Information Management and Ecosystem Research

金恒鑣　　Hen-biau KING

Jane Goodall Institute of Taiwan

Wenzhou, ZheJiang, China

October 21, 2013

The VISION

Long-term science helps understand ecosystems and their services

Data is the foundation to better understand ecosystems and their services

Two Focuses

1. Data Management

Data are one of the most valuable products of the Forest Dynamics Plot program

2. Ecosystem Research

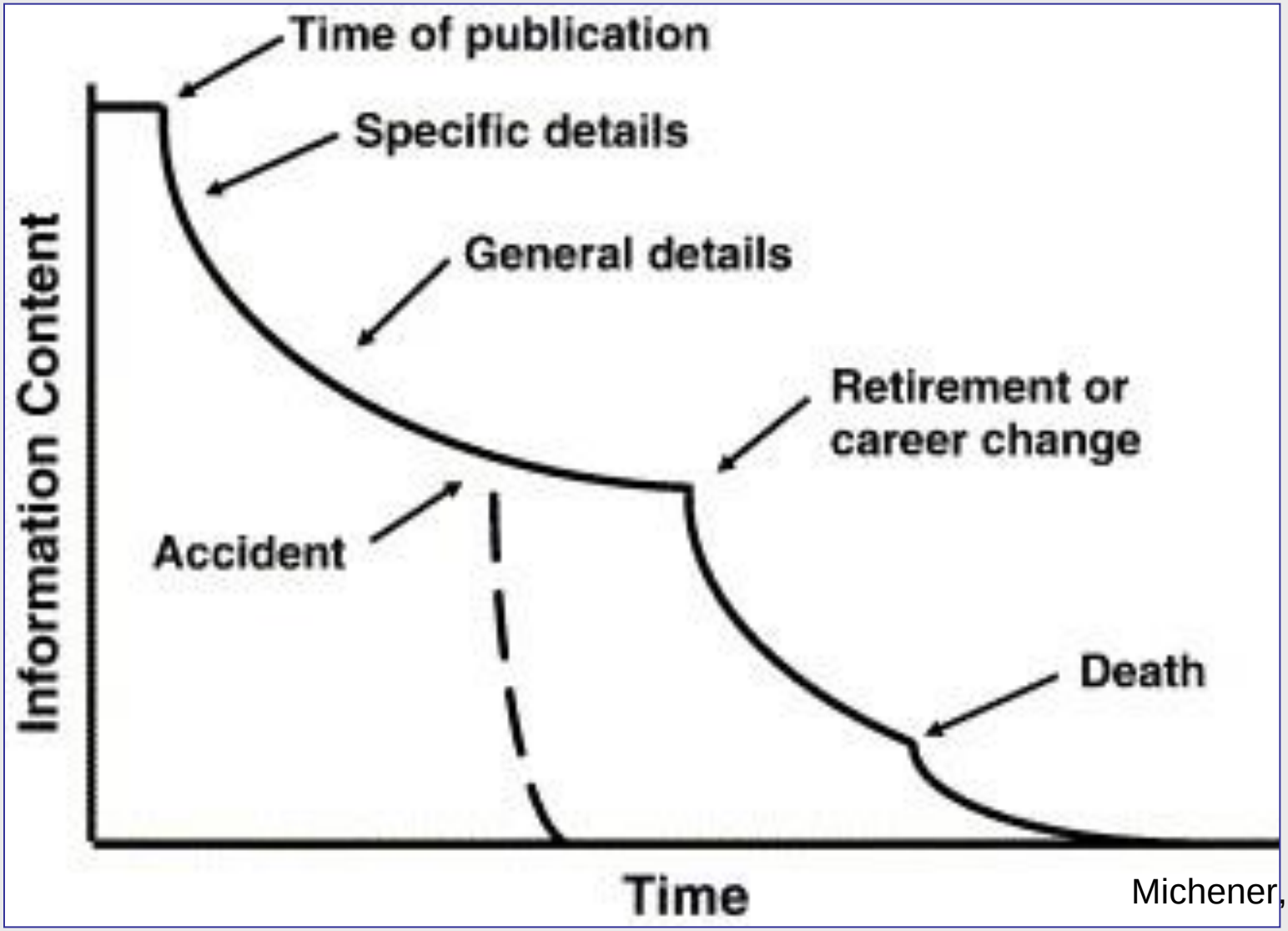
Ecologists should research ecosystem services for sciences, policies and societies

Data Management

We ceremoniously dump our data
upon retirement as there are no metadata.
The data are useless once their owners are gone.

http://besurbanlexicon.blogspot.tw/2012/07/metadata.html

Data Entropy



Michener, 1997

Why Data Matters?

Past information is used to help forecast the future

- The long-term value and the utility of data depend on the availability of suitable and adequate **metadata**.
- Structured metadata can maximizie the value of **Data**
- Large datasets are indispensible for scientists to discover **patterns and relationships** that are unknown at present.

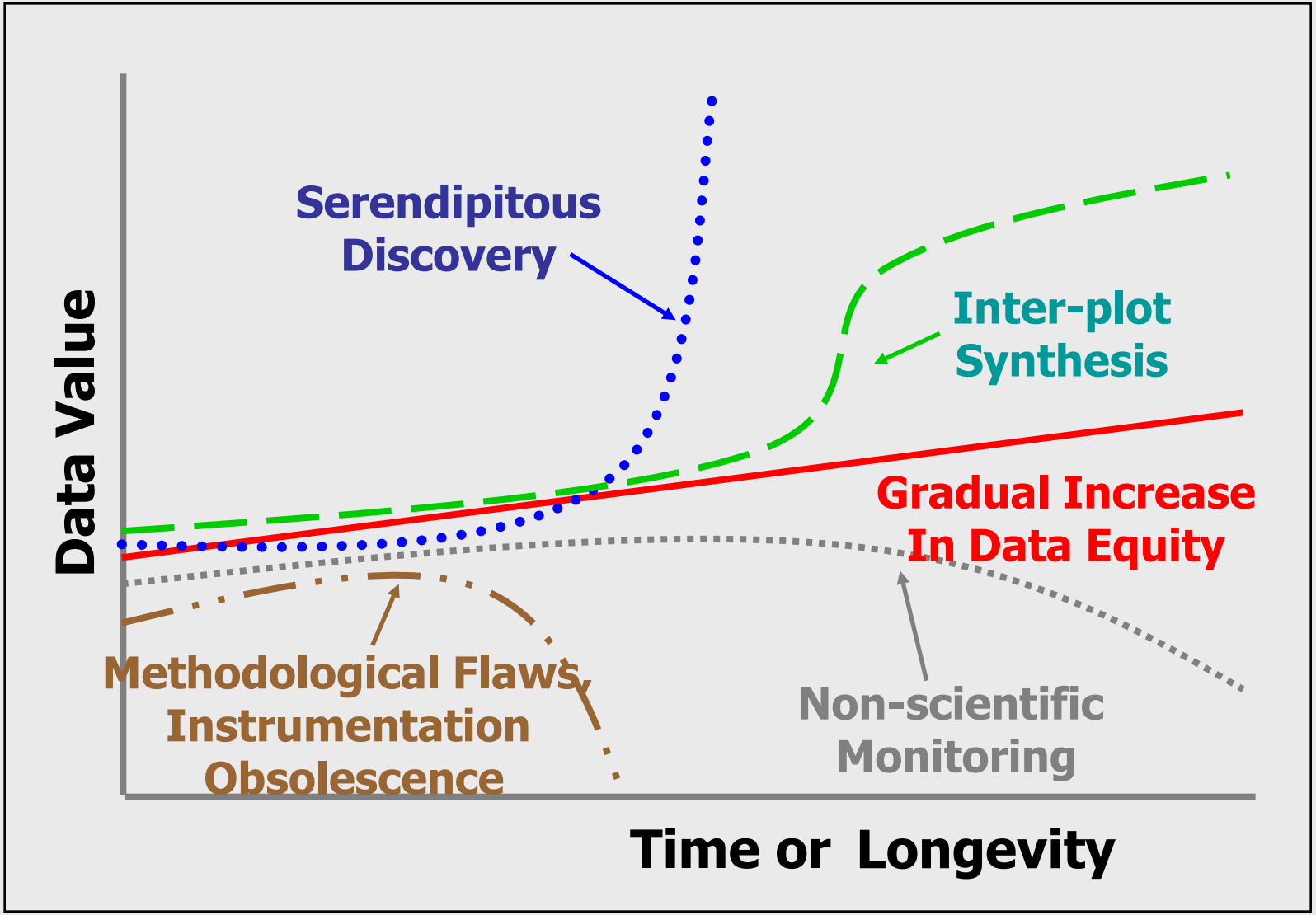
Michener et al. 1997

Why should data be managed?

- Well managed data can delay data entropy and increase datasets longevity. Data become Legacy.
- Data may be used to expand the scale of ecological inquiry and support valid comparisons in space and time
- Shared data can be reused to generated new knowledge.

Michener et al. 1997

Changes of Data value over time



Slide from James Brunt

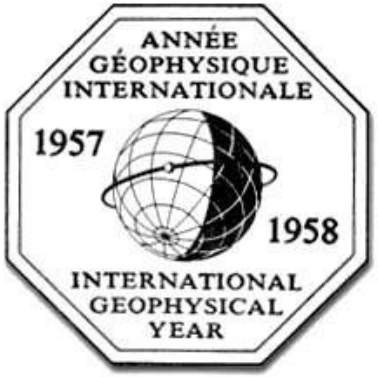
Managing and then Sharing

- We know the advantages of sharing data, but it still remains an issue. Why?
- Whose data? Why mine? What rewards?
- Data sharing guidelines & policies?
- Standards for metadata?
- Tools to support data sharing?
- Data repositories (and cyberinfrastructures)?

Milestones of Data Management

Collaborative and international research efforts had already been made.

Data-driven Scientific Initiatives



- The IPY (1932-1933)
- The IGY (1957-1958)
- The IBP (1964-1974)
- The LTER (1980+)

Collection of data on regional or global scale

Milestones of US-LTER IM Development (1/2)

1980	Initial LTER sites were funded (BCI Plot was established).
1989	A Data Manager was hired & email system was developed The personnel database was incorporated.
1990	Core Data Set Catalog was developed and published. GPS/RS Laboratory was established at UW *LTER Databits* was developed
1991	A Communications System Analyst was hired.
1993	Site-specific data-sharing policy was in place An Internet (gopher) server was established. The LTER All-Site Bibliography, Core Data Set Catalog & Personnel Database were put online in searchable form.
1994	LTER Network World Wide Web site was established.

Milestone of US-LTER IM Development (2/2)

1999	Knowledge and Distributed Intelligence Grant was Funded
2001	the LTER community accept the sharing policy
2005	the LTER Network also accept Data Access Policy, Data Access Requirements, and General Data Use Agreement
2006	The 1st Information Management Committee was formed
2007	A decadal plan was released to promote eco-social science
2008	The 1st Ecological IM Conference was held
2010	Became a member node of the dataONE (Data Observation Network for Earth)
2011	LTER's Network News was published its first online version

Three Decades of IM Development in US LTER

1980-1990

Decade of long-term
Decade of time-series data

Role:
Data Management

1990-2000

Decade of large-scale
Decade of data sharing

Role:
Information Management

2000-2010

Decade of synthesis
Decade of data integration

Role:
Informatics

Courtesy Chau-chin LIN

The Development of TERN IM

- 1995/03 Participated in IM Workshop in UW
- 2004/08 Built IM Capacity
- 2006/12 Set up IM System
- 2007/02 Established EAP-ILTER IM Committee
- 2008/04 Hosted EML Workshop for EAP-ILTER Regional Network
- 2010/05 Joined DataOne Initiative
- 2012/09 Developed Ontology-driven IMS

EAP-ILTER IM: Regional Efforts

2005: CERN Organized the 1st International Workshop on Ecological IM, July in Beijing

2006: **TERN** hosted the 2nd International Workshop on Ecological IM, February in Taipei

2006: **TERN** helped to form the EAP-ILTER Regional IM Committee

2007: KLTERN organized the 3rd International Workshop on Ecological IM, October in Seoul, **(KLTERN)**

Courtesy CC LIN

Beijing, China 2005


Taipei, Taiwan 2006

Seoul, Korea 2007

Kunchung, Taiwan 2007

Data-intensive Researches

- Twenty-first century research is more **data-intensive** than ever due to the proliferation of digital technologies and the demand for answers in this era of fast-paced innovation.
- **Data sharing** has increasingly become a global imperative due to the changing nature of research, available data and tools.

Science must move from data to information to knowledge

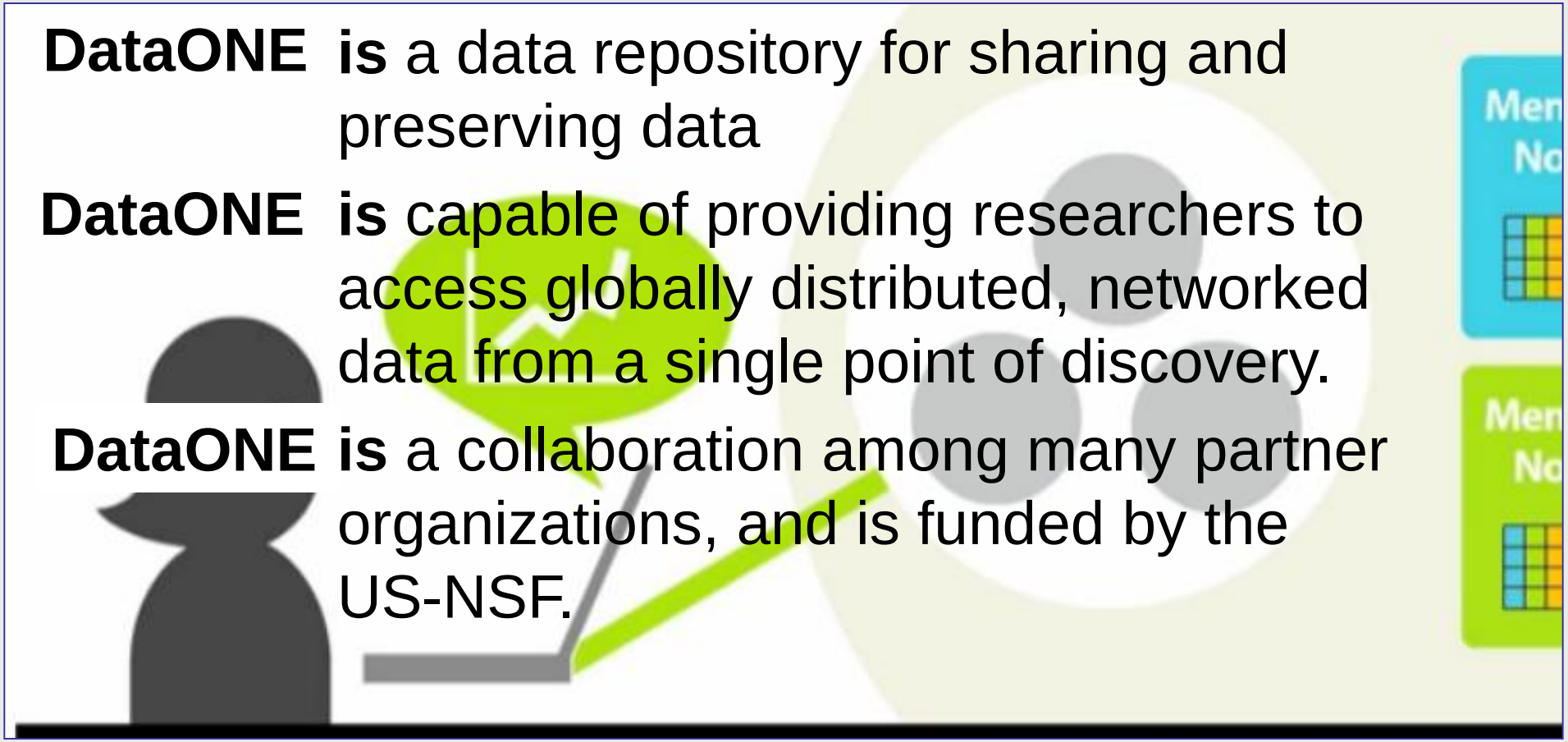
What is DataONE ?

Data Observation Network for Earth

DataONE **is** a data repository for sharing and preserving data

DataONE **is** capable of providing researchers to access globally distributed, networked data from a single point of discovery.

DataONE **is** a collaboration among many partner organizations, and is funded by the US-NSF.



[Through the knowledge and infrastructure integrates information]

National Center for Ecological Analysis and Synthesis (NCEAS), U.S.A; ; *http://www.dataone.org/what-dataone*.

Data Citation Index

Connecting the data to the research it forms

WHAT IT DELIVERS

- Digital research is discoverable, citable and linked to primary research literature.
- There are nearly 2 million records from quality repositories around the world.
- Records are built from descriptive metadata to create bibliographic records and cited references for digital research.
- The scholarly community for promoting standard citation formats for digital research is recommended.

http://wokinfo.com//products_tools/multidisciplinary/dci/

Data Descriptor

http://www.nature.com/scientificdata/?WT.mc_id=EMI_SCIDATA_1310_ISIOct

- **Scientific Data** is a new open-access, online-only publication for descriptions of scientifically valuable datasets. It makes your data more discoverable, interpretable and reusable.
- **Data Descriptors** are aimed at promoting data sharing and reuse but are not expected to contain new scientific conclusions or interpretations, which means you can publish about valuable data in the absence of specific new findings or when the data might be considered confirmatory.

Helping you publish, discover, and reuse research data

Credit

Credit, through a citable publication, for depositing & sharing your data

Reuse

Complete, curated & standardized descriptions enable the reuse of your data

Quality

Rigorous community based peer review

Discovery

Find datasets relevant to your research

Open

Promotes & endorses open science principles & available to all through a Creative Commons license

Service

In-house curation, rapid peer review & publication of your data descriptions

Data Descriptor

http://www.nature.com/scientificdata/?WT.mc_id=EMI_SCIDATA_1310_ISIOct

- **Data Descriptors** must describe ‘primary’ datasets: data that were directly produced by an experimental or observational procedure.
- **Scientific Data** publishes descriptions of scientifically valuable datasets, and requires that authors make described datasets available to editors and referees at the time of submission and share these datasets with the scientific community as a condition of publication.
- **Data Descriptors** are now accepting submissions More information can be found on their **website**

Data Sharing

- It is a difficult process
- It is a slow process
- It is a must process
- It requires a stepwise approach
- It requires a scientific paradigm
- It requires a cultural paradigm

A culture of presenting metada
A culture of data-intensive science
A culture of data curation and sharing
A culture of transparent data exchange and aggregation

OJ Reichman et al 2011; SE Hampton, …J Porter 2013

Measures to be taken

Data is the new currency for research

1. To organize and preserve our data for posterity,
2. To share data more routinely & through publicly accessible databases (such as dataONE).
3. To collaborate in networks where data are shared.
4. To address data management issue to junior researchers in our labs.
5. To use big data to answer the big ecological questions.
6. To put knowledge based on data into action.
7. To promote standard citation formats for digital research (such as Data Citation Index)

Moving to a world where all data is linked .

and can be stored/analyzed

in the Cloud

Data/information

is interconnected through

machine - interpretable information

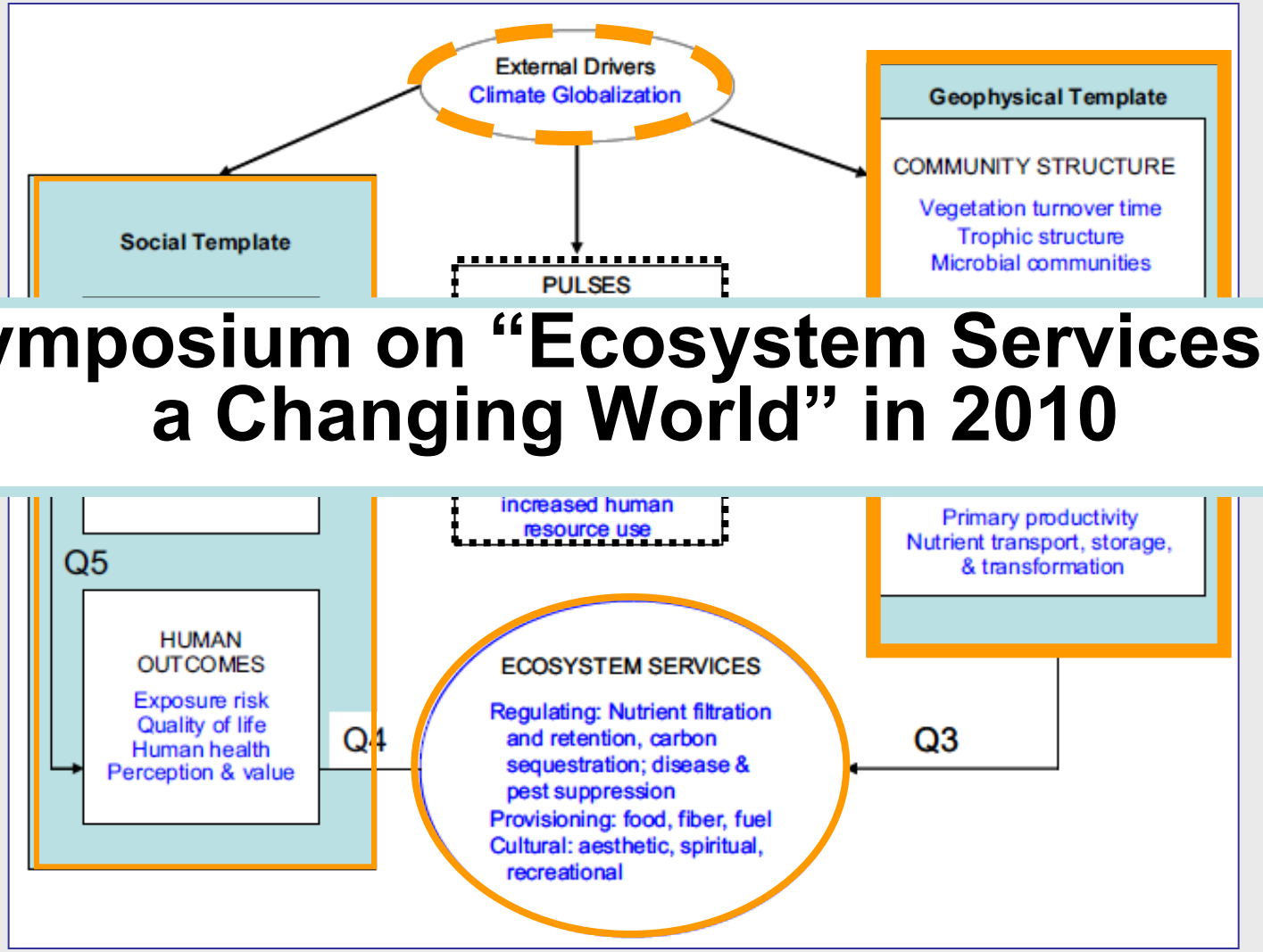
Tony Hey/Microsoft External Research

Ecosystem Management for Providing Services

Ecosystem services are defined as the benefits that human obtained form ecosystems. (Westman 1977)

The benefits that people derive from the structures and processes generated by ecological systems.

Conceptual Framework and Theoretical Foundations

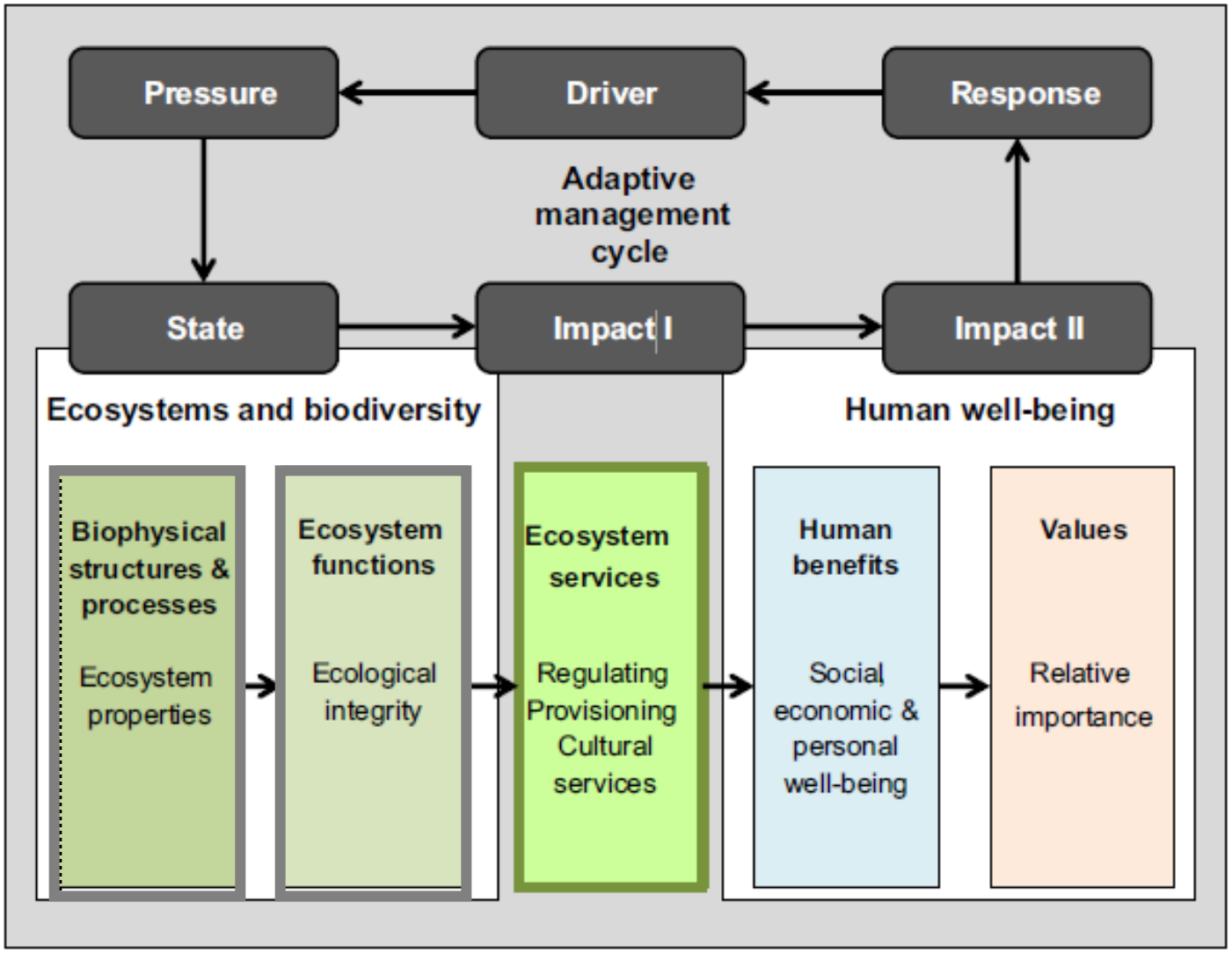


A symposium on "Ecosystem Services in a Changing World" in 2010

Integrative Science for Society and the Environment Decadal Plan, US-LTER, 2007

Human-environment System

***D*rivers**
***P*ressure**
***S*tate**
***I*mpact**
***R*esponse**

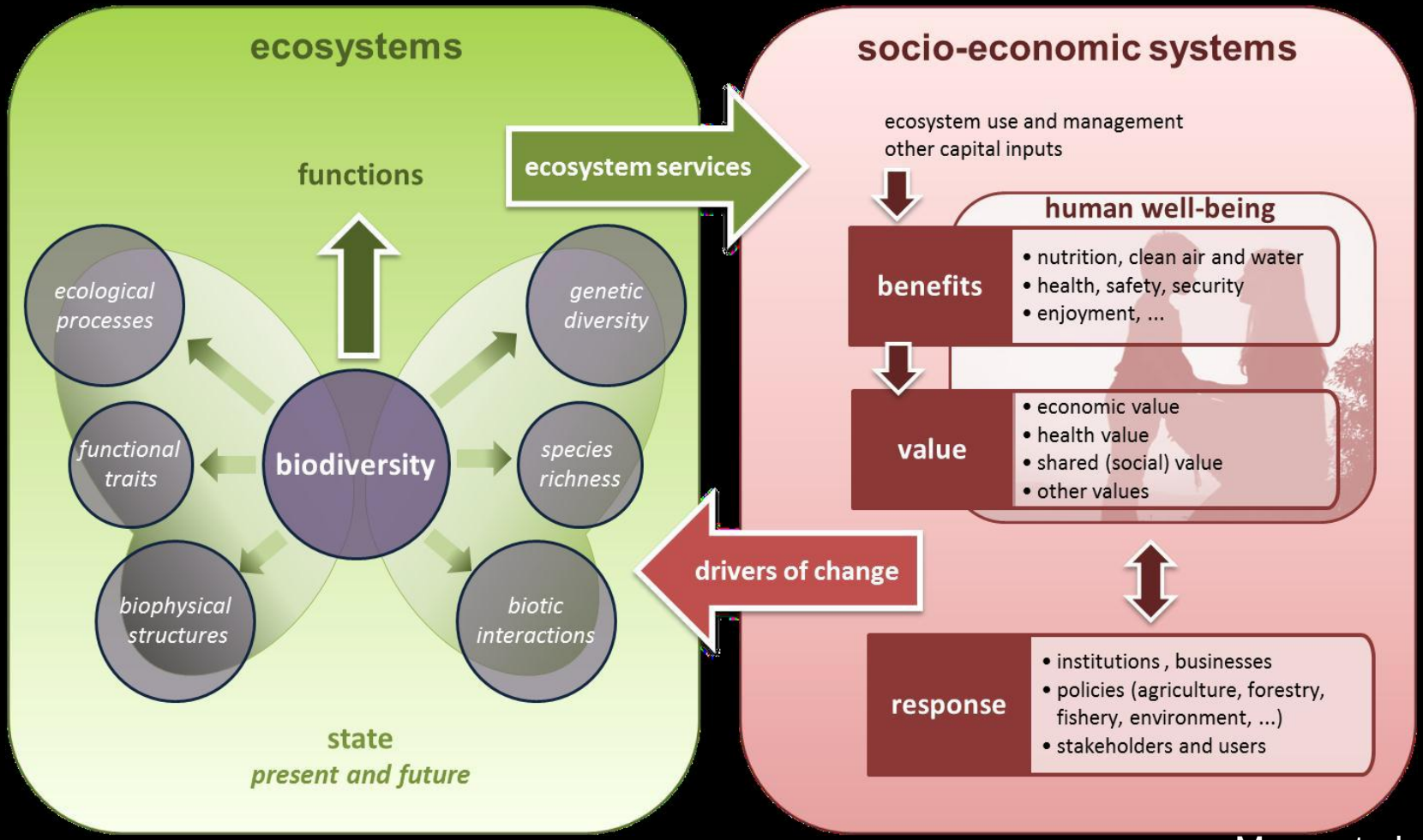


***DPSIR* Framework**

Muller & Burkhard, 2012

Ecosystems and Their Services

To map & assess on E & ESs to support the maintenance & restoration of ESs and their services



The EU 2012 Biodiversity Baseline (EEA 2012)

Maes et al. 2013

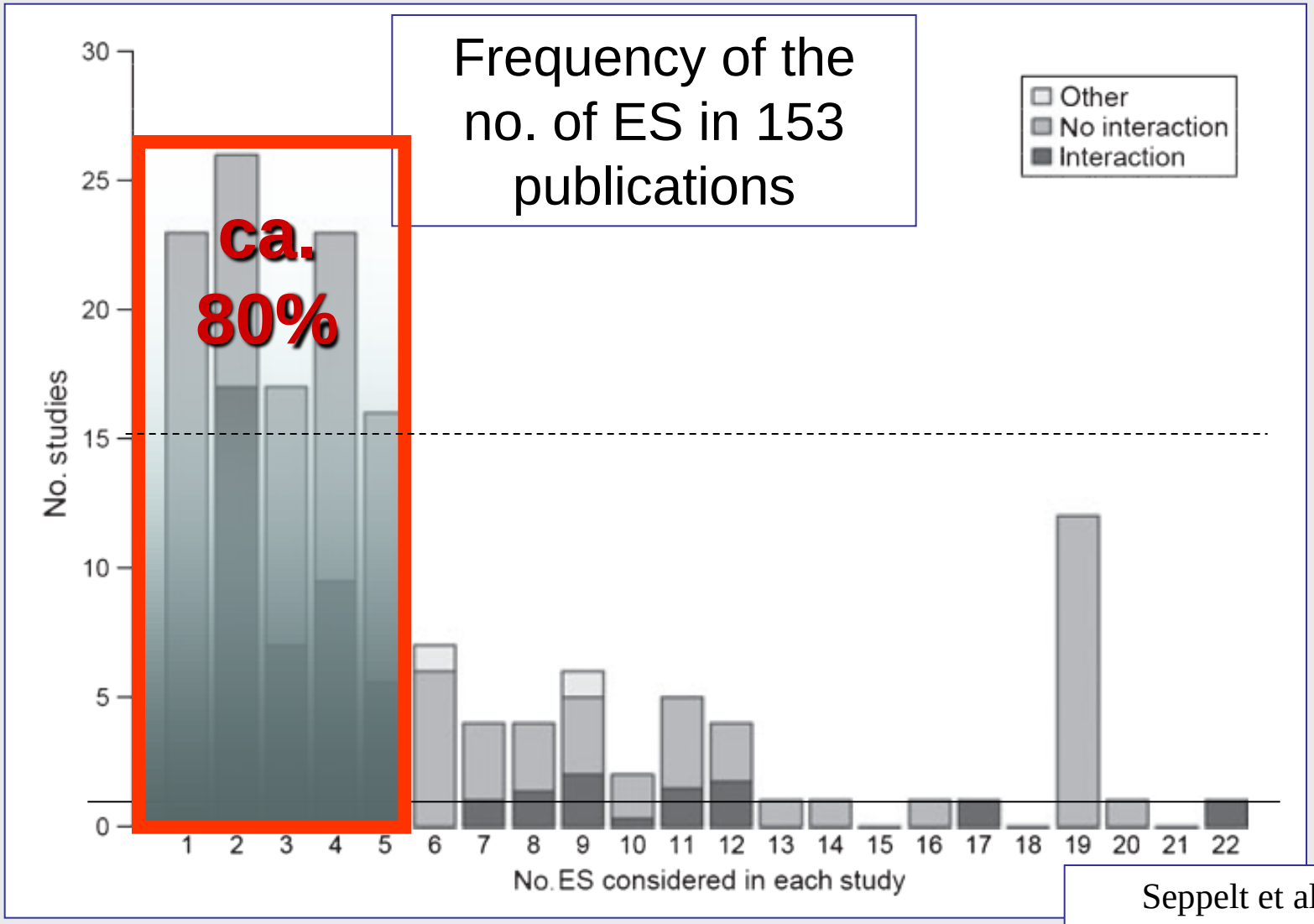
Ecosystems & Humans

Benefits that humans obtain from ecosystems

- The **MEA (2005)** contributed substantially in bring forward the ESs **concept** as a policy to achieve the sustainable use of natural resources.
- But a fully operational **method** to implement the concept was not delivered.
- **Mapping and assessing** the state of ecosystems and their services are major research topics.
- Provisioning of a critical **evaluation** of the best available information for guiding decision on complex public issues is urgently required.

Westman, W.E. (1977) How much are nature's services worth? Science,197,960–964.
(Seppelt 2013).;European Union 2013

Current Research on ESs



Seppelt et al 2011 *JAE*

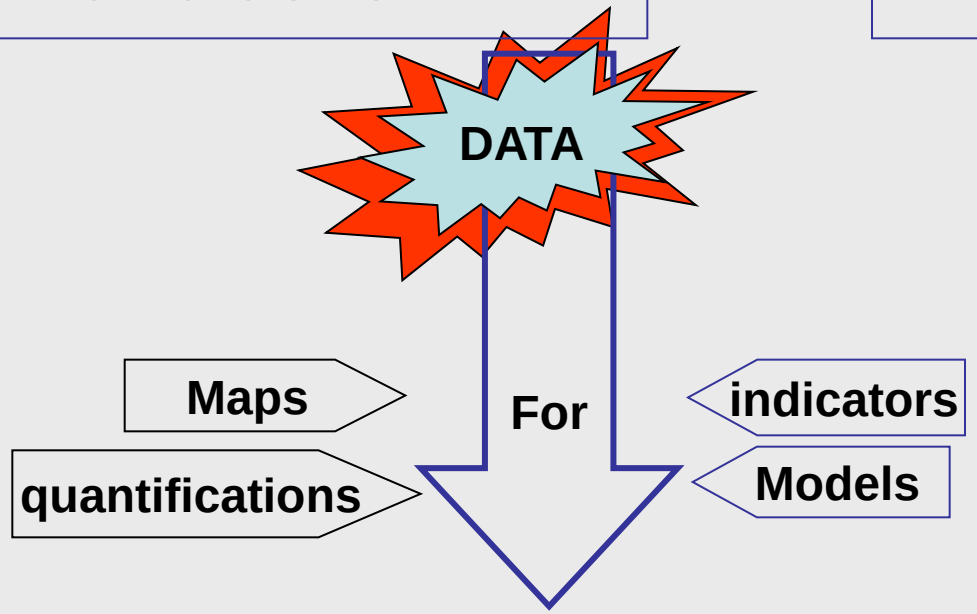
The Strength of FDPs
in Ecosystem Studies

- Our plots are in great numbers; varying in ecological settings; major ecosystems of each FDP site; distributed globally, …
- We have long term primary data (including matadata) and employ agreeable methodologies.
- The plots' financial support is relatively stable.
- We are linked and work closely within and with other ecological communities and beyond.

Three Major Topics

- To maintain and enhance
- To map & assess the state of
- To assess the value of

ecosystems and their services



Sciences, Policies and Practices

Big Data for Ecosystem Science

- We have large amount of distributed, high-quality information. We must manage them to improve data's accessibility and sharing capacity.
- Big data present unprecedented opportunities for advancing ecosystem studies through data-intensive approaches.
- We have to step into the age of information.
- Individual researcher should think big with sheer volume of accessible datasets.

Hampton, …, Porter 2013 ***ESA***

Action Plans

- To inform decisions with ecological knowledge
- To advance innovative and anticipatory research
- To convey the knowledge from science to be translatable to actions
- To stimulate cultural changes for a forward-looking and international ecology

- **We need a culture of data sharing.**
- **We should participate in collaboration to address large-scale questions**

- Htpp://www.esa.org.ecovisions; Hampton, …,and Porter. 2013

Positive thinkings are brighter

THANK YOU

Negative ones are darker

STEVE LOHR