# Mirroring Technology through the World Data Centers

David Clark

WDC Panel

18th International CODATA Conference

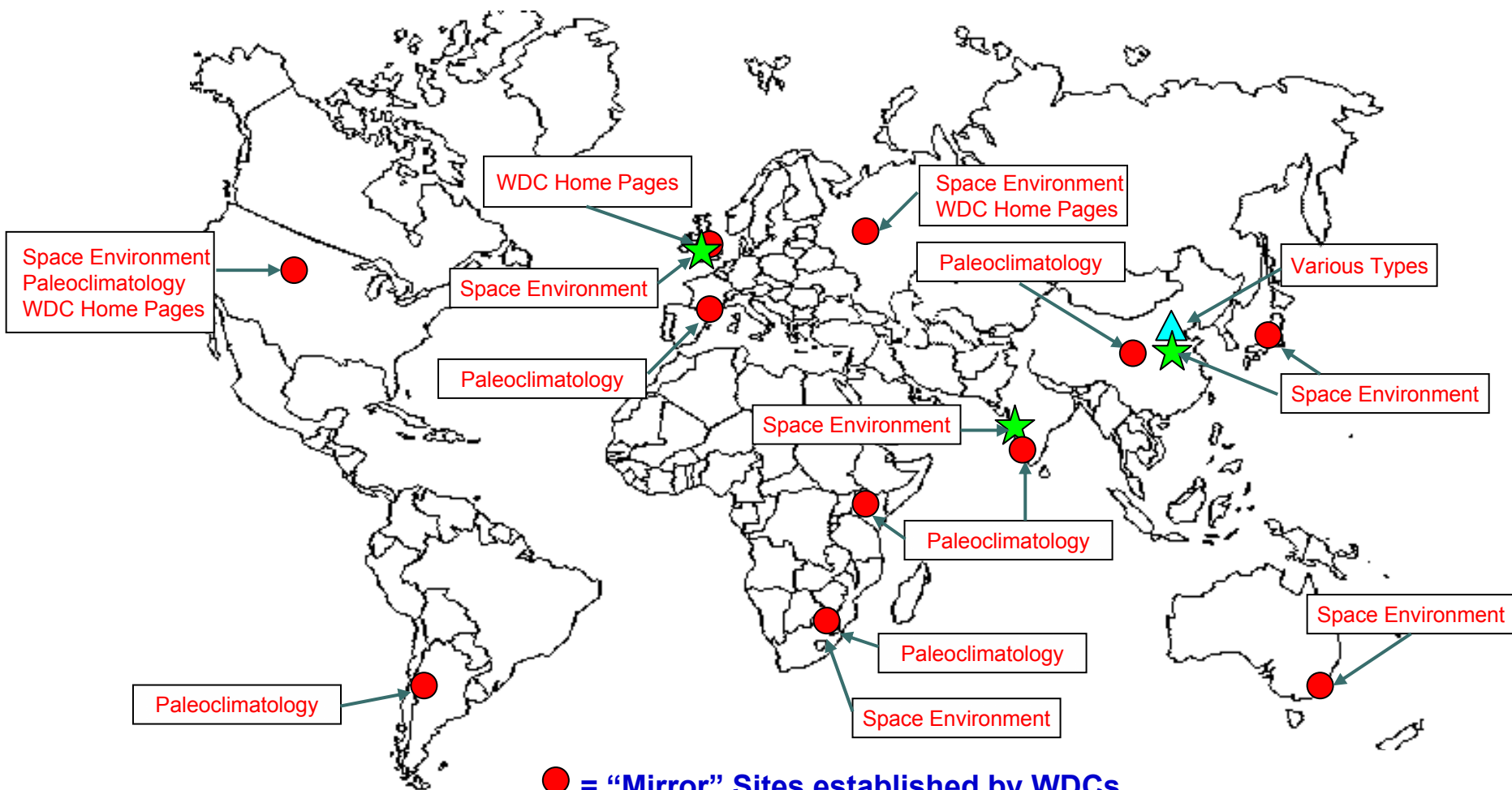October 1, 2002

# Why Establish Mirror Sites?

- Improves access between geographically separated sites
- Encourages data exchange
- Encourages new data set compilations
- Adds a regional aspects
- Builds capacity at mirror sites

# Three Types of "Mirrors"

- 1- Exact copy, i.e. true mirror
- 2- Duplicates content, mirror site designed locally to reflect regional/cultural/organizational aspects
- 3- Includes some aspects of main site which acts in a "mirror" mode; local and regional data added which can also be mirrored as appropriate

# World-wide Connectivity



WDC Home Pages

Space Environment
WDC Home Pages

Space Environment
Paleoclimatology
WDC Home Pages

Paleoclimatology

Various Types

Space Environment

Space Environment

Paleoclimatology

Space Environment

Paleoclimatology

Space Environment

Paleoclimatology

Space Environment

Space Environment

Paleoclimatology

⬤ = "Mirror" Sites established by WDCs

★ = "Mirror" Sites soon to established by WDCs

▲ = "Mirror" Sites under discussion

9/2002

# WDC Mirror sites

| WDC Paleoclimatology Program Mirrors | |
|---|---|
| *Location* | *URL* |
| NGDC, Boulder, Colorado, USA | http://www.ngdc.noaa.gov/paleo/paleo.html |
| Médias, Toulouse, France | http://medias.meteo.fr/paleo/paleo.html |
| University of Nairobi, Nairobi, Kenya | http://wdc.uonbi.ac.ke/ |
| University of the Witwatersrand, Johannesburg, South Africa | http://sunsite.wits.ac.za/paleo/paleo.html |
| Indian Institute of Tropical Meteorology, Pune, India | http://wdc.tropmet.res.in/paleo/ |
| Cold and Arid Regions Environmental and Engineering Research Institute, Lanzhou, China | http://wdc.casnw.net/paleo/ |
| **WDC Home Pages Mirrors** | |
| *Location* | *URL* |
| NGDC, Boulder, Colorado, USA | http://www.ngdc.noaa.gov/wdc/ |
| Moscow, Russia | http://plato.wdcb.rssi.ru/wdc/ |
| Chilton, UK | http://www.wdc.rl.ac.uk/wdcmain/ |
| **WDC SPIDR Mirrors** | |
| *Location* | *URL* |
| NGDC, Boulder, Colorado, USA | http://spidr.ngdc.noaa.gov/ |
| Russia | http://spidr.wdcb.ru/ |
| South Africa | http://spidr.ru.ac.za/ |
| Australia | http://spidr.ips.gov.au |
| Japan | http://gedas22.stelab.nagoya-u.ac.jp/spidr/ |

# What is *mirroring*?

**What gets mirrored in the Paleoclimatology site from Boulder?**

- 4000      Web pages (HTML)
- 4000      Images (graphics, figures, slide sets)
- 100      CGI programs (WebMapper, search forms, model output comparisons)
- 12      Java animations (temperature, climate, drought reconstructions)
- 110,000      FTP files

**What does not get mirrored**

- Oracle database searches (metadata queries; but results *are* localized)
- IDL "on-the-fly" graphics (model output comparisons)
- ArcIMS (GIS) data access

# Requirements (ideally)...

- **Unix server with (good!) Internet access**
- **10 Gb disk space (but can be less: "server minimal")**
- **Software**
  - Apache web server
  - Perl (programming language)
  - Java2 (programming language)
  - SSH (secure shell)
  - rsync (a faster, flexible remote copy program)
- Updates through JavaMail-based mirror system

# There will be days...

- **Server availability**
  - Internet connectivity: slow to very slow to non-existent
  - Electrical power problems: frequent on-battery, occasional shutdown
- **System administrators**
  - Security concerns: sudden loss of access to the server
  - Unannounced changes, e.g. Domain Name Service reorganizations
- **Sometimes at the main site!**
  - Changes that don't get mirrored correctly
  - Failure to verify that things work on the mirrors

# How it works...

- **Analyze** our web- and ftp- sites
  - Discover and correct problems, e.g. bad links or absolute addresses
- **Stage** the mirror locally
  - Localize headers for each mirror site
  - Change FTP hostnames (these are absolute references)
  - Change script paths
  - Exclude specific pages, text, or images
- **Copy** the staged material to the mirror site
- **Check** that mirroring occurred correctly

# Examples of Type One Site

- Exact mirror copies

- mostly to aid access in geographically separate locations

- WDC pages

- Paleoclimatology

- STP Sites

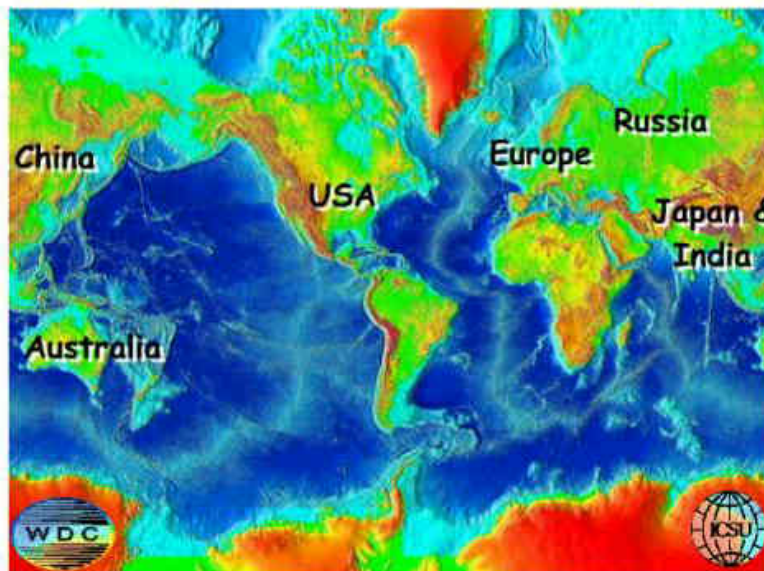*International Council for Science* - *World Data Center System*

# The World Data Center System

This web site is mirrored in Moscow, Russia and in Chilton, UK.

Click on the name of the World Data Center in the image map below to see more about that Center
The presentations from the November 2000 All WDC are available on-line.

WDC GUIDE                    WDC HOME PAGES



Russia

China    Europe

USA    Japan &
India

Australia

WDC    ICSU

LIST OF WORLD DATA CENTERS

[List of WDC's] WDC's in [USA] [RUSSIA] [EUROPE] [JAPAN & INDIA] [CHINA] [AUSTRALIA]
[WDC System Guide] [Introduction] [System] [PANEL] [APPENDIX] [WWW HOMES]
[MinutesBirmingham 1999 | [November 2000 All WDC]
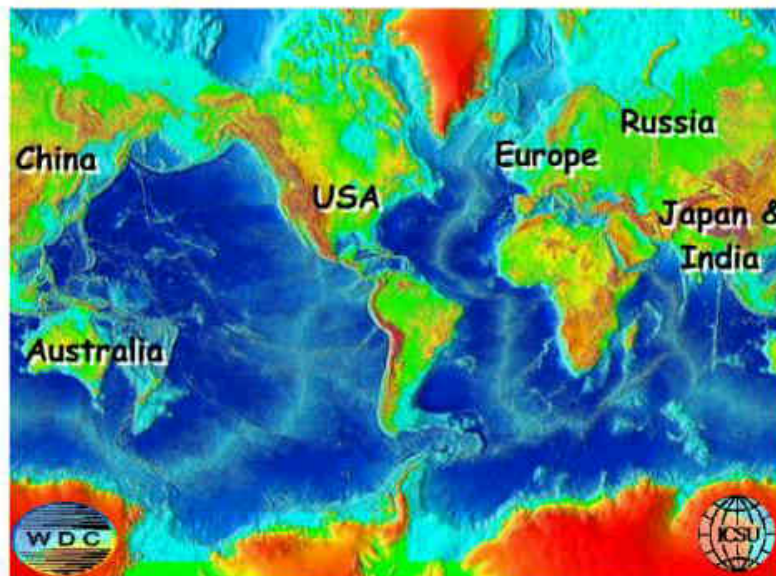
*International Council for Science* - *World Data Center System*

# The World Data Center System

This web site is mirrored in Moscow, Russia and in Chilton, UK.

Click on the name of the World Data Center in the image map below to see more about that Center
The presentations from the November 2000 All WDC are available on-line.

**WDC GUIDE**                    **WDC HOME PAGES**



China

Russia

Europe

USA

Japan & India

Australia

WDC

ICSU

**LIST OF WORLD DATA CENTERS**

[List of WDC's] WDC's in [USA] [RUSSIA] [EUROPE] [JAPAN & INDIA] [CHINA] [AUSTRALIA]
[WDC System Guide] [Introduction] [System] [PANEL] [APPENDIX] [WWW HOMES]
[Minutes Birmingham 1999 | [November 2000 All WDC]

# Examples of Type Two Site

- Content mostly identical
- Layout similar or identical
- Reflects regional data sets in addition to other data from main site
- Implemented to encourage regional data exchange
- "Selective Mirroring"
- SPIDR site
- Paleoclimatology site

**S**pace
**P**hysics
**I**nteractive
**D**ata
**R**esource

Boulder

If you are a registered SPIDR user, please specify your login here

anonymous

**Login**

If you have not used SPIDR before, you need to register

**Registration**

| Data Types | Available Dates | Initial Sampling | Coverage | Data Server |
|---|---|---|---|---|
| GOES - Geosynchronos Operational Enviromental Satellites | 1986 - 2001 | 1, 5 min | 7 satellites | Boulder |
| OMNI IMF - Interplanetary Magnetic Field | 1973 - 2001 | 1 hr | global | Boulder |
| IMF minute data | 1992 - 2001 | 1, 5 min | 3 satellites | Boulder |
| Geomagnetic Indices | 1932 - 2001 | 1, 3 hr, 1 day | global | Boulder |
| Solar Data | 1610 - 2001 | 1 day | global | Boulder |
| Ionospheric Data | 1900 - 2001 | floating: 15 min, 1 hr | 206 stations | Boulder |
| Geomagnetic Data | 1900 - 2001 | 1 min, 1 hr | 461 stations | Boulder |
| HPI NOAA Data | 1978 - 2001 | floating: about 50 min | 9 satellites | Boulder |
| HPI DMSP Data | 1983 - 2001 | floating: about 50 min | 10 satellites | Boulder |
| Cosmic Ray Data (preliminary) | 1999 - 2001 | 1 hr | 5 stations | Boulder |
| Cosmic Ray Data (4096 format) | 1953 - 2001 | 1 hr | 117 stations | Boulder |
| Cosmic Ray Data (general format) | 1951 - 1999 | 1 hr | 39 stations | Boulder |
| Sun Images | 1999 - 2000 | 1 day | global | Boulder |
| DMSP Images | 1999 - 2001 | floating: a few photos per day | 6 satellites | Boulder |
| DMSP SSJ4 Images | 2001 | floating: about 100 min | 4 satellites | Boulder |
| Nighttime Lights of the World | 2000 | no sampling | global | Boulder |
| Space Weather Events | 1975 - 2000 | random | global | Boulder |

*Review date: September 5, 2001*
*SPIDR version 2.2 of June 26, 2002*

space
physics
interactive
data
resource

This interactive documentation will help you to work with **SPIDR system.** Now please enter your login or register if you are a new user.

Moscow

If you are a registered SPIDR user, please specify your login here

anonymous

**Login**

**If you have not used SPIDR before, you need to register**

**Registration**

| Data Types | Available Dates | Initial Sampling | Coverage | Data Server |
|---|---|---|---|---|
| GOES - Geosynchronos Operational Enviromental Satellites | Jan, 1986 - Jul, 2002 | 1, 5 min | 7 satellites | Moscow |
| OMNI IMF - Interplanetary Magnetic Field | Jan, 1973 - Apr, 2001 | 1 hr | global | Moscow |
| IMF minute data | Jul, 1992 - Sep, 2001 | 1, 5 min | 4 satellites | Moscow |
| Geomagnetic Indices | Jan, 1932 - Mar, 2002 | 1, 3 hr, 1 day | global | Moscow |
| Solar Data | Jan, 1610 - Apr, 2002 | 1 day | global | Moscow |
| Ionospheric Data | May, 1900 - Sep, 2002 | floating: 15 min, 1 hr | 218 stations | Moscow |
| Geomagnetic Data | Jan, 1900 - Sep, 2002 | 1 min, 1 hr | 361 stations (separate for hr & min) | Moscow |
| SWR Geomagnetic Variations | Mar, 1996 - Dec, 1998 | 1 min | 153 stations | Moscow |
| HPI NOAA Data | Nov, 1978 - Feb, 2001 | floating: about 100 min | 9 satellites | Moscow |
| HPI DMSP Data | Jan, 1983 - Feb, 2001 | floating: about 50 min | 10 satellites | Moscow |
| Cosmic Ray Data (preliminary) | Jan, 1999 - Mar, 2002 | 1 hr | 5 stations | Moscow |
| Cosmic Ray Data (4096 format) | Jan, 1953 - Jun, 2001 | 1 hr | 120 stations | Moscow |
| Cosmic Ray Data (general format) | Jan, 1951 - Dec, 1999 | 1 hr | 39 stations | Moscow |
| Sun Images | 1991 - 2002 | 1 day | global | Moscow |
| DMSP Images | Jun, 1992 - Dec, 2001 | floating: a few photos per day | 7 satellites | Boulder |
| DMSP SSJ4 Images | Feb, 2001 - Sep, 2002 | floating: about 100 min | 4 satellites | Boulder |
| Nighttime Lights of the World | 2000 | no sampling | global | Moscow |
| Space Weather Events | 1975 - 2000 | random | global | Moscow |

SPIDR version 2.3 of September 7, 2002

# Examples of Type Three Site

- Content not identical
- Layout reflects regional aspects and programs
- Implemented to encourage regional data exchange
- Builds capacity at mirror site
- Paleoclimatology mirror site

# NOAA Paleoclimatology Program

Site Map   What's New!

**North American Drought: A Paleo Perspective**
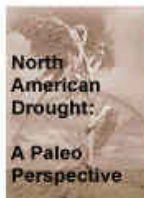
Welcome to the NOAA Paleoclimatology Program of the *National Climatic Data Center*, a central location for paleoclimatic data, research, and education located at the *National Geophysical Data Center* in Boulder, Colorado.

NOAA Paleoclimatology helps the world share scientific data and information related to climate system variability and predictability.

Our mission is to ensure the international paleoclimate research community meets the scientific goals of programs including *IPCC*, *IGBP PAGES*, *WCRP CLIVAR*, and *NOAA's Climate and Global Change Program*.

## ► Research Programs

Goals, National and International Initiatives, Funding Opportunities, Publications, Staff Directory. The Paleoclimatology Program is now a Branch of the National Climatic Data Center.

## ► Paleoclimatic Data

Access and Submit Data, Reconstructions, Search by Contributor, Search by Proxy, WDC Mirror Sites

## ► Paleo Perspectives

How paleoclimatology relates to societally relevant climate issues including Drought and Global Warming.

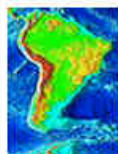## ► Education & Outreach

Introduction to Paleoclimatology, Slidesets, Related Educational Sites

## ► Other Features

Site Map, Address Exchange, Free Software, Places of Interest, What's New, Paleoclimatology Discussion List, Awards, Privacy

WDC   GLOBAL CHANGE   PAGES   CLIVAR   NSF

# Servidor Regional de los Centros Mundiales de Datos ICSU

**ICSU World Data Centers Regional Server**

▶ WDC de Paleoclimatología

▶ Centro Regional de Investigaciones Cientificas y Tecnologicas (CRICYT)

▶ Latin American Pollen Database (LAPD) Newsletter Volumen 2, No. 1 - Junio 2002

▶ IAI Climate Variability from Treeline Environments

## WDC for Paleoclimatology Mirror Site

Home • Research • Data • Education • What's New • Features • Perspectives • Site Map • Mirrors

CRICYT, Mendoza

# Data Access & Data Submission

Please Cite Data Contributors!!!

**Contributing Data**

**Obtaining Climatic Reconstructions**

**Obtaining Data by Contributor**

**Unable to find a particular Dataset?**

**Obtaining Data by Discipline**

- Borehole Data
- Climate Forcing
- Corals and Sclerosponges
- Fauna
- Historical
- Ice Cores
- Insecta
- Lake Levels
- Loess
- Paleoclimatic Modeling
- Paleolimnology
- Paleomagnetism
- Paleoceanography
- Plant Macrofossils
- Pollen
- Speleothems
- Tree Ring
- Other Paleo Data

**Searching for Subsets of Data**

- All Paleoclimatology Data Search

NEW Paleotempestology Resource Center.
Featured new data sets are listed on our
What's New page.
Recent contributions are also Listed
Chronologically.

WDC Paleo Data is also mirrored at several
sites around the world.

## Please Cite Data Contributors!

All data in the WDC-A archive is contributed by
research scientists. Contributors and
references are provided in Readme files,
PaleoVu memo files, or data file headers. Click
for a Sample Data Citation.

## How to Find Data:

All the data we have are ordered in directories
(data by contributor) that can be accessed via
the web and by anonymous FTP. Much (but not
all) of our data have been placed in relational
databases that can be queried to retrieve data
via our www forms. The contents of the
relational databases are also available as
ASCII files in directories (data by discipline)
that can be accessed via the Web and
anonymous FTP.

PAGES