



# Scientific Data: The Engine of Innovation

*- New Opportunities and Challenges for CODATA*

**GUO Huadong**

Center for Earth Observation and Digital Earth  
Chinese Academy of Sciences, Beijing 100094, China  
hdguo@ceode.ac.cn

22<sup>nd</sup> International CODATA Conference  
27 October 2010, Stellenbosch, Cape Town, South Africa



Data SIO, NOAA, U.S. Navy, NGA, GEBCO

Geo

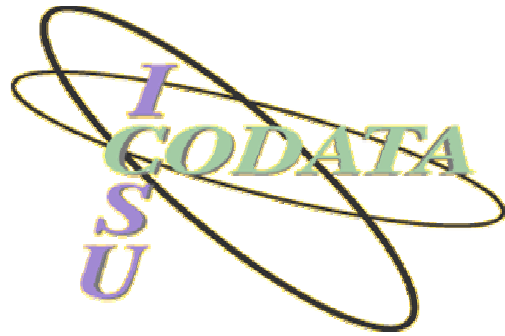
**“Data are the common wealth of humanity”**

*Adama Samassekou, Convener of the UN World Summit on the Information Society*

**“Data and information are essential building blocks of science”**

*Scientific Data and Information, ICSU*

**Data are part of life.**



**44 years ago** scientific data were recognized as one of the critical and interdisciplinary issues among the ICSU family.



- Data Policy
- Reduce Digital Divide
- Global Information Commons
- Drive the Development of New Disciplines
- Establish and Promote Data Science
- Active National Committees and Union Members



12 Sep. 2003



21 Oct. 2005



# OUTLINE

- 1. The Coming Era – Data Intensive Discovery**
- 2. New Opportunities and Challenges for CODATA**
- 3. CODATA Future – Vision, Priorities and Actions**

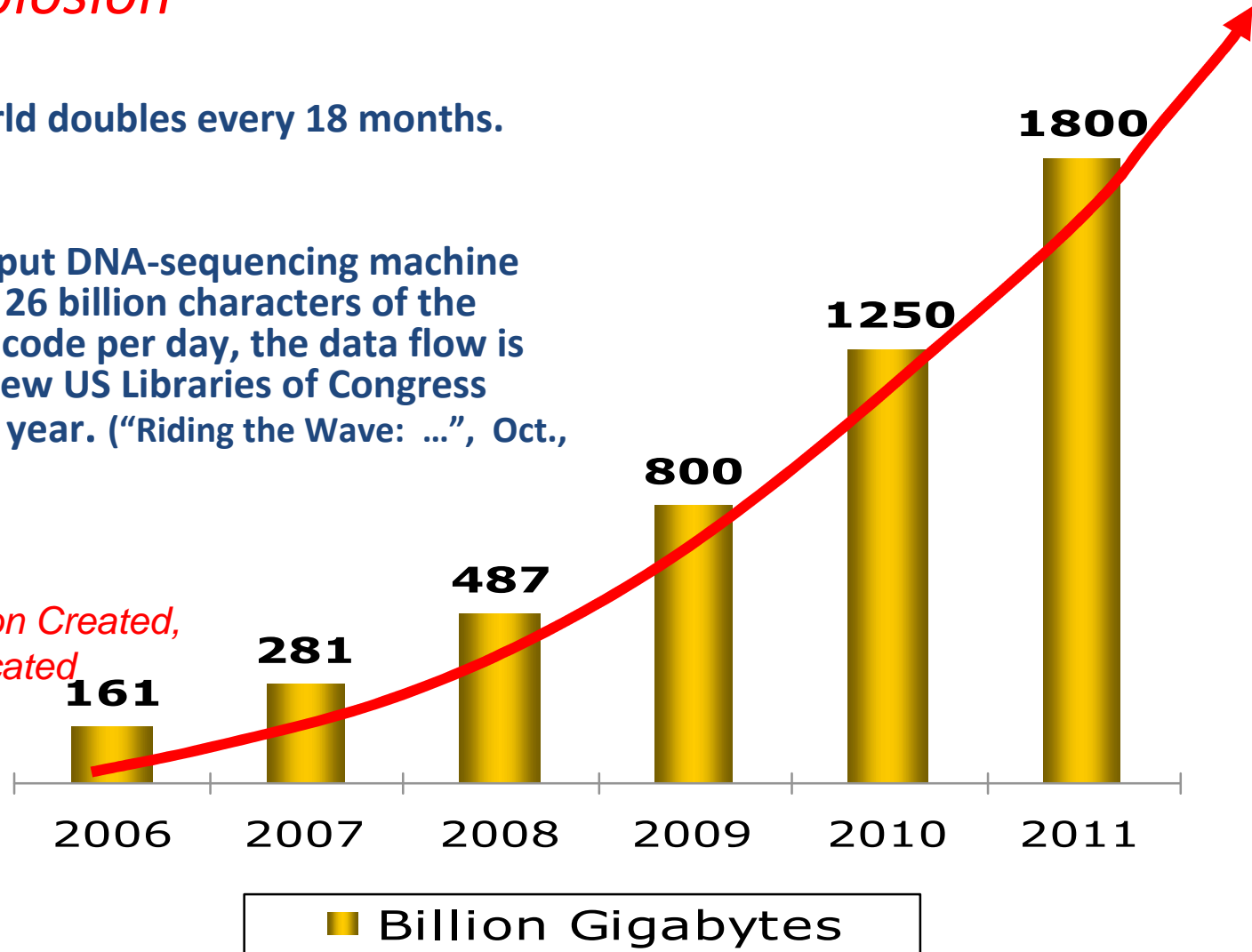
# 1. The Coming Era – Data Intensive Discovery

## ❖ *Data Explosion*

- Data in the world doubles every 18 months. (IDC, 2009)
- A high-throughput DNA-sequencing machine can read about 26 billion characters of the human genetic code per day, the data flow is more than 20 new US Libraries of Congress each and every year. (“Riding the Wave: ...”, Oct., 2010)

*Digital Information Created,  
Captured, Replicated  
Worldwide*

(Data From IDC)

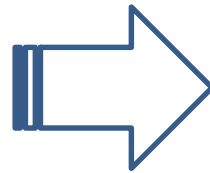


# 1. The Coming Era – Data Intensive Discovery



## *Scientific Data and Methods Change*

Characteristics of  
Scientific Data in the  
New Era



Changes to Scientific  
Methods in the New  
Era

**Complexity**

**Comprehension**

**Globalization**

**ICT Integration**

From	To
Disciplinary Research	Disciplinary, Interdisciplinary & Trans-disciplinary
Strong Emphasis on Natural Sciences	Full Range of Sciences and Humanities
Individual or Small Groups	Worldwide Communities
Research Instruments	ICT + Research Instruments

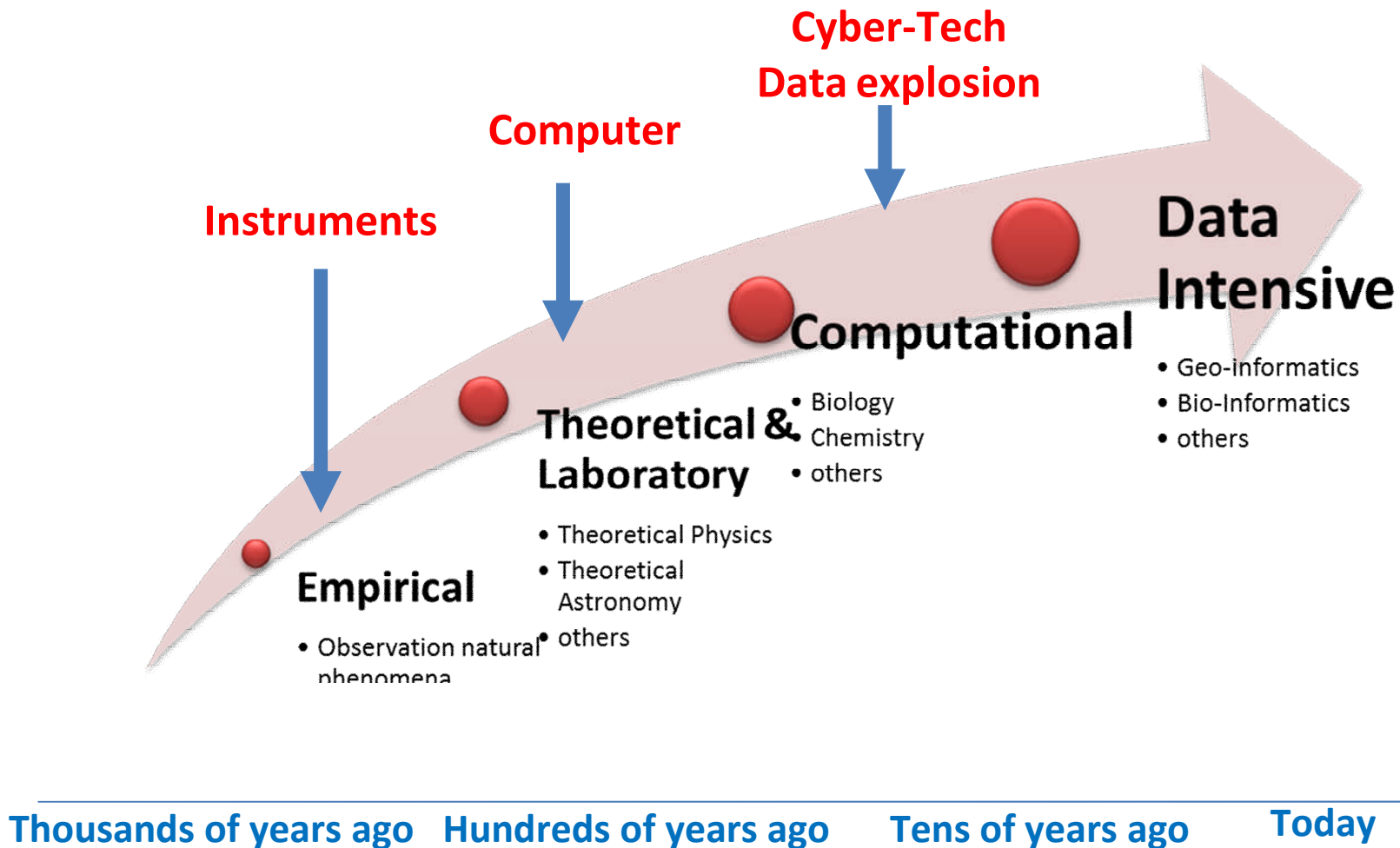


# 1. The Coming Era – Data-Intensive Discovery

## *Scientific Method Transformed*

Disciplinary Research

Disciplinary, Inter & Trans-disciplinary



# 1. The Coming Era – Data-Intensive Discovery



## *Innovations within Data Intensive Program*



DNA helix

## • Human Genome Project (HGP)



### DATA-INTENSIVE

- HGP is an international scientific research project to produce a reference sequence of human genome, which occupies over **3 billion** DNA base pairs corresponding to about **691.4 megabytes** of data.

### INNOVATIONS

- The project is anticipated to sequence and identify chemical units in the human genetic instruction set
- uncovering genetic information from such tremendous amount of data to revolutionize the ways to diagnose
- treat and even prevent a number of diseases.



The first printout of the human genome to be presented as a series of books, displayed at the Wellcome Collection, London

# 1. The Coming Era – Data-Intensive Discovery



## *Innovations Within Data Intensive Programs*



Human\_Genome



.....

**More and more innovations are happening in data intensive programs.**

## 2. New Opportunities and Challenges for CODATA



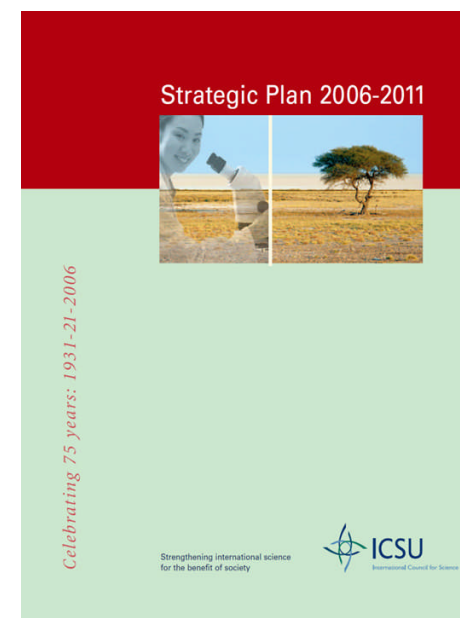
*CODATA Can Play a Key Role in the New Era*

### CODATA'S MISSION

To strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.

### ICSU STRATEGIC VISION

“The long-term ICSU strategic vision: ... In such a world, universal and equitable access to high quality scientific data and information is a reality and all countries have the scientific capacity to use these and to contribute to generating the new knowledge that is necessary to establish their own development pathways in a sustainable manner.”



## 2. New Opportunities and Challenges for CODATA



### *CODATA SWOT Analysis*

#### **S**trengths

- Cross-disciplinary
- Membership
- Task Group
- Data Science Journal
- Biannual Conference
- Open Data Policy

#### **W**eaknesses

- Human Resources
- Limited Membership
- Finances
- Organizational Structure
- Lack of Resources

#### **O**pportunities

- More Awareness to Scientific Data
- ICT & Cyber-infrastructure
- Global and Inter-Disciplinary International Programs

#### **T**hreats

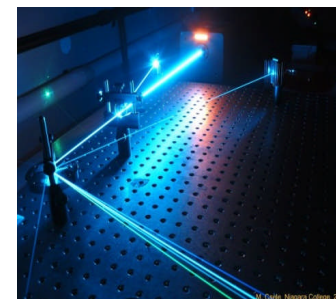
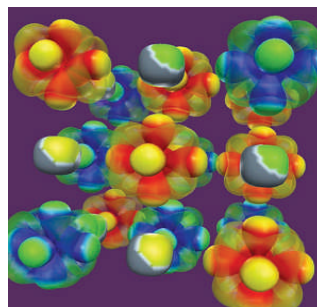
- Duplication
- Strategy
- Financial Resources
- Cultural and Policy Barriers

## 2. New Opportunities and Challenges for CODATA



### *CODATA SWOT Analysis: STRENGTHS*

- **Cross-disciplinary focus on scientific data**
- **Broad international and multi-disciplinary membership**
- **Effective Secretariat & a number of strong National Committees**
- **Active Task Groups**
- **Data Science Journal**
- **Biennial Conference**
- **Open Data Policy**



## 2. New Opportunities and Challenges for CODATA



### *CODATA SWOT Analysis: WEAKNESSES*

- **Limited human resources**
- **Membership**
  - lack of members from western Europe and developing world.
- **Finance**
  - does not now have a steady source of supplement income almost entirely on the dues paid by the national members.
- **Organizational structure**
  - limits its flexibility to respond quickly to time-sensitive issues and to adapt quickly to changing needs.
- **Lacks the service capacity and resources**

## 2. New Opportunities and Challenges for CODATA



### *CODATA SWOT Analysis: OPPORTUNITIES*

- Scientific data explosion
- Data-intensive scientific discovery as the 4th paradigm transformed scientific method
- ICT & Cyber-infrastructure provide new ways to harness the power of data
- Global scale science requires the ability to share global data and information
- Global challenges (e.g disasters, public health) call for new knowledge mining from data
- The awareness of open access to scientific data is widely accepted

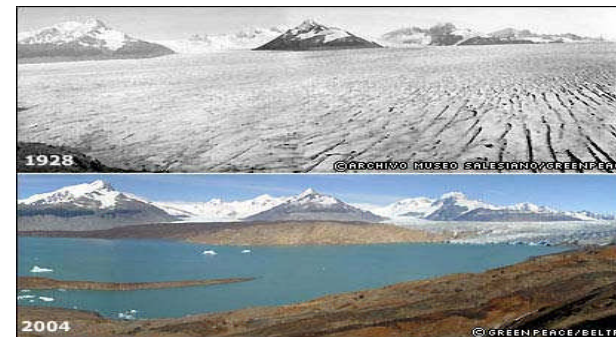


## 2. New Opportunities and Challenges for CODATA



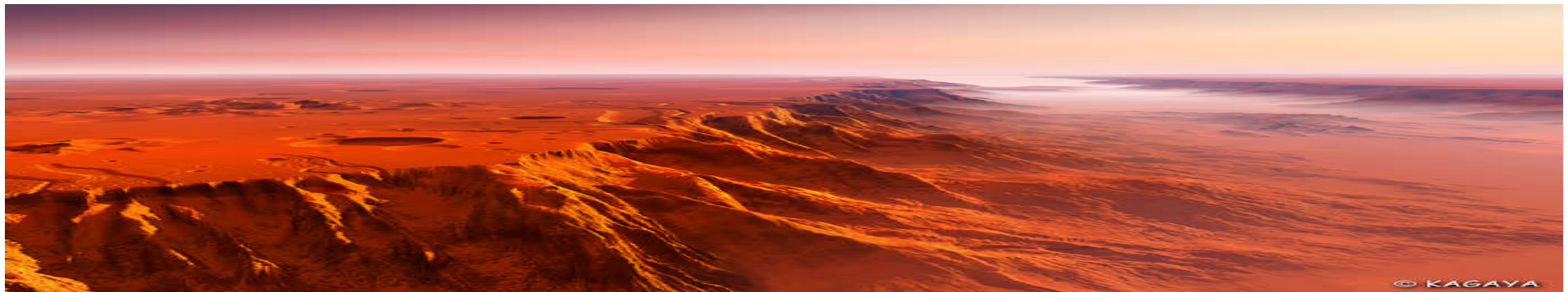
### *CODATA SWOT Analysis: THREATS*

- Duplication with other similar organizations
- Long-term clearly Strategy
- Limited Resources
- Cultural and Policy Barriers



## 2. New Opportunities and Challenges for CODATA

A clear Long-Term Strategy for CODATA is required and must be developed.



### 3. CODATA Future - Vision, Priorities and Actions



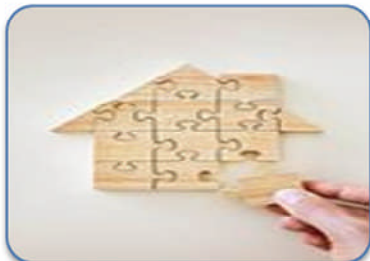
*Vision: A World Leader of Data Communities*



**The lead organization promoting globally coherent data policies.**



**Sink and source of data, benefiting science, society and the ICSU Community.**



**Home of data scientists and the platform for research and education communities.**

# 3. CODATA Future - Vision, Priorities and Actions



## *Priorities*

- **Initiating and coordinating International Programs and data intensive research worldwide.**
- **Developing coherent data policies at all levels and continuing to reduce the Digital Divide.**
- **Improving CODATA's financial situation.**
- **Demonstrating utility of CODATA to broaden services.**

# 3. CODATA Future - Vision, Priorities and Actions



## *Action 1*

- **The Hand-in-Hand Program: accelerate co-operation with other International Programs and Organizations: Global Change, Natural Disaster, Public Health, New Energy...**
  - **ICSU Programs in the coming decade**
    - IRDR , PECS ,IPY,COSPAR...
  - **Global Change and Other Programs**
    - ESSP, GEOSS ...
  - **Other Data Bodies**
    - WDS, INASP ...
  - **United Nations Programs and Governments**
  - **CODATA National Members and Unions**

# 3. CODATA Future - Vision, Priorities and Actions

## The Hand-in-Hand Program



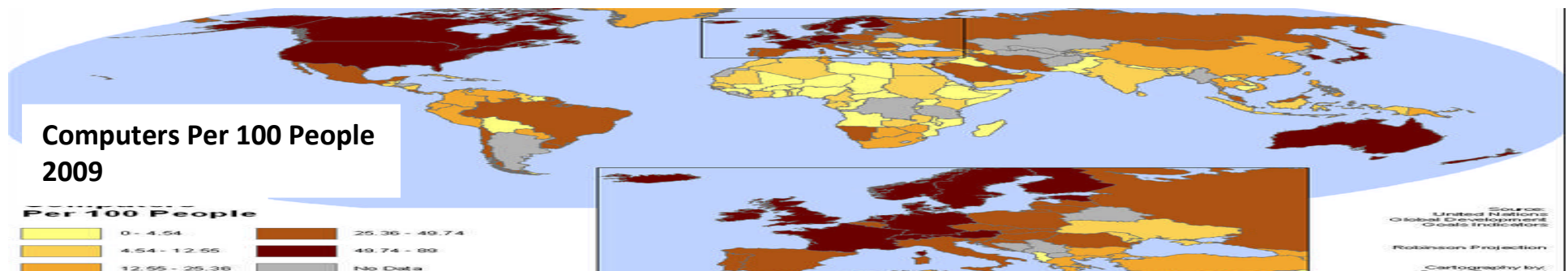
Together with CODATA National Members, Union Members ...

# 3. CODATA Future - Vision, Priorities and Actions

## *Action 2*

### Enhancing Global Coherence Data Policy

- Identify the barriers of data access policies at different levels
- Develop Global Data-Sharing Principles, Frameworks and Policies
- Demonstrate the benefits of rapid data access with real examples
- Show these results to decision makers



# 3. CODATA Future - Vision, Priorities and Actions



## *Action 3*

**Extending the cooperative network to International Financial and other Institutions ,etc.**

**Diversity of membership**

- **Attract new members and institutions – focus on Developing Regions**
- **Expanding in new regions and countries, in partnership with Regional bodies**
- **Establish new Member Categories**

**Approach National Governments for Long-Term Support**



# 3. CODATA Future - Vision, Priorities and Actions



## *Action 4*

- **Establish an International Data Academy**
- **Task Groups**
  - Publicize recommendations & positions – Press Releases
  - Encourage Task Groups in National Committees
- **Publications**
  - Increase the impact of the Data Science Journal
  - Produce Strategy Reports and Position Papers
- **CODATA Conference**
  - Wider coverage and participation
- **Involve more Young Scientists**

# Closing Remarks

**CODATA has to expand and extend its Programs to grow, for the benefit all Governments and Societies, by:**

- **Initiating new Programs and Data-Intensive research on a Global scale;**
- **Developing sound Data Policies at all levels, and reduce the Digital Divide;**
- **Demonstrating the uses of CODATA beyond the Scientific community;**
- **Enhancing National and International Cooperation among and between Governments and Institutions.**

**When in place, these will greatly improve CODATA's abilities and financial outlook, and generate more opportunities.**



**Thank you**