CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, Beijing , China, 5-20 June 2014

Overview of things learned

Presentation at NeDICC Meeting on 16 July 2014

Johann van Wyk

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Participants

Participants

21 People from the following countries:

South Africa 3

Brazil 1

Kenya 4

Tanzania/Zanzibar 1

Uganda 1

India 5

Vietnam 2

Indonesia 2

Mongolia 2

Programme

The programme was held at the Computer Network Information Center, of the Chinese Aacademy of Sciences, and consisted of lectures from various international scholars and scholars from the Academy of Sciences in China.







Workshop on Big Data for International Scientific Programmes: Challenges and Opportunities, 8-9 June 2014

We also attended this 2 day Workshop/Conference on 8-9 June 2014 in Beijing, where experts on Big Data Management from across the world came together

Organisations represented

futureearth
research for global sustainability

ICSU CODATA

RDA
RESEARCH DATA ALLIANCE

Center for International Earth Science Information Network, EARTH INSTITUTE, COLUMBIA UNIVERSITY

GEO GROUP ON EARTH OBSERVATIONS

NATIONAL ACADEMY OF SCIENCES

Computer Network Information Center, CAS

UNIVERSITY OF GOTHENBURG
Dept of Earth Sciences

WDCM World Data Center for Microorganisms

Thetherless World Constellation

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

UNITED NATIONS UNIVERSITY
Institute for environment and Human Security

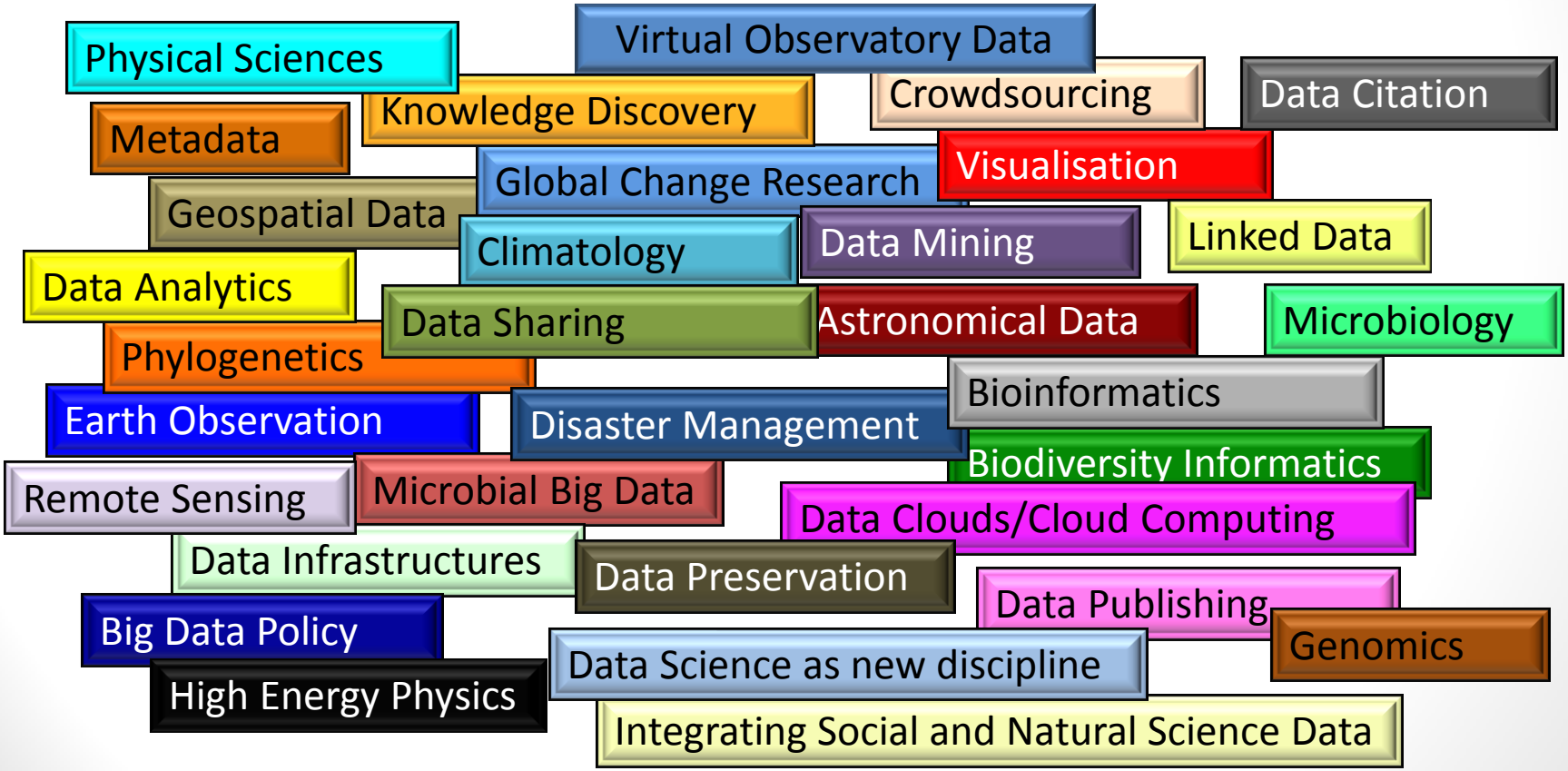
RADI
Institute of Remote Sensing and Digital Earth, CAS

IRDR
Integrated Research on Disaster Risk

International Society for Digital Earth
ISDE

Topics Covered

A very good overview of international collaborations, developments and applications of big data and its management in various disciplines internationally and in China



Things learned/of value/of interest

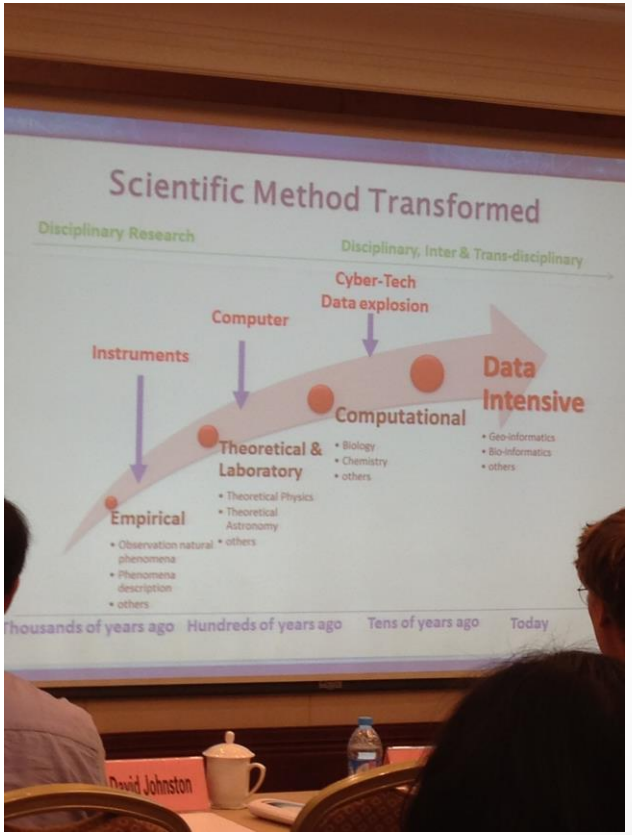
A clearer understanding of the concept of Big Data and all its facets

Prof GUO Huadong's presentation on 8 June at Workshop on Big Data:

"Current characteristics of big data:

- Relative characteristics: denotes those datasets which cannot be acquired, managed or processed on common devices within an acceptable time
- Abolute chacteristics defines big data through Volume, Variety, Veracity and Velocity"

Paradigm shift from model-driven to data-driven science



Data Intensive Research - Inter- & Intra-disciplinary

Emergence of a Fourth Research Paradigm

Thousand years ago – **Experimental Science**

- Description of natural phenomena

Last few hundred years – **Theoretical Science**

- Newton's Laws, Maxwell's Equations…

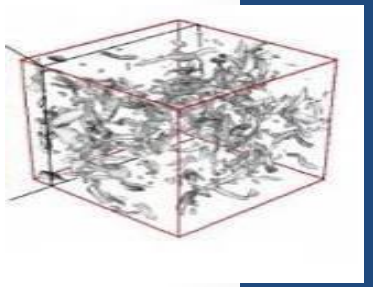
Last few decades – **Computational Science**

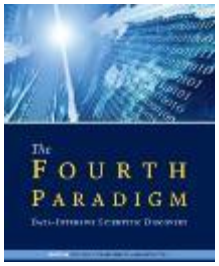
- Simulation of complex phenomena

Today – **Data-Intensive Science**

- Scientists overwhelmed with data sets from many different sources
  - Captured by instruments
  - Generated by simulations
  - Generated by sensor networks

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$





eScience is the set of tools and technologies to support data federation and collaboration

- For analysis and data mining
- For data visualization and exploration
- For scholarly communication and dissemination

Thanks to Jim Gray

Courtesy of Prof. Tony Hey

From Presentation by Chenzhou Cui on 18 June 2014 at CODATA Workshop, Beijing China

A Modern Scientific Discovery Process

Data Gathering (e.g., from sensor networks, telescopes…)

↳ **Data Farming:**

Storage/Archiving
Indexing, Searchability
Data Fusion, Interoperability
} Database Technologies

↳ **Data Mining** (or **K**nowledge **D**iscovery in **D**atabases):

Pattern or correlation search
Clustering analysis, classification
Outlier / anomaly searches
Hyperdimensional visualization

Key Technical Challenges

Key Methodological Challenges

↳ **Data Understanding**

↳ **New Knowledge**

+feedback

Courtesy of Prof. S. G. Djorgovski

From Presentation by Chenzhou Cui on 18 June 2014 at CODATA Workshop, Beijing China

Things learned/of value/of interest

More deliberations on the concept of Big Data:

- What is your understanding on Big Data & Data Science?
- What are changing or will be changed in the Big data era when we do science?
- What should we do in order to embrace this new opportunity?

From Presentation by Jianhui LI, 11 June 2014, at CODATA Workshop, Beijing China

Some Comments/Views on Big Data

- **High dimensionality** may lead to **wrong** statistical inference and **false** scientific conclusions.
- Big Data creates issues of heterogeneity, experimental variations, and statistical biases, and requires us to develop **more adaptive and robust procedures** to handle the challenges of Big Data, we **need new statistical thinking and computational methods**.
- Old style science coped with nature's complexities by seeking the underlying simplicities in the sparse data acquired by experiments. But Big Data **forces scientists to confront the entire repertoire of nature's nuances and all their complexities.**

https://www.sciencenews.org/blog/context/why-big-data-bad-science

From Presentation by Jianhui LI, 11 June 2014, at CODATA Workshop, Beijing China

Some Comments/Views on Big Data

- Prediction and understanding have been intimately tied together in science, but the influence of big data in science is now breaking them apart. **HOW CAN YOU PREDICT something without understanding it**? Simple: Find some other phenomenon that tends to occur with the event you're trying to predict.
- With big data, it turns out that almost **everything in nature and society has a tale**, one that can be discovered with sophisticated computer models that run on inexpensive hardware and crunch through terabytes of data. If you measure enough variables, it doesn't matter whether you understand the **relationship between cause and effect**; all you need is **a relationship between one variable and another**.

http://www.psmag.com/navigation/nature-and-technology/big-data- changing-science-society-69650/

From Presentation by Jianhui LI, 11 June 2014, at CODATA Workshop, Beijing China

Some Comments/Views on Big Data

- **Hypothesis-driven research** is designed to answer specific questions about **cause and effect**
- A big data scientist considers hypothesis- driven science too limiting. For example cancer is a complex disease, involving many genes, and **we'll never understand it if we get bogged down in the time-consuming process of testing cause-and-effect relationships one at a time**. Instead, we can tackle cancer much more effectively by measuring as many variables as possible in as many cancers as we can collect, **without being biased** by preconceived ideas
- It's **not a question of data correlation versus theory**. The use of data for correlations allows one to test theories and refine them.

http://www.psmag.com/navigation/nature-and-technology/big-data- changing-science-society-69650/

The promise and peril of big data by aspen institute

From Presentation by Jianhui LI, 11 June 2014, at CODATA Workshop, Beijing China

Some Comments/Views on Big Data

- Data is not just a back-office, accounts-settling tool any more.

  It is increasingly used as **a real-time decision-making tool**.
- New forms of computation combining statistical analysis, optimization and artificial intelligence are able to construct statistical models from large collections of data to **infer how the system should respond to new data.**
- Big data suddenly changes the whole game of how you look at the ethereal odd data sets. Instead of identifying **outliers** and "cleaning" datasets, theory formation using big data allows you to "craft an ontology and subject it to tests to see what its predictive value is.
- In other words, the data once perceived as "noise" can now be reconsidered with the rest of the data, leading to new ways to develop theories and ontologies. Look at how can you invent the **"theory behind the noise"** in order to de-convolve it, and to find the pattern that you weren't supposed to find.

From Presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

Some Comments/Views on Big Data

- Big data are big in four ways: Volume, Variety, Veracity and Velocity.
- **Volume:** The scale of data that systems must ingest, process and disseminate;
- **Variety:** the complexity of the types of information handled (many sources and types of data both structured and unstructured)
- **Velocity:** the pace at which data flows in and out from sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices
- **Veracity:** refers to the biases, noise and abnormality in data

http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/

4 H's of Scientific Big Data

- **Heterogeneity and Non-reproducibility**

  "Everything changes and nothing remains still ... and ... you cannot step twice into the same stream" – *Heraclitus of Ephesus*

  Statistics may or may not work
- **High uncertainty**

  Observation, sampling, record, uncertain models, . . .
- **High dimension Multi-sourced**

  Mathematical models: Fourier Transform, Wavelet, Sparse representation, Machine learning, . . .
- **High computational complexity**

From Lizhe Wang's presentation at CODATA Workshop on 9 June 2014, Beijing, China

Data as Resource

- **Ubiquitous:** available anytime and anywhere Non-rivalrous: one person's use of it does not impede another 's
- **Hyper-renewable:** data do not diminish when it is used; it can be processed again and again and its consumption creates even more data
- **Accumulative:** Data's value usually increases when it is used
- Big Data is like Oil or Soil?
- Data is Resource

Big Data is like Oil or Soil?

From Jianhui LI's presentation at CODATA Workshop presentation, 11 June 2014

- He used Information from Yike GUO' lecture

How to use this resource

- **“Mining ”**
  to extract the actionable knowledge from the data by understanding the correlations, causalities and to predict future, just like digging for nuggets
- **“Transduction”**
  to explore the optional use of the data to gain new value, just like transforming energy
- **“Interaction”**
  to enable the “data chemistry” approach to unleash the value of combined data resource; just like organic reaction and mineral composition

From presentation by Jianhui Li at CODATA 11 June 2014, Beijing, China

- He used information from Lecture in CNIC by Yike GUO

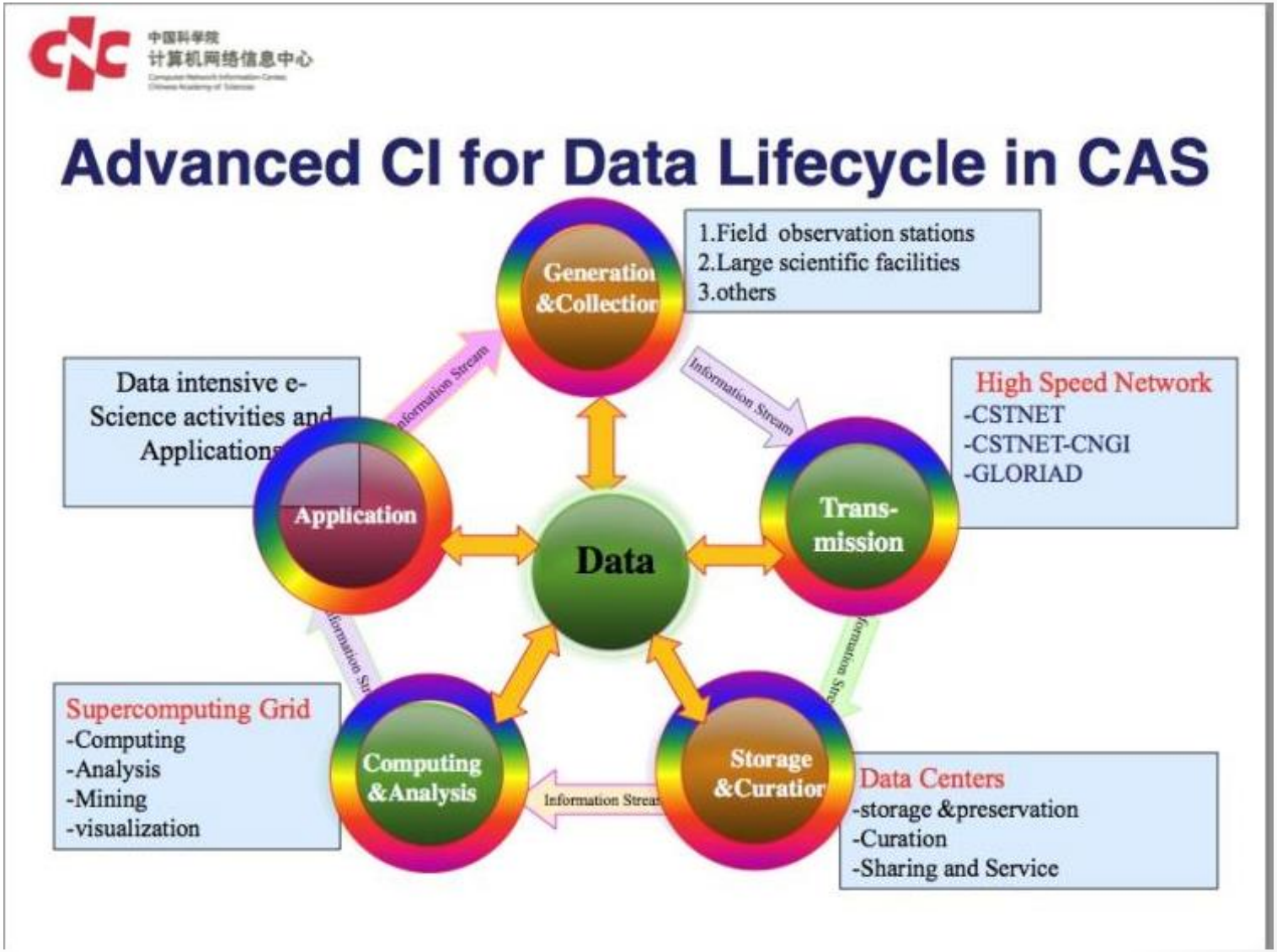
Discovery science – we need Abduction!

- a method of *logical inference* introduced by C. S. Peirce which comes prior to induction and deduction for which the colloquial name is to have a "hunch"

Human intuition is needed in interacting with large-scale data

From presentation by Peter Fox on 8 June at CODATA Workshop on Big Data for International Scientific Programmes, Beijing, China

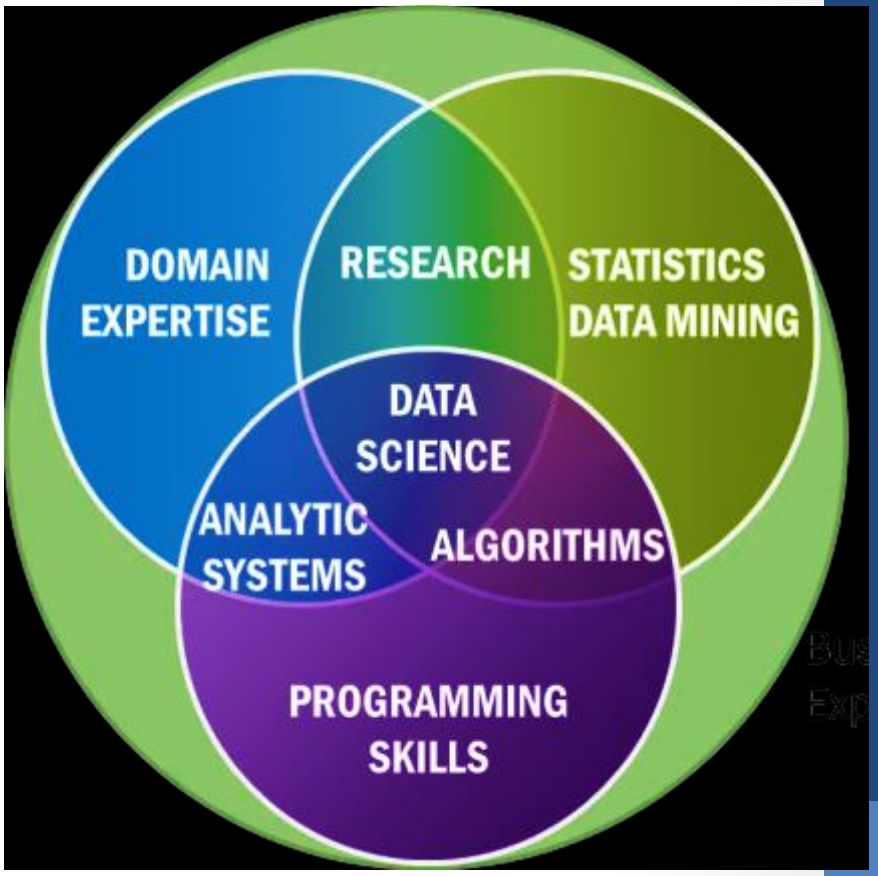
Managing Big Data - CAS



From presentation by Danhuai Guo at CODATA Workshop on 13 June 2014, Beijing, China

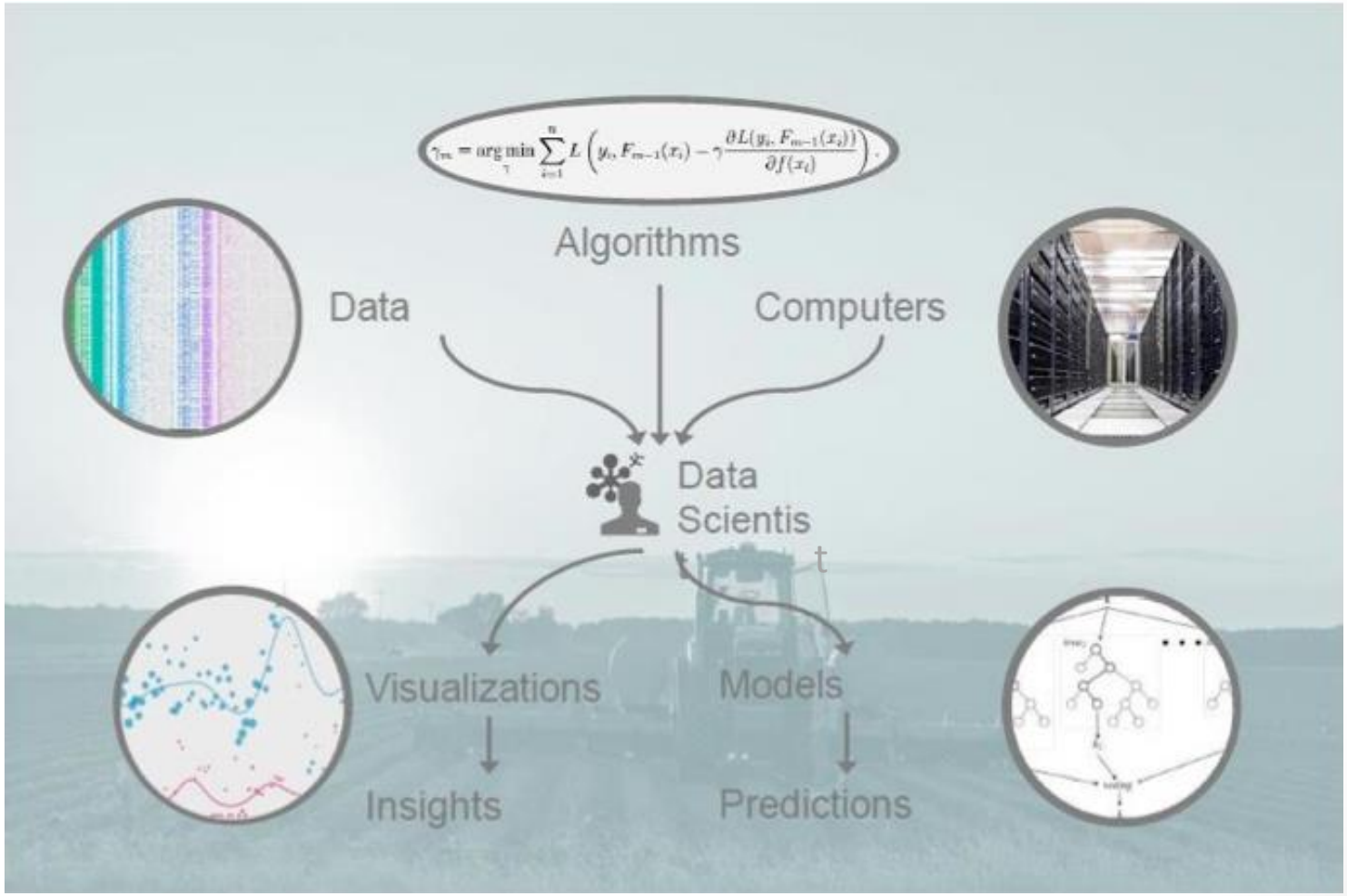
Data Science as a new discipline

- **Data Science** is the extraction of actionable knowledge directly from data through a process of discovery, hypothesis, and analytical hypothesis analysis.
- **Data Scientist** is a practitioner who has sufficient knowledge of the overlapping regimes of expertise in business needs, domain knowledge, analytical skills and programming expertise to manage the end-to-end scientific method process through each stage in the Big Data lifecycle (through action) to deliver value.



From presentation by Wo Chang on 9 June at CODATA Workshop on Big Data for International Scientific Programmes, Beijing, China

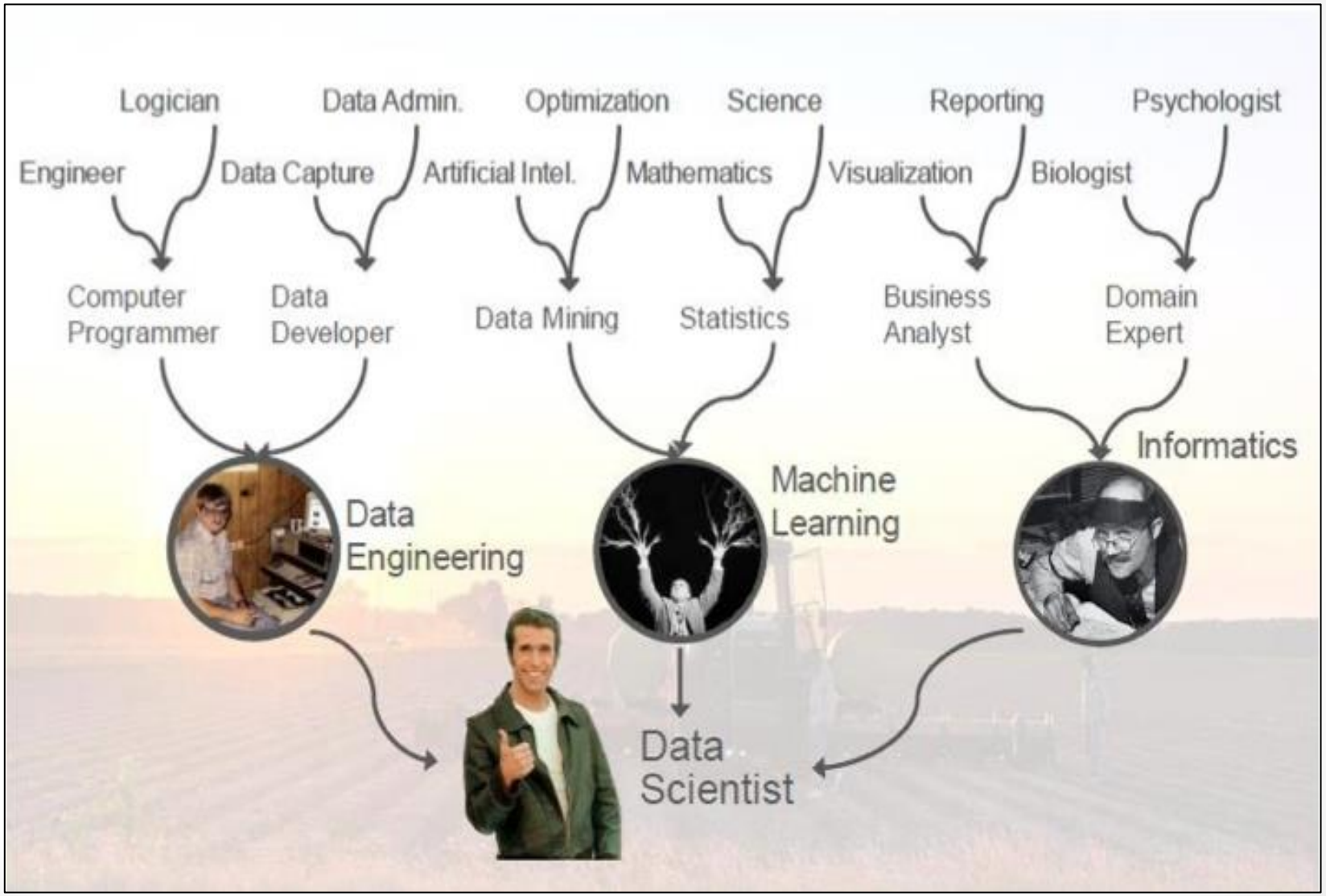
Data Scientist



From presentation by Jianhui Li at CODATA Training Workshop on Big Data for Science on 11 June 2014, Beijing, China

He used information from Lecture in CNIC by Yike GUO

How to be a data scientist



From presentation by Jianhui Li at CODATA Training Workshop on Big Data for Science on 11 June 2014, Beijing, China

He used information from Lecture in CNIC by Yike GUO

Cloud Computing for CAS data services

- Chinese Academy of Sciences Data Cloud (CAS Data Cloud) is focused on cloud technology to provide facilitated ways for scientists to make use of powerful information infrastructure, massive scientific data and rich scientific software
- It is a mixed evolution of grid computing, distributed computing, parallel computing, utility computing, network storage technologies, virtualization, and etc.

From presentation by Yuanchun Zhou at CODATA Workshop on 6 June 2014, Beijing, China

3 Cloud Service Models

- **Cloud Infrastructure as a Service (IaaS)**

The capability provided to the consumer is to **rent** processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has **control over operating systems, storage, deployed applications, and possibly select networking components** (e.g., firewalls, load balancers).

- **Cloud Platform as a Service (PaaS)**

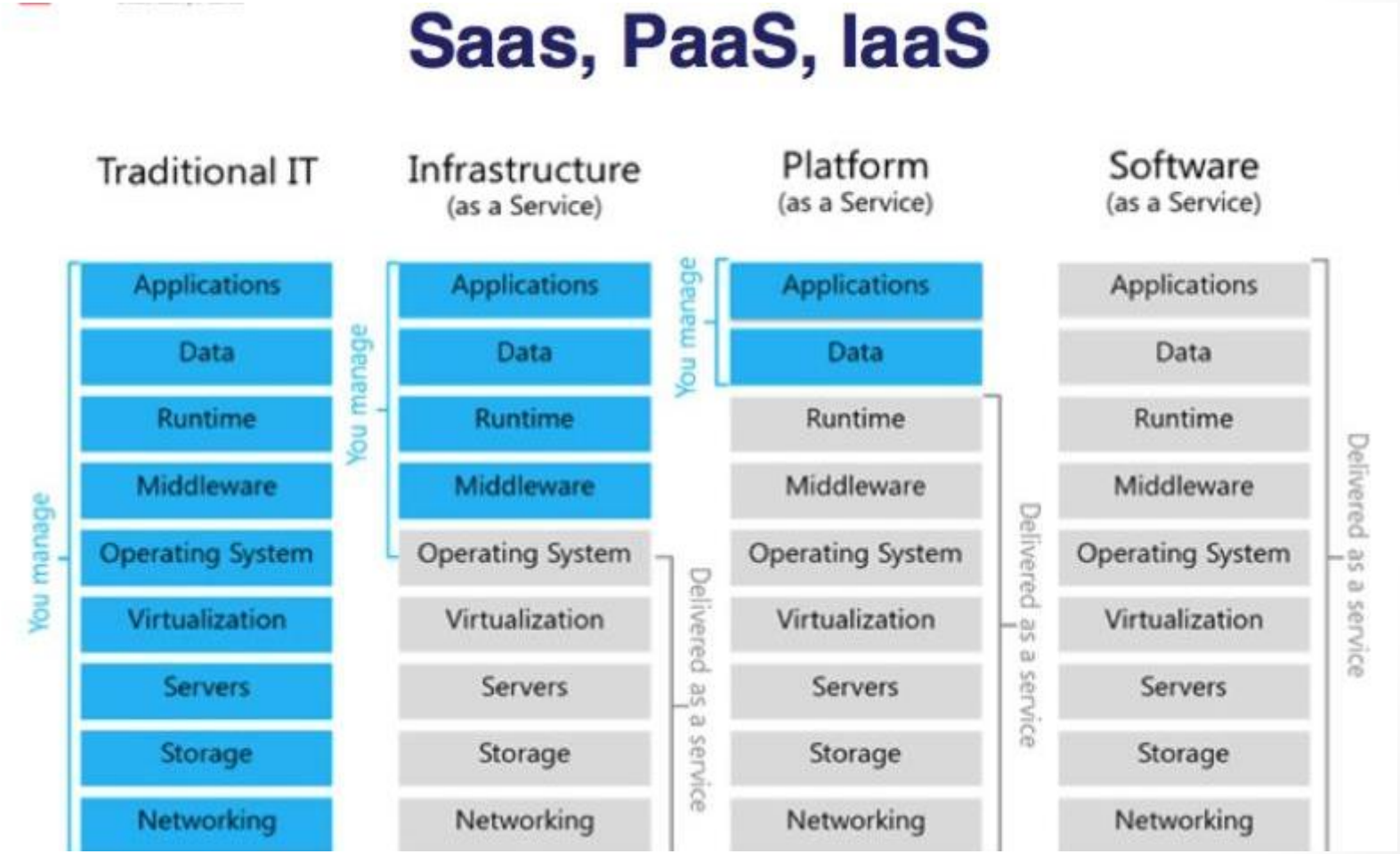
The capability provided to the consumer is to **deploy onto the cloud infrastructure** consumer- created applications using programming languages and tools supported by the provider (e.g., Java, Python, .Net). The consumer does not manage or control the underlying cloud infrastructure, network, servers, operating systems, or storage, but the **consumer has control over the deployed applications and possibly application hosting environment configurations.**

- **Cloud Software as a Service (SaaS)**

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure and accessible from various client devices through a thin client interface such as a Web browser (e.g., web-based email)**. The consumer does not manage or control the underlying cloud infrastructure, network, servers, operating systems, storage, or even individual application capabilities**, with the possible exception of limited user-specific application configuration settings.

From presentation by Yuanchun Zhou at CODATA Workshop on 6 June 2014, Beijing, China

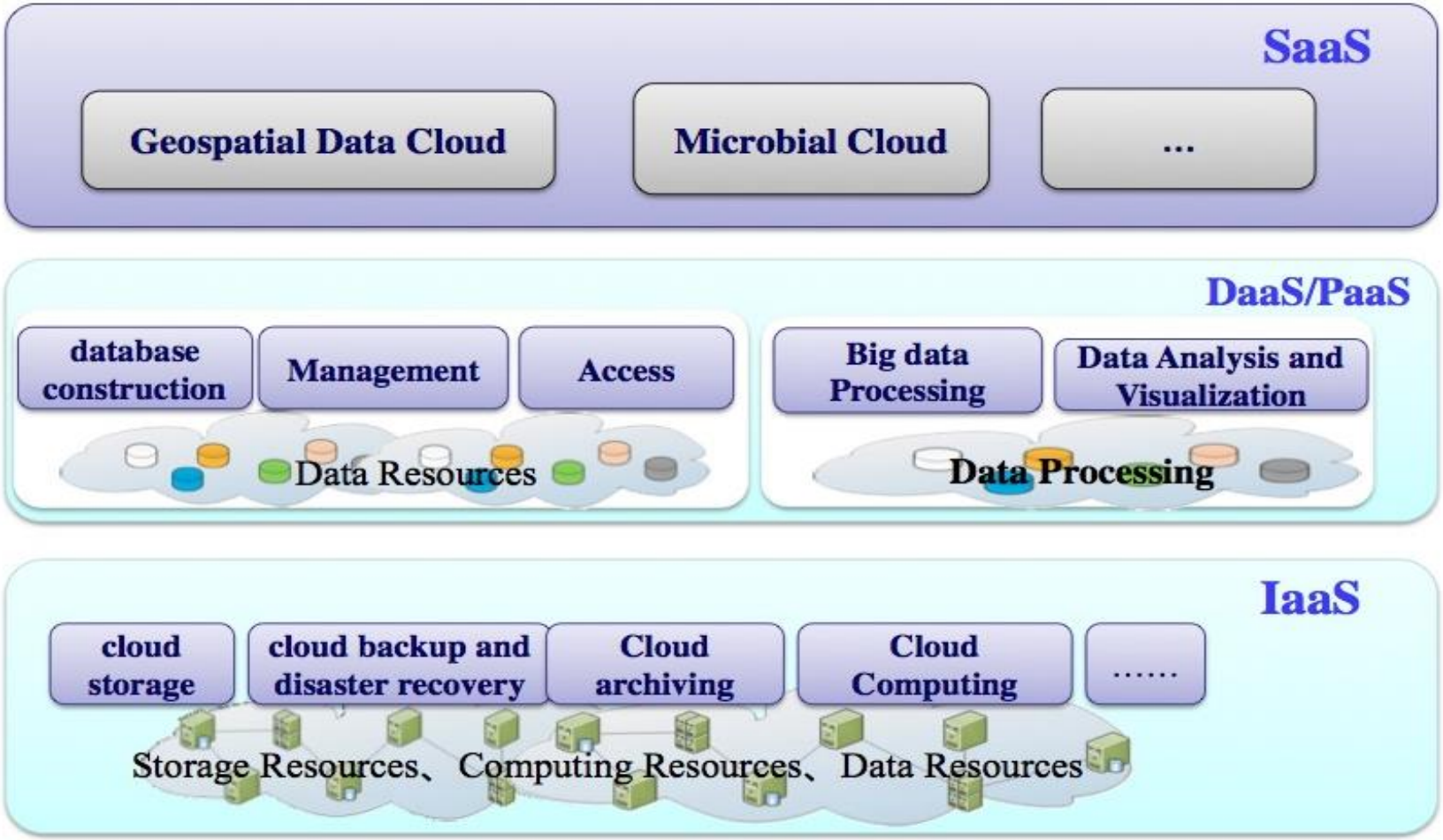
3 Cloud Service Models



From presentation by Yuanchun Zhou at CODATA Workshop 6 June 2014, Beijing, China

中国科学院
计算机网络信息中心
Computer Network Information Center, Chinese Academy of Sciences

CAS Data Cloud--Architecture



From presentation by Yuanchun Zhou at CODATA Workshop on 6 June 2014, Beijing, China

Data Analysis

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

- maximize insight into a data set;
- uncover underlying structure;
- extract important variables;
- detect outliers and anomalies;
- test underlying assumptions;
- develop parsimonious models;
- and determine optimal factor settings.

EDA Techniques

- Most EDA techniques are graphical in nature with a few quantitative techniques.

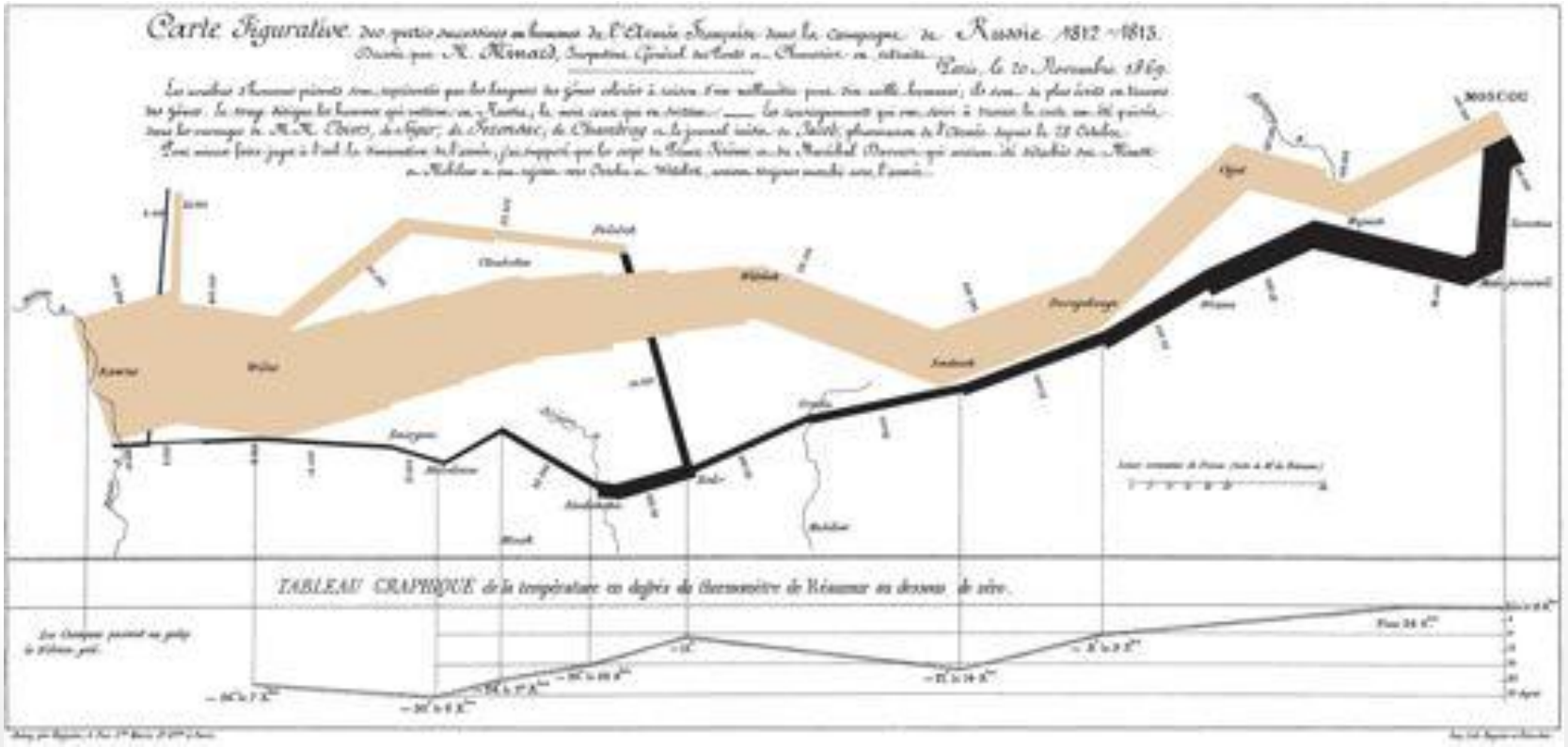
  The main role of EDA is to open-mindedly explore, and graphics gives the analysts unparalleled power to do so, enticing the data to reveal its structural secrets, and being always ready to gain some new, often unsuspected, insight into the data.

- Graphics provide, unparalleled power to apply natural pattern-recognition capabilities. The particular graphical techniques employed in EDA are:
  - Plotting the raw data (such as data traces, histograms, bihistograms, probability plots, lag plots, block plots, and Youden plots.
  - Plotting simple statistics such as mean plots, standard deviation plots, box plots, and main effects plots of the raw data.
  - Positioning such plots so as to maximize our natural pattern-recognition abilities, such as using multiple plots per page.

Visualisation of Data

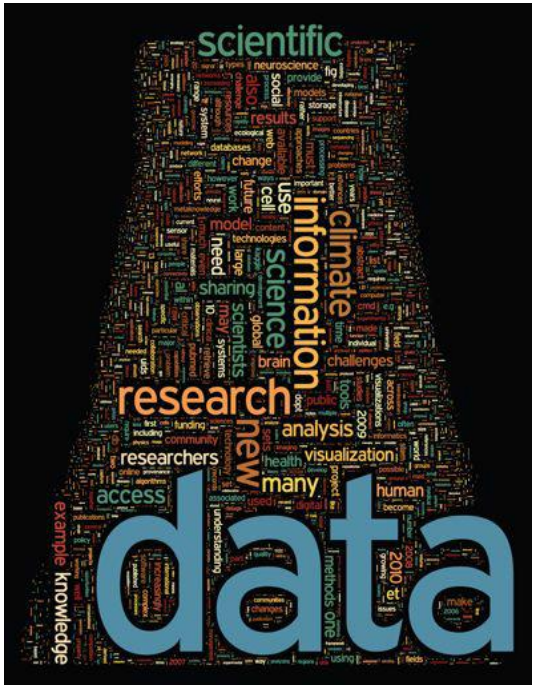
Not Something New

Napoleon's March to Moscow, Charles J. Minard, 1869

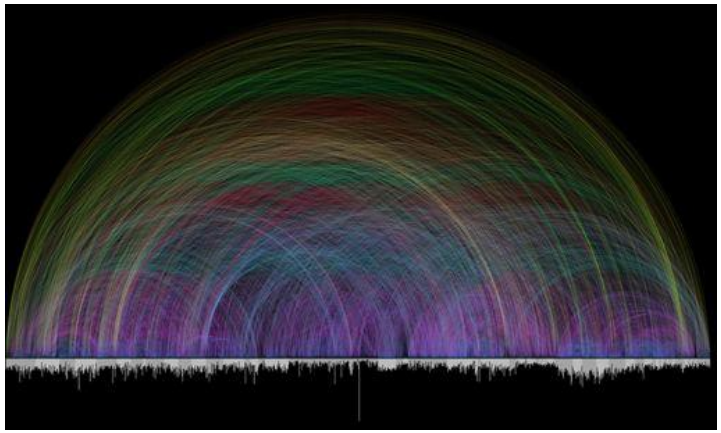


From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Visualisation of Data: Examples



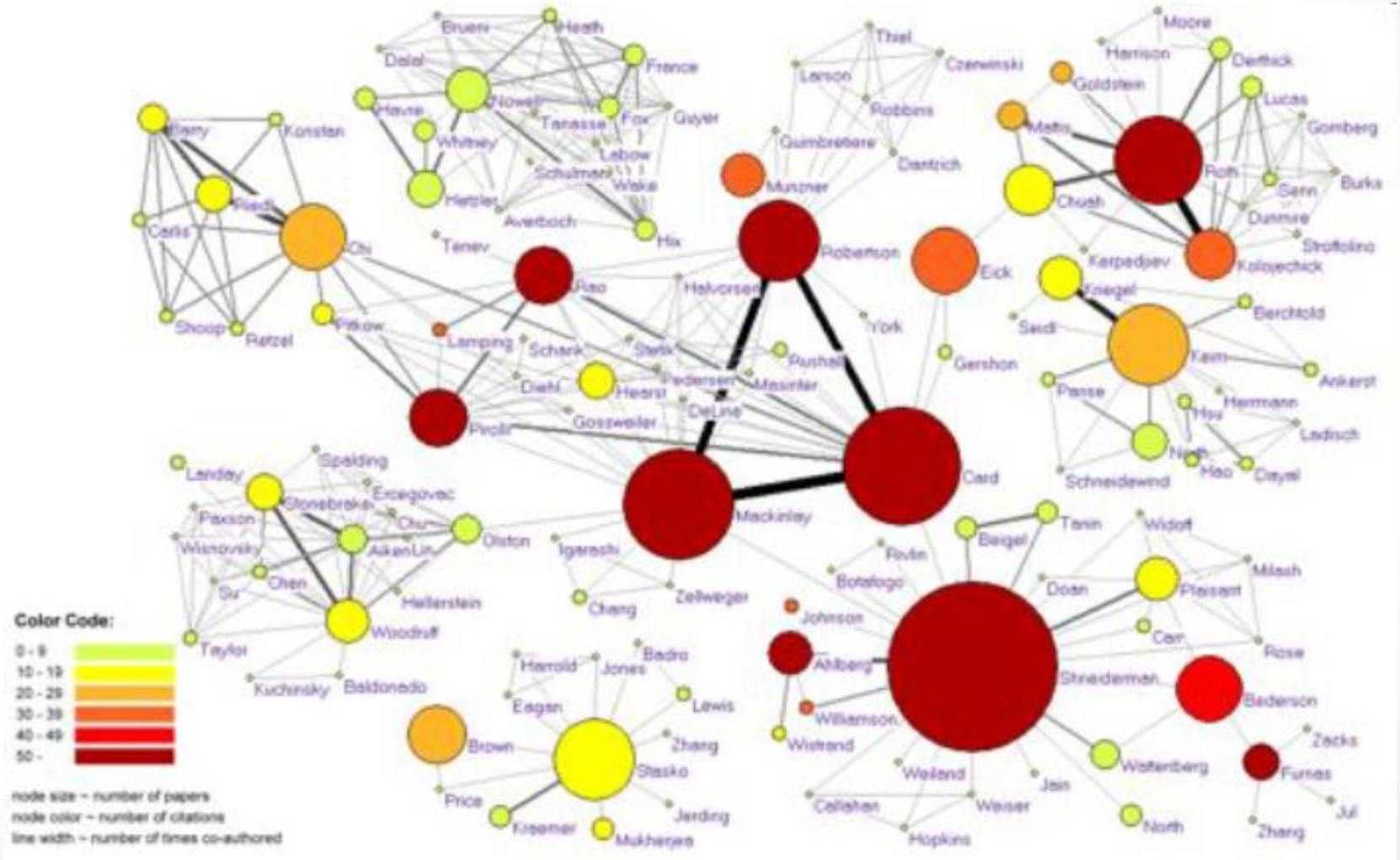
Cross-References in Bible



Kevin Hulsey Illustration, INC

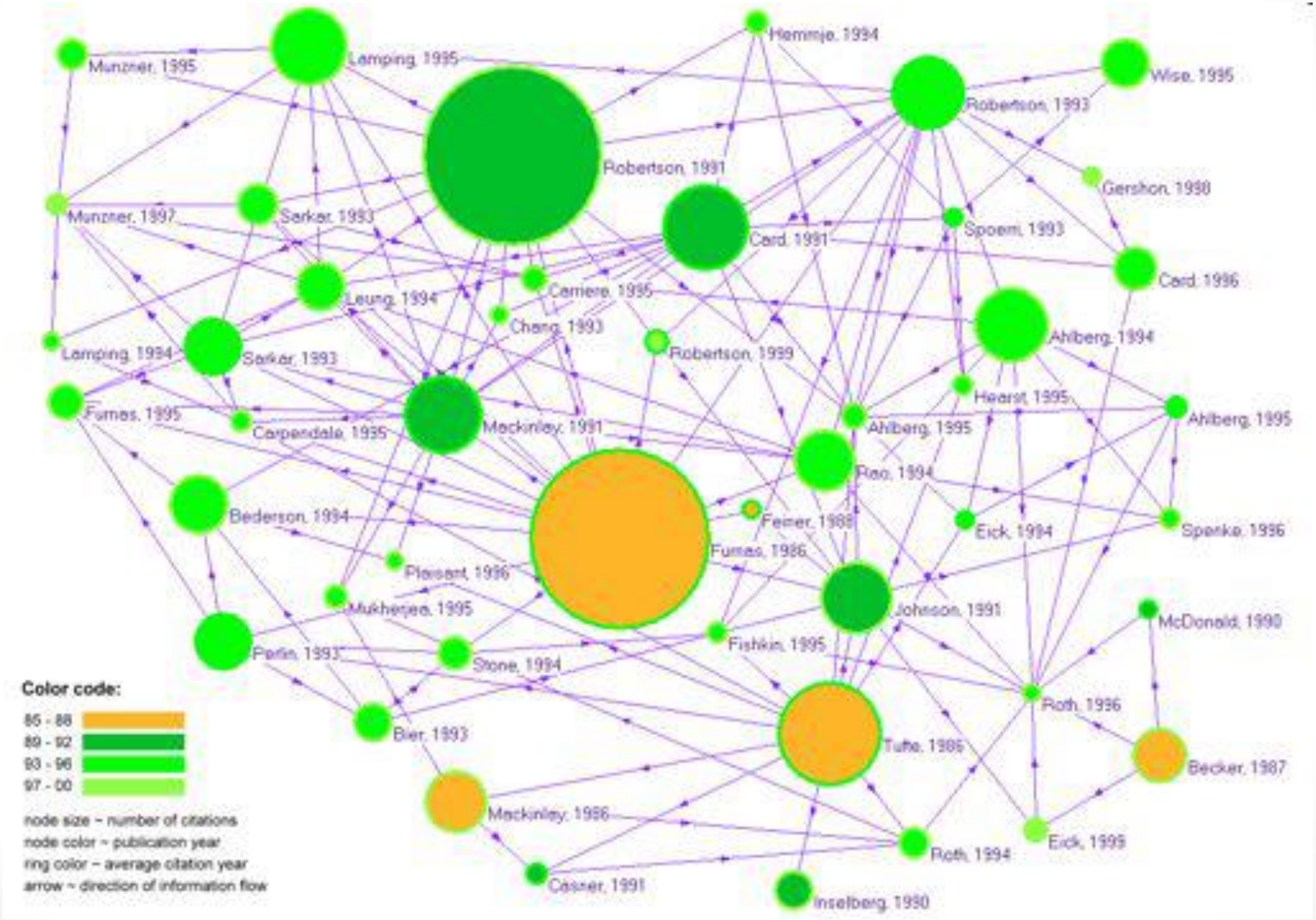
From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Visualisation of Data: Examples



From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

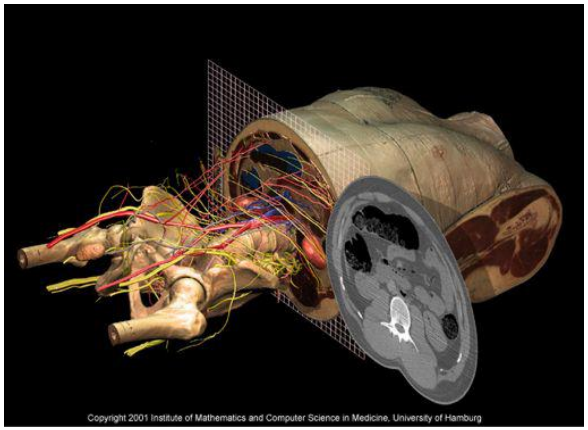
Visualisation of Data: Examples



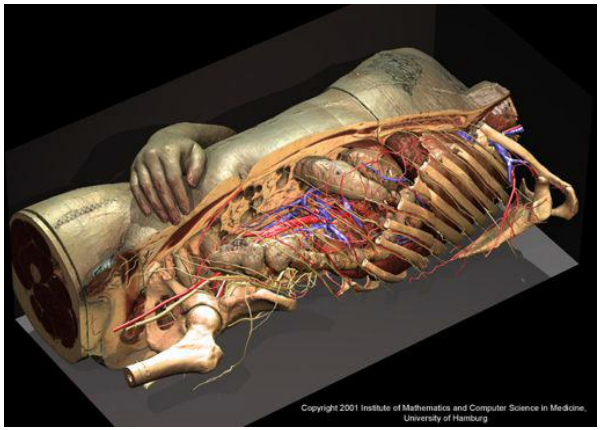
From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

More examples

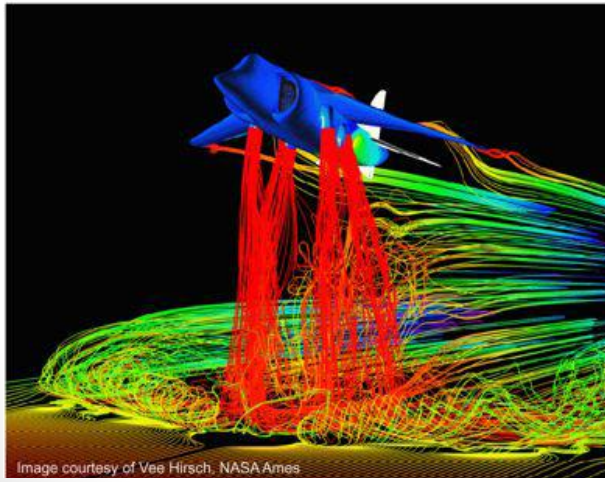
Health Sciences Field



Health Sciences Field



Wind flow



Scalable Multi-variate Analytics of Seismic and Satellite-based Observational Data

From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Challenges!

- More and more unseen data
- Faster Creation and Collection
- Fast Dissemination
- 5 exabytes of new information in 2002 [lyman 03] - 37000 Libraries of Congress
- 161 exabytes in 2006 [Gantz07]
- 988 exabytes in 2010
- Need better tools and algorithms to visually convey information

From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Why Visualisations?

Reasons for Visualisations	
• To expose ideas/relationships	• To answer questions
• To make an argument	• To make decisions
• To observe trends	• To see data in context
• To summarize/aggregate data	• To expand memory
• For archiving	• To support graphical calculation
• To create trust	• To find patterns
• To advertise ideas	• To present an argument
• For exploratory data analysis	• To tell a story
• To inspire	

From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Three Functions of Visualisations

- **Record information**
  - Photographs
  - Blueprints, etc
- **Support reasoning about information (analyse)**
  - Processing and calculations
  - Reasoning about data
  - Feedback and interaction
- **Convey information to others (present)**
  - Share and persuade
  - Collaborate and revise
  - Emphasize important aspects of data

From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Peking University Visualisation Lab

From Xiaoru Yuan's presentation at CODATA Workshop on 12 June 2014, Beijing, China

Data Publishing: Re-using World data sources

- Have a clear idea of what you want to do

  → creative idea

- Look for existing datasets (world data infrastructure – GEOSS, WDS, GCMD …)

  → access the data

- Evaluate how this data is suitable for re-using data

  → processing for integrating

- Add your local knowledge and integrate it with the data

  → algorithm

- Get your data product analysed

  → discovery

- Publish your research products

  → publish paper and data

From LIU Chuang's presentation at CODATA Workshop on 16 June 2014, Beijing, China

Data Publishing

Why should data be published?

- Big data not only means huge of volumes of data, but large of numbers of scientists using and creating data.
- Every scientist in their research uses data, also creates new data.
- Most of the datasets are at “sleep” in individual scientists’ offices
- Scientists’ IP is recognized and rewarded – data publishing is the best way to get data author(s)’ contribution to science recognized and rewarded.

From LIU Chuang’s presentation at CODATA Workshop on 16 June 2014, Beijing, China

Data Publishing- A New Research Data Sharing Model

- Make data persistently available on the Web:
  - maybe with a procedure of quality control;
  - So that they can be accessed, downloaded, analysed and reused by anyone for research or other purposes
  - Including
    - Academic Publication
    - Commercial Publication

Data Publishing: A Different Publication Model

- **Resource Sharing Model**
  - Data collected by public funding and should be shared by public
  - Data Centres and Libraries (Public Repository)
  - Big Science, and especially observation data held by Big Science Facilities (Astronomy, High Energy Physics) or Organisations (USGS/NOAA etc.)
- **Evidence/Supplement Publishing Model**
  - Data used by scientific paper should be opened as evidence reviewed by peer reviewers, accessed by readers (even reanalysis)
  - Data as supplements to the scientific paper in Journal should be submitted to the public repository (Genebank, etc), before the paper is published
- **Result Publishing Model (Data paper)**
  - Data paper + data repository
  - Rewards to data author

From presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

Data Publishing: Fundamental requirements for a Research Data Publication

- **Persistent Unified Identifier**
- **Persistent Accessible**
  - Archived in one, or preferably, more than one independent document, or data repository, so that it will be available to be accessed persistently
- **Understandable**
  - Necessary Metadata that are Machine Readable, and Understandable
  - Rich Metadata/description that are human readable and understandable
- **Reusable**
  - Quality, fit for use
  - Easy for reuse (format, approach)
- **Citable**

From presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

Data Journals

- Ecological Archives
  - From 2000 to now, published 88 data papers - data paper's average citation number is almost 11
- Earth System Science Data
  - From 2009 to now, published 89 data papers – data paper's average citation number is almost 12
- Biodiversity Data Journal
  - From 2011 to now, published 11 data papers
- Nature Scientific Data
  - Launched last month

From presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

Registry of Data Repositories

Popular Data Registries:- Databib and re3data.org

- Databib connects to 978 data repositories and databases (agriculture,Geo-sciences,social Sciences,Biological sciences)
- re3data.org currently lists 634 research data repositories from different disciplines and 586 of these are described in detail using the re3data.org schema.
- In future, Databib and re3data.org will be merged into one service.

From ARD Prasad's presentation at CODATA Workshop on 9 June 2014, Beijing, China

Public Research Data Repository

- **Dryad**
- **PANGAEN**
- **Dataverse Network**
- **FigShare**

From presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

	Dryad	PANGAEN	DataVerse
Domain	Any, now more in ecology and life science	Earth and Life Sciences	All disciplines worldwide
Founders	Non-profit membership organisation	AWI and MARUM in Germany hosted, supported by several projects	Harvard University, Institute of Quantitative Social Science
Data type/Format	Any type	Any type	All file formats with maximum size of 2GB per file
Quality Control	Journal Peer-reviewed	Data Editorial ensures integrity and authenticity	By data owners
Data Providers	Research article authors	Authors for journal ESSD and various scientific journals related to earth system research	Any researcher worldwide (faculty, postdoc, student or staff)
Business Model	Member fees and DPCs	Free of charge, but appreciate financial contribution of average 300 €	Free to deposit data
Data Identifier	DOI	DOI	DOI
Data Package Download	Yes	Yes	Yes
API	A few	None	Data sharing API/Data deposit API
Citable	Yes, to data and paper	Yes, to data directly	Yes, to data directly

From presentation by Jianhui LI at CODATA Workshop on 11 June 2014, Beijing, China

Challenge: how to evaluate scientists' contributions with regards to the data

- Who is the data author? Not very clear in some cases in China
- Is the data reliable? Not very clear in some datasets in China
- How to evaluate scientists' contribution with regards to their data? No standard and no practices even;
- ….;

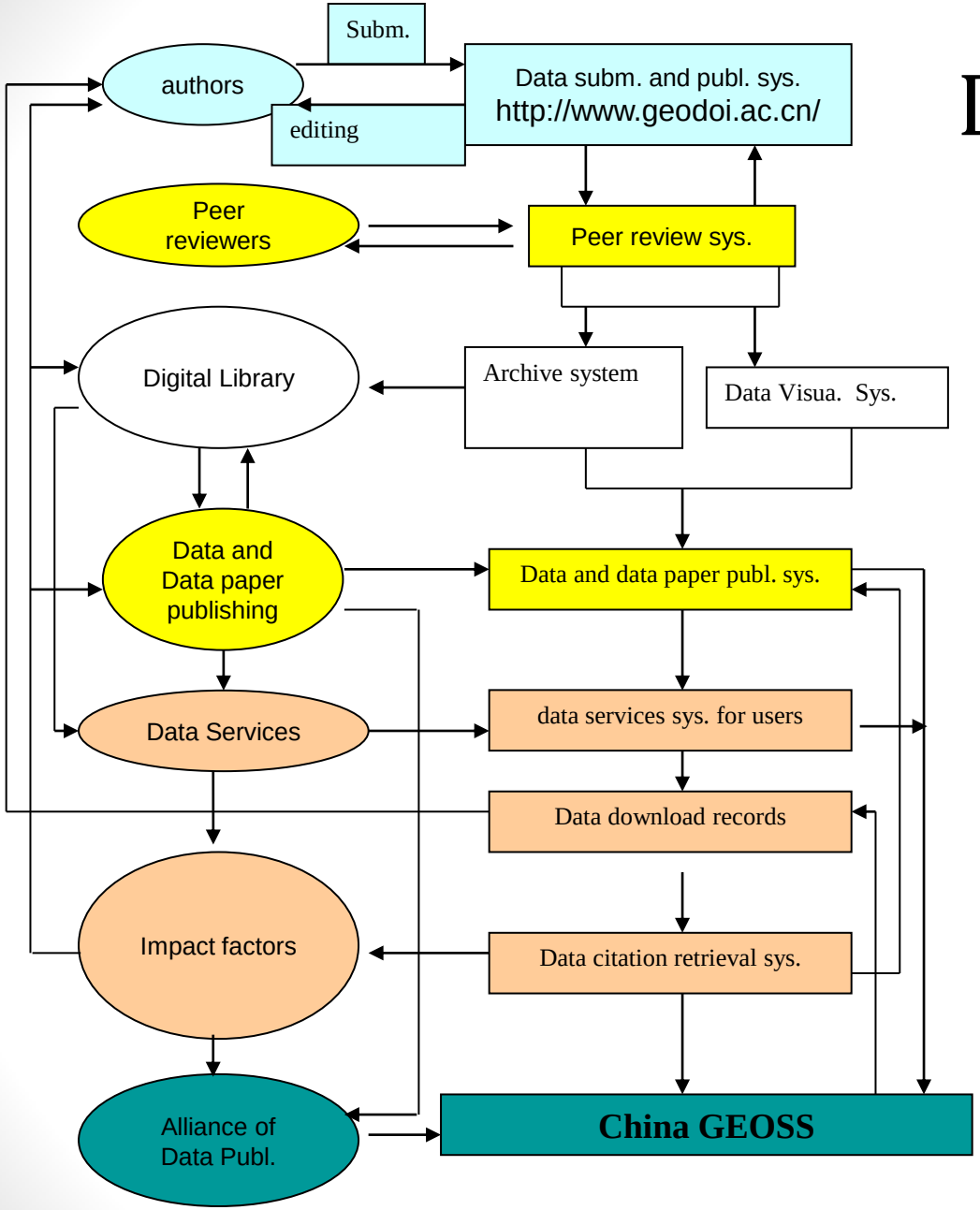
From LIU Chuang's presentation at CODATA Workshop on 16 June 2014, Beijing, China

DOI Technology and Standard

- DOI: ISO 26324 (May 2012)
- DOI: DOI/Handle system and technology worldwide available
- This made data publishing possible and is the key solution for data sharing in big data science

From LIU Chuang's presentation at CODATA Workshop on 16 June 2014, Beijing, China

Data Publishing



DOI:10.3974

Procedure and Flowchart of Global Change Research Data Publishing and Repository

From LIU Chuang's presentation at CODATA Workshop on 16 June 2014, Beijing, China

Example of Repository for Data Publishing

www.geodoi.ac.cn

Things learned/of value/of interest

A clearer understanding of CODATA, all its activities, Workgroups, and Task groups

Data Policy

- International and national aspects of data policy
- Data policy committee: setting an international agenda for data policy, expert forum, advice and consultancy
- Coordinating with national committees

Data Science

- Long-standing activities: fundamental constants.
- Strategic working groups; community-driven task groups
- Disciplinary and interdisciplinary data challenges, Big Data

Capacity Building

- Longstanding work on data preservation and access with developing countries
- Executive Committee Task Force on Capacity Building: setting an international agenda for capacity building; Early Career WG

Data for International Science

- Support for ICSU Mission
- Data issues and challenges in international, interdisciplinary science programmes

From presentation by Simon Hodson at CODATA Workshop on 6 June 2014, in Beijing, China

Task groups in CODATA

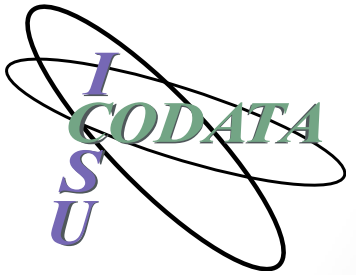
- Advancing Informatics for Microbiology
- Anthropometric Data and Engineering
- Data at Risk
- Data Citation Standards and Practices
- Earth and Space Science Data Interoperability
- Exchangeable Materials Data Representation to Support Scientific Research and Education
- Fundamental Physical Constants
- Global Information Commons for Science Initiative
- Linked Open Data for Global Disaster Risk Research
- Octopus: Mining Space and Terrestrial Data for Improved Weather, Climate and Agriculture Predictions
- Global Roads Data Development
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

From presentation by Simon Hodson at CODATA Workshop on 6 June 2014, in Beijing, China

Invitation

Apply for corresponding membership to CODATA Task group on Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

http://www.codata.org/task-groups/preservation-of-and-access-to-scientific-and-technical-data-in-for-with-developing-countries-pastd



Invitation

Join CODATA Early Career Data Professionals Group (Initiative of CODATA EC Task Force on Capacity Building)

- This initiative is in line with two of the aims of NICIS (National Integrated Cyberinfrastructure System) of South Africa

  "develop training for the human expertise necessary for data scientists, including data management, sharing and analysis"; and

  "actively participate in international forums related to Data Services to promote South African activities and gain knowledge from other international efforts"

- This initiative is in line with one of the aims of NeDICC (Network of Data and Information Curation Centres)

  "to train and develop skills in research data management"

Events/Opportunities coming up

International Workshop on Open Data for Science and Sustainability in Developing Countries, Nairobi, Kenya, 4-8 August 2014: http://www.codata-pastd.org

Events/Opportunities coming up

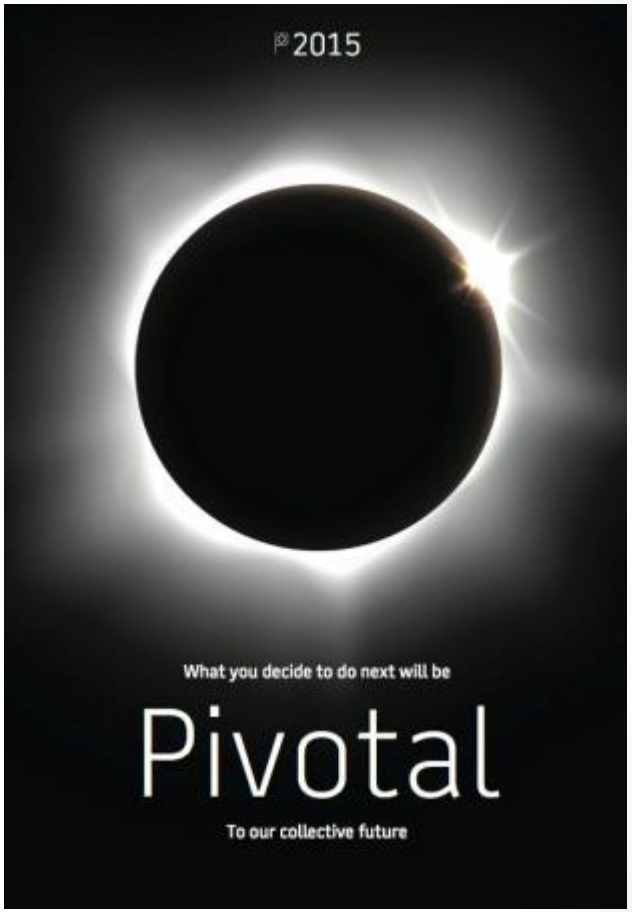
SciDataCon 2014, New Delhi, 2-5 November 2014: http://www.scidatacon2014.org

Events/Opportunities coming up

- **Pivotal – The Spatial Edge Master Class for Resilient and Rapid Response, Brisbane, Australia: 29 June – July 2015**

  Hands-on Workshop on Visualisation grounded on real-world solutions

  http://pivotal2015.org

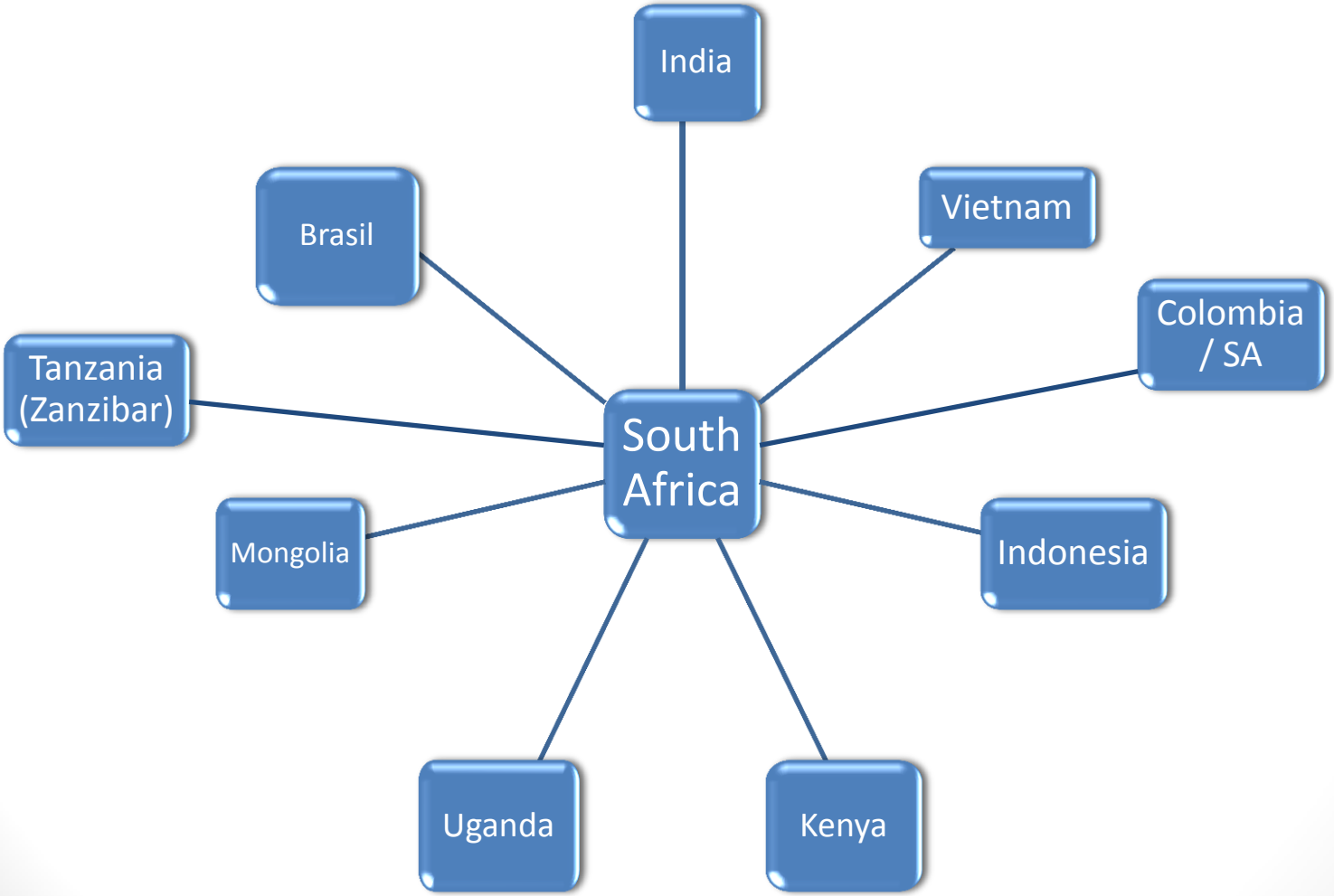


Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Mon	Tues	Wed	Thurs	Fri
Local & World View	Foundation	Build & Design	Nouveau Spatial	Operations For Humanity	Design Retreat		Creative Workshop & Capacity Building				

NASA ISDE SIBA QUT

Actions

Network with participants in the CODATA Workshop

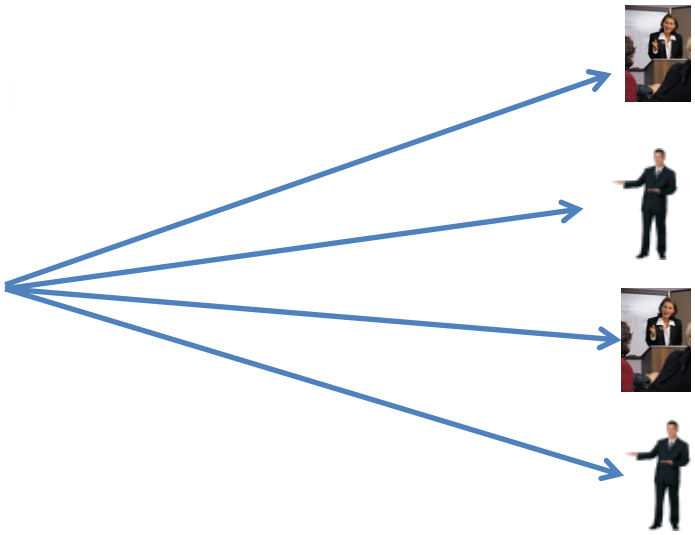


Actions

Act as bridge between the various researchers in different disciplines at University of Pretoria and the presenters at the CODATA Training Workshop

- create awareness of initiatives
- link-up researchers with presenters

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



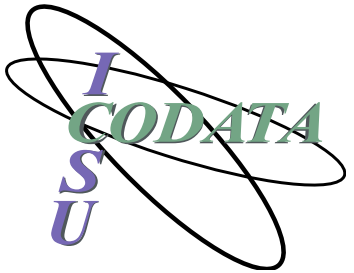
ICSU CODATA

Actions

Create awareness among networks in South Africa about the activities, workgroups and taskgroups of CODATA

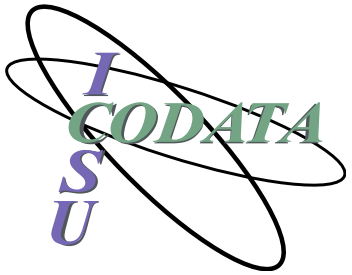
NICIS

DIRISA



Actions

Work in closer collaboration with the National Research Foundation (the national member of CODATA) with regards to Data Management Initiatives



Actions

Include knowledge gained in the formulation of the new UP Data Management Policy

Bibliography

- CHANG, WO. 2014. ***NIST Big Data Public Working Group and standardization activities.*** Presentation on 8 June 2014 at the CODATA Workshop on Big Data for International Scientific Programmes: challenges and opportunities, held in Beijing, China, 8-9 June 2014.
- CHENZHOU, CUI. 2014. ***Virtual Observatory, a global platform for astronomical data.*** Presentation on 18 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.
- CHUANG, LIU. 2014. ***Data re-use and publishing for sciences and sustainability in developing countries.*** Presentation on 16 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.
- DANHUAI, GUO. 2014. ***Exploratory analysis and visualization of spatio-temporal data.*** Presentation on 13 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.
- ***The Fourth Paradigm: data-intensive scientific discovery.*** Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Redmond, WA : Microsoft Research, c2009.

Bibliography (2)

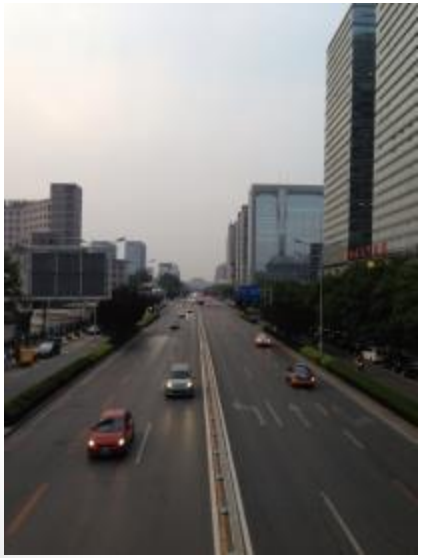
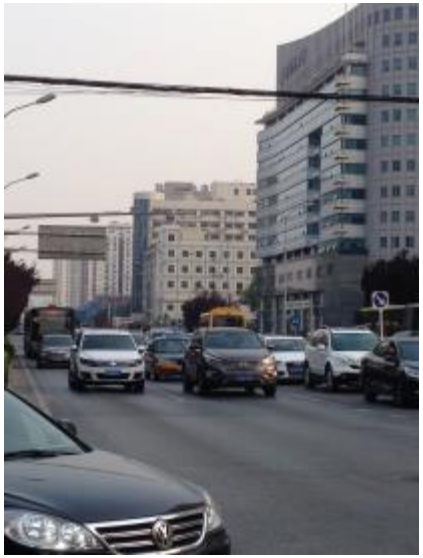
- FOX, PETER. 2014. ***Environmental Informatics: leveraging big data to solve big problems.*** Presentation on 8 June 2014 at the CODATA Workshop on Big Data for International Scientific Programmes: challenges and opportunities, held in Beijing, China, 8-9 June 2014.
- HODSON, SIMON. 2014. ***Global collaboration in data science: an introduction to CODATA.*** Presentation on 6 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.
- HUADONG, GUO. 2014. ***Big data, big science, towards big discovery.*** Presentation on 8 June 2014 at the CODATA Workshop on Big Data for International Scientific Programmes: challenges and opportunities, held in Beijing, China, 8-9 June 2014.
- JIANHUI, LI. 2014. ***Understanding big data and data science: a dialogue with participants of the training workshop.*** Presentation on 11 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.
- NORMANDEAU, KEVIN. 2013. *Beyond volume, variety and velocity is the issue of big data veracity.* ***Inside BIGDATA***, 12 September 2013. [Online] available at http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/ (Accessed 15 July 2014).

Bibliography (3)

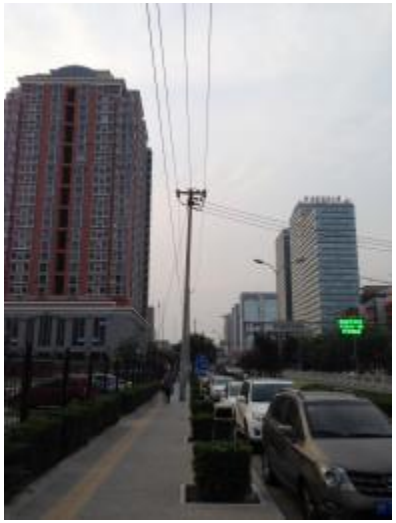
- PRASAD, ARD. 2014. ***Metadata in big data.*** Presentation on 9 June 2014 at the CODATA Workshop on Big Data for International Scientific Programmes: challenges and opportunities, held in Beijing, China, 8-9 June 2014.
- SIEGFRIED, TOM. 2013. *Why big data is bad for science*. ***Science News***, 26 November 2013. [Online] available at https://www.sciencenews.org/blog/context/why-big-data-bad-science (Accessed 15 July 2014).
- WANG, LIZHE. 2014. ***Data-intensive scientific discovery in Digital Earth.*** Presentation on 9 June 2014 at the CODATA Workshop on Big Data for International Scientific Programmes: challenges and opportunities, held in Beijing, China, 8-9 June 2014.
- WHITE, MICHAEL. 2013. *How big data is changing science and [society]*. ***Pacific Standard***, 8 November 2013. [Online] available at http://www.psmag.com/navigation/nature-and-technology/big-data-changing-science-society-69650/ (Accessed 15 July 2014).
- YUANCHUN, ZHOU. 2014. ***Cloud concept, overview and CAS Data Cloud.*** Presentation on 6 June 2014 at the CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, held in Beijing, China, 5-20 June 2014.

Sightseeing: Modern City



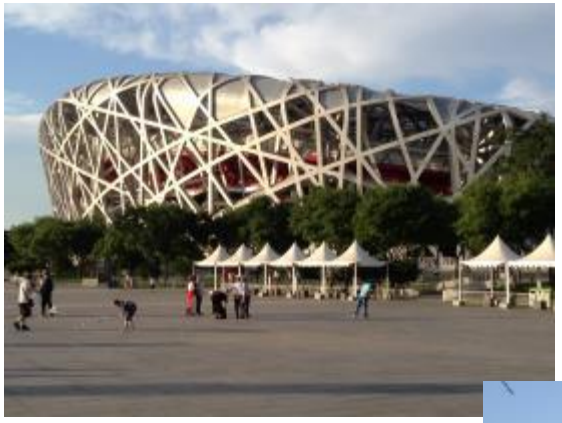




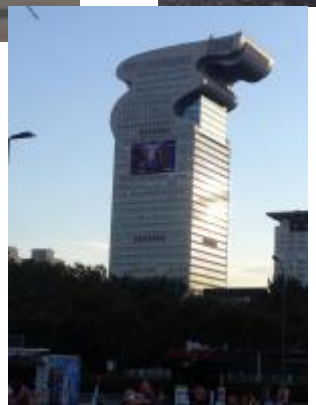

Sightseeing: Modern City



Sightseeing: Subway System

Sightseeing: Olympic Park

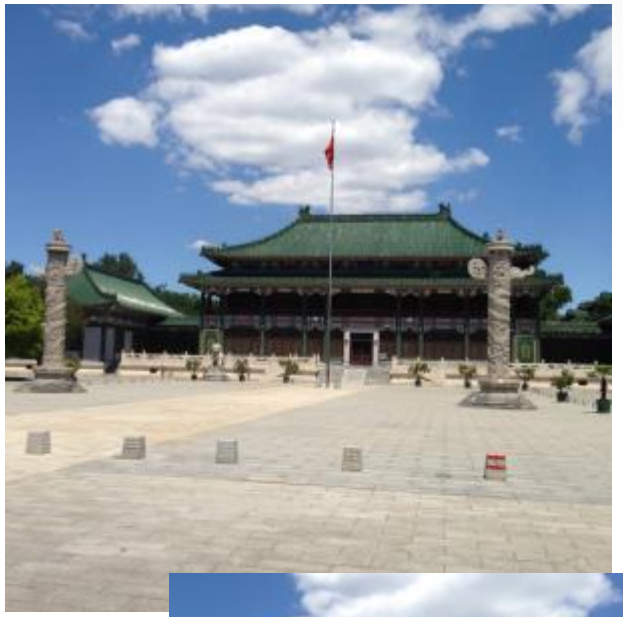


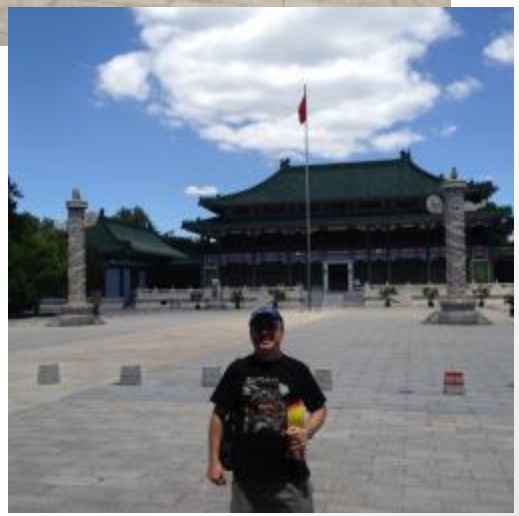


Sightseeing: Forbidden City

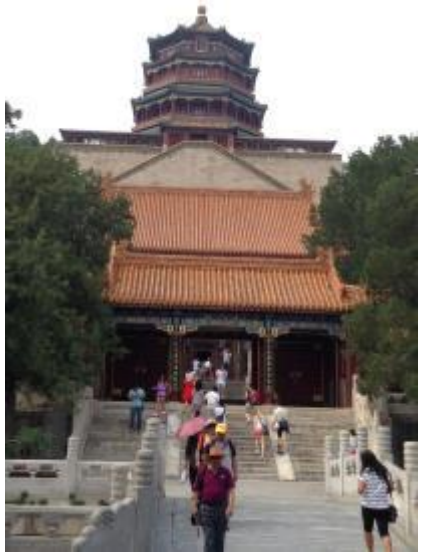
Sightseeing: Forbidden City

Sightseeing: National Library of China





Sightseeing: Summer Palace



Sight Seeing: Great Wall of China









Sightseeing: Shopping

Jade Factory

Silk Market

Ming Tombs & Beijing Art District

Ming Tombs

Art District

Beijing Meals

谢谢 - Xièxiè! – Thank You