jan 2017 à 0 jours après OU Toutes les dates où la région NOAA numéro est observée	
Ephémérides	Longueur d'onde ou fréquence, instrument	
Spectre solaire Externes Soleil live & webcams Logiciels	Tous chromosphère photosphère DU DU Tous H_Alpha K1v DU Tous H_Alpha CU CU Tous Spectrohéliographe de Meudon Shift+clic ou clic pour une sélection multiple	
GUIDES	Options de sortie	
Instruments Données Logiciels Ressources éducatives	Imagettes: Ooui®Non Résultats/page: Un jour v Trier les résultats par: date -> v	
Collection avant 1980 Web Solaire Galerie Multimédia	Ou obtenir les observations du mois en cours pour l'instrument sélectionné	

Issues and risks

Collected and the set of a state of the second of the seco

17

Weaknesses

- Isolated initiatives:
 - by individuals or single institute
 - No coordination or common standards
 - Long-term storage depends only on local resources of a single hosting institute.
- Low accessibility and discoverability:
 - Data not accessible in-digital form
 - Basic or no on-line browsing tools (FTP)
- Users reluctant to spend efforts on the time-consuming and "invisible" data recovery
- Data at risk!
 - Improper document storage conditions
 - Important data falling into oblivion:
 - Disappearance of know-how as dedicated staff is not renewed
- Most problems resulting from:
 - Minimal or no funding for data curation work (one-person teams)
 - Disappearance of long-term structural funding schemes

Issues for **paper documents**

- Assets:
 - Information accessible without equipment (visual reading)
 - Permanently decodable (no compression or coding layer)
- Risks:
 - Slow paper and ink degradation (acid paper)
 - Accidental destruction (single copy): fire, floods
 - Archives maintained by individual scientists in local institution:
 - loss of know-how after retirements, etc.
 - Data loss (trashing of archives considered as obsolete)
- Inaccessible to direct scientific exploitation:
 - Digitization required:
 - Scanning: fast step, ensure protection against risks
 - Encoding (tables) or measuring (drawings)
 - Dispersed and non-discoverable:
 - Original historical documents scattered in their home institutions, libraries
 - Deep storage and lack of metadata
 - Restricted or paying access policy in academic libraries (property)

Issues for photographic documents

- Assets:
 - Information accessible without equipment
 - Permanently decodable/measurable (no compression or coding layer)
- Risks:
 - Slow degradation (humidity, mold, delamination): improper storage
 - Accidental destruction (single copy): fire, floods
 - Archives maintained locally in the host institution: loss of know-how, changes in institution priorities
 - Loss: trashing of collections or isolated plates considered as obsolete
- Inaccessible to direct scientific exploitation:
 - Digitization required:
 - Scanning (photometric quality): ensure protection against risks
 - Calibration: missing metadata (e.g. emulsion type and properties)
 - (Exploitation similar to other digital images)
 - Dispersed and non-discoverable:
 - Original historical documents scattered in their home institutions

Issues for old **digital-born data**

- Assets:
 - Data already in digital form (ready for computer processing)
 - Data sets often documented (readme files) and discoverable
- Risks:
 - Slow media degradation (magnetic tapes, CDs, DVDs, hard disks)
 - Accidental destruction: media erased, fire, floods
 - Archives maintained locally in the host institution: loss of knowhow, changes in institution priorities
 - Loss (uncatalogued media considered as obsolete)
- Long-term readability of the archived data:
 - Data in old compressed form
 - Obsolete file formats (sometimes proprietary): unsupported by newer software
 - Obsolete media (no more readers)

 \Rightarrow Need for an active periodic transfer to new media/formats.

Low and shrinking funding: causes

- Funding limited to short-term projects with immediate return
 - Requirement of immediate identifiable science return: excludes long data recovery projects
 - Political attraction for Big Science (big data): ignores old small-volume high-significance data sets
- Competition-based science policy
 - Quick return, mainstream topics (e.g. space weather)
 - Deters young scientists ("slow science" kills your career!)
- Unidimensional science productivity criteria:
 - Bibliometry as unique measure
 - Data production not considered as a true scientific activity
 - Science as driver of technology and economy:
 - Past data only require off-the-shelf commercial equipment/software

Long-term data recovery and maintenance is out of the picture

Typical "survival" solutions

- End-of-career project of senior scientists:
 - Small side-budgets (from annual dotation of host institute)
 - One-shot money for small investments from project participations
 - Use of volunteers, part-time colleagues, student summer jobs
 - Part-time side work >>> slooooow

Highly precarious (homogeneity and completion not guaranteed)

- Governmental digitization programs:
 - Often destined to museums and ministries:
 - Paper documents: texts
 - Pieces of art (cf. Europeana)
 - Special quantitative techniques are not foreseen (e.g. densitometers)
 - Focus on commercial return (public-private contracts):
 - On-line access to popular items (paintings, etc.) for a fee
 - Focus on data/artifact preservation:
 - Survival to steady decay
 - No more manipulation of the original

No budget for measurements or catalog construction

Broader data problems and challenges

16/10/2017

International science data organizations

A much broader problem

- Three worldwide organizations dedicated to science data
- World Data System, WDS (ICSU/UNESCO)
 - since 2008: www.icsu-wds.org
 - Coordination and accreditation of World Data Centers (data repositories & services, origins back in 1957)
 - Focus on identifying, creating, and sustaining institutions that provide stewardship, long-term preservation, and access to data.
- Committee on Data for Science and Technology, CODATA (ICSU) since 1966: www.codata.eu
 - Members: academies, science councils, international scientific unions
 - Focus on data technologies and standards: framework of standards, agreements and protocols that enable data to be shared and reused.
 - Data citation standards and practices
 - Interoperability standards
 - Citizen science and crowdsourced data



WORLD DATA SYSTEM

A much broader problem

- Research Data Alliance RDA since 2013:
 - www.rd-alliance.org



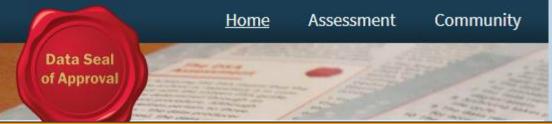
- Membership: private & non-profit organizations (publishers), individuals
- Focus on data curation practices and policies:
 - Data discoverability (permanent identifiers)
 - Data security, versioning, interoperability
 - Data repositories, metadata standards, disciplinary data registers
 - Data rescue, data service sustainability
 - Long-tail data
 - Legal and ethical aspects (Health and social sciences)
- Organization of International Data Weeks (Sept. 2016, Denver, USA): synergies, common initiatives: INTERNATION
 - Challenge of Big Data



- Trans-disciplinary data standards and interoperability
- Data traceability and permanent identifiers (CrossRef, DataCite)

Some initiatives under development

- Generalization of permanent identifiers (DOI codes):
 - Uniquely identifies an item regardless of its physical status/location
 - Also for identification of data sets
 - Tracking of different versions
- Data Seal of Approval (DSA): https://www.datasealofapproval.org
 - Created by the Data Archiving and Networked Services (DANS, Netherlands)
 - Focused on data curation practices
 - Now adopted by WDS as base for future accreditation of World Data Centers



Value of data production = peer-review publications

Putting solar data in a broader context

- Advantages:
 - Not reinventing the wheel: adopting standards, practices already in development for other domains
 - Compatibility and interoperability with other types of data
 - Leveraging effect: benefiting from the weight of a large international data community
- Disadvantages:
 - Small solar community:
 - Limited influence on adopted standards (not well adapted to the case of solar physics?)
 - Large bureaucratic organizations:
 - Heavy decisional process
 - Emphasis on advanced data techniques (large repositories) and on service accreditation:
 - Data rescue is secondary.

Putting solar data in a broader context

- Risks for small historical data repositories:
 - Higher requirements on data services (advanced standards):
 - More work for already understaffed services
 - Lack of existing know-how in the latest data technologies (small teams)
 - Organization working only as specifying bodies:
 - No corresponding support or funding for the implementation
- Only data curation standards:
 - Science validation of the data is under responsibility of disciplinary science unions

Business models

Classical institutional repository:

- Funding, staff and logistic provided by a hosting institution (University, Governmental research)
- Funding source: running annual budgets of institutions
- Open-access public data (no revenue)
- Dominant (exclusive?) model for solar data

Commercial revenues from paying users:

- All data (typically: statistical data, geographically distributed data)
- Or part of the data: value-added data products (mapping, etc.)
- Not applicable to solar data: No market:
 - Users = small community of (non-paying) scientists
 - Indirectly feeds services to end-users (ionosphere: telecom, GNSS; magnetosphere: energy distribution, space assets)

Business models: other options?

Cost-recovery mechanisms:

- Charging licence fees for any commercial downstream service
- Still open for all non-profit public-service uses:

Creative Commons type of licences



Global funding scheme for international data services:

- International source: Supervising worldwide institution provides the funding to accredited services
- Most logical solution: full superposition
- benefiting science community = accreditation body = financial source
 - Primary overarching organisations for solar physics:
 - IAU
 - ICSU
 - AGU, EGU, IUGG

Questions as conclusion

- Which is the best way to standardized-browsablediscoverable multi-site archives?
 - Promoting the implementation of standards in existing local archives?
 - Which extra budget?
 - How to acquire the required technical know-how? (training?)
 - Creation/development of central repositories/portals?
 - Where and how ?
 - How to address data property issues ?
 - Will data stored in those centers have a better life expectancy?
- Local institutional solar archives: a financially sustainable business model?
 - Alternative options? Which ones?
 - Working examples?
- Solar-specific advocacy for solar archive services or integration in international data organizations?
 - Role of IAU or other scientific unions?

Thank you for your attention

and the second second

16/10/2017