DataONE Webinar Series

Open Science and Data:
Towards Optimizing the Research
Process

Dr Jean Claude Guédon
Université de Montréal

DataONE

DataONE

Enable new science and knowledge creation
through universal access to data about life on earth
and the environment that sustains it

DataONE network of Member Nodes:
www.dataone.org/current-member-nodes

DataONE search tool:
https://cn.dataone.org/onemercury/

DataONE

Lesson 3: Data Management Planning

View all Education Modules at https://www.dataone.org/education-modules

The Data Management Plan (DMP)

A DMP outlines what you will do with your data during and after you complete your research project. It is a formal (but evolving) document in which you lay out a plan... the present...

Component 2: Metadata content & format

- What metadata are needed?
- How will metadata be created and/or captured?

DataONE

Lesson 2: Data Sharing

View all Education Modules at https://www.dataone.org/education-modules

Address data sharing throughout the data lifecycle

Describe data content, character, and process.
Deposit in a location from which it can be accessed.
Preserve in formats & on media good for long term
Publish info... discover it e...

Concerns about data sharing

Concern	Solution

DataONE

Lesson 1: Data Management

View all Education Modules at https://www.dataone.org/education-modules

The world of data around us

The data deluge has created a surge of information that needs to be well-managed, discoverable, and accessible.

The amount of available storage is not keeping pace with the amount of data being produced.

Information vs. Available Storage

- Information
- Available Storage

Petabytes

2005 2006 2007 2008 2009 2010
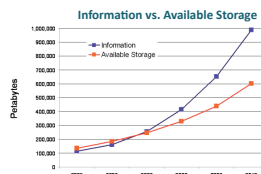
Gantz, The Expanding Digital Universe

Why manage data: the researcher perspective

- Keep yourself organized ⇒ find your own files!
- Track your processes for reproducibility
- Better version control of data
- More efficient data quality control
- More backups to avoid data loss
- Format your data for reuse by yourself & others
- Document your data for understability and reuse
- Prepare it to share it & gain credibility and recognition for your scientific efforts

Data Reuse
Data Sharing
Data Management

Data management facilitates sharing and reuse.

Causes of data loss

- Natural disasters
- Facilities infrastructure failures
- Storage failure
- Server hardware or software failure
- Application software failure
- Human errors
- Malicious attack
- Format obsolescence
- Loss of competencies
- Loss of funding
- Loss of institutional commitment

Costs of not doing data management can be very high!

Data Reuse Example

Researchers reused and aggregated data from several different sources to determine migration routes for specific bird species.

© Jerry & Sherry Liguori

The Case for Data Management

If data are:
- Well-organized
- Documented
- Preserved
- Accessible
- Verified as to accuracy & validity

The results are:
- High quality data
- Data that is easy to share and reuse
- Citation & credibility to researcher
- Cost savings to further science

The Data Lifecycle

The stages through which well-managed data passes from project inception to conclusion.

Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze

The Data Lifecycle

Local contact information

DataONE

Embed
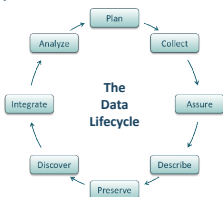
Presentations 10

Tutorials on Data Management

Lesson 10: Analysis and Workflows

Get content → check → store → Make Available

Analysis and Workflows

DataONE

◄ 1 of 34 ►

Documents 0
Infographics 0
Videos 0

#DWS2015
@DataONEorg
www.DataONE.org

DwS

DataONE Webinar Series
    www.dataone.org/webinars
Upcoming Webinar Event
    www.dataone.org/upcoming-webinar
Previous Webinar Events (Recording and Forum)
    www.dataone.org/previous-webinars

Next Webinar
    Tuesday March 10th 0900 PT / 12 noon ET
    Dr Cameron Neylon, Public Library of Science

#DWS2015
@DataONEorg

Quick Poll

Open Science and Data: Towards Optimizing the Research Process

Dr Jean Claude Guédon

Université de Montréal

Open Data and Science

Jean-Claude Guédon

In the beginning:
The ethos of science

Open Data and Science

Jean-Claude Guédon

Building on the shoulders of giants

(and others, not so giant...)

Open Data and Science

Jean-Claude Guédon

The scientific process is a mixture of

Collaboration

Communication

Competition

Open Data and Science
Jean-Claude Guédon

Competition is not the unique motivation for doing science

Collaborating and sharing through <u>publishing</u> are also fundamental

Open Data and Science

Jean-Claude Guédon

Sharing research results
is one of the hallmarks
of the Scientific Revolution

Open Data and Science

Jean-Claude Guédon

It was not always easy!

Here is how Galileo told Kepler about Saturn's satellites:

smaismrmilmepoetaleumibunenugttauiras

Open Data and Science
Jean-Claude Guédon

In case you did not get it, it means:

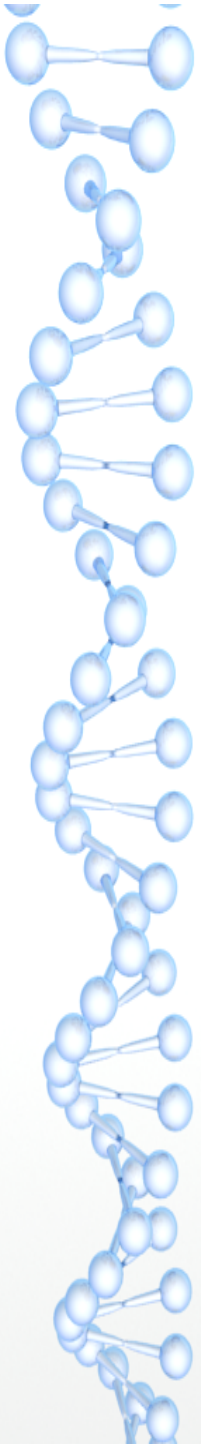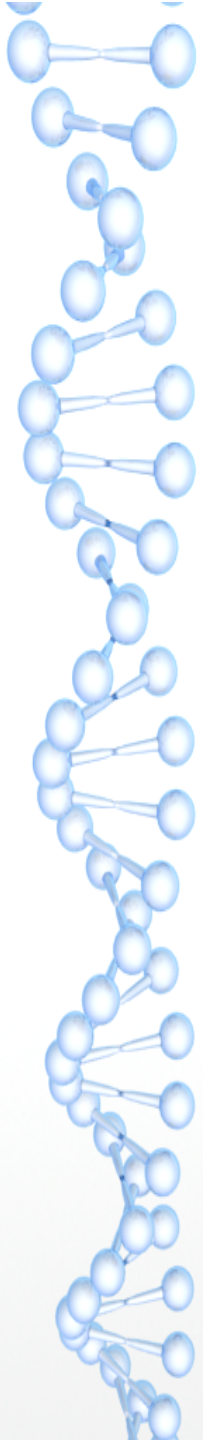## Altissimum planetam tergeminum observavi

Open Data and Science
Jean-Claude Guédon

Still having difficulties?

Well, it simply means:

I have observed the most distant planet to have a triple form

14

Open Data and Science
Jean-Claude Guédon

Now, that is clearer...

Don't you agree?

Open Data and Science

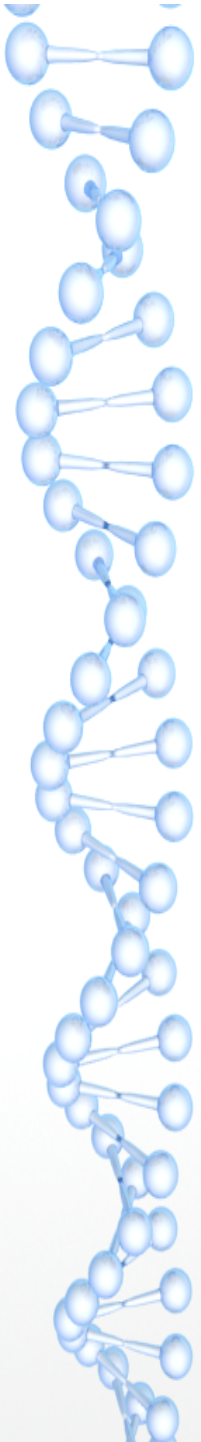Jean-Claude Guédon

You don't agree?

Well, you are right

Open Data and Science
Jean-Claude Guédon

This is the language of alchemists,

not that of modern scientists
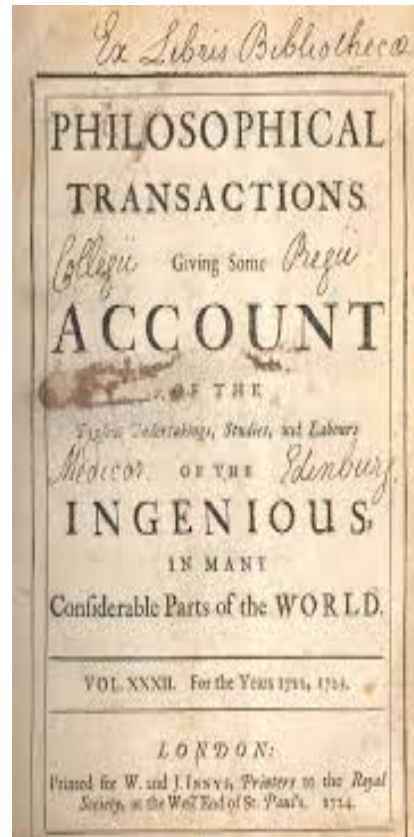
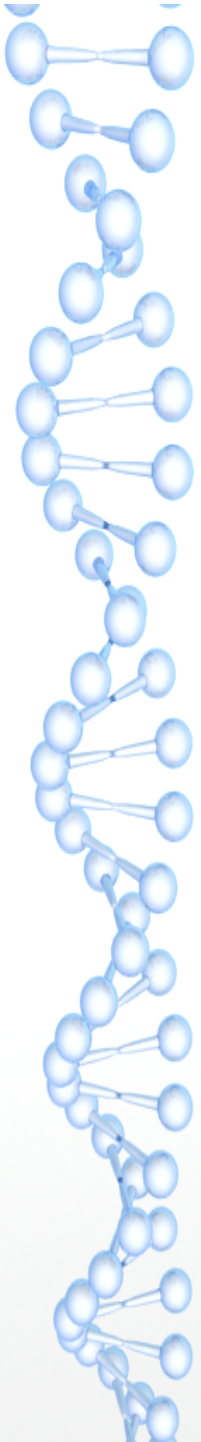# Open Data and Science

Jean-Claude Guédon

Open Data and Science
Jean-Claude Guédon

Something better was needed

Open Data and Science
Jean-Claude Guédon

Open Data and Science
Jean-Claude Guédon

Thus, the modern scientific journal

was invented

Open Data and Science
Jean-Claude Guédon

Nowadays, a new revolution is breeding:

Modelling!

Open Data and Science

Jean-Claude Guédon

An example?

Check out this weather model

vs. Actual observations:

http://vimeo.com/117256684

Open Data and Science
Jean-Claude Guédon

Across the centuries, researchers have learned to share their papers

Now, they must learn to share their data

Open Data and Science
Jean-Claude Guédon

HOW?

Open Data and Science
Jean-Claude Guédon

Two problems
related to this question:

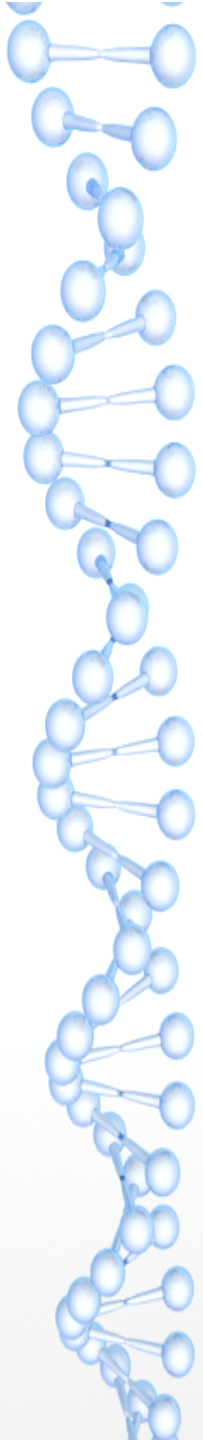- Scientific ethos
- Technical problems

Open Data and Science

Jean-Claude Guédon

Scientific ethos

The need to reward sharing by visibility and prestige

The need to preserve competitive edge where it matters

Open Data and Science

Jean-Claude Guédon

Technical problems

Retrieval

Interoperability

Cross-disciplinary capacity

Open Data and Science
Jean-Claude Guédon

First steps:

Data sets
linked with published papers
e.g. OpenAIRE+ in Europe

Open Data and Science

Jean-Claude Guédon

Second steps

Curation and metadata
e.g. EUDAT in Europe

Open Data and Science

Jean-Claude Guédon

Third steps

Preservation

e.g. Lockss in the USA

Open Data and Science

Jean-Claude Guédon

Fourth Steps

The means to evolve:

Standards

procedures

Over all this:

A governance model

Open Data and Science

Jean-Claude Guédon

Possible models:

The Internet (e.g. IETF)
Free software
(e.g. Free software Foundation)

Open Data and Science
Jean-Claude Guédon

In the Works

Research Data Alliance

Open Data and Science
Jean-Claude Guédon

Possible dangers?

Privatization of data

with dire consequences

for the prospects of optimizing research

for the common good

Reminders

Recording and Forum:
www.dataone.org/previous-webinars

Registration opening soon for:

TBA
Dr Cameron Neylon
Public Library of Science
Tuesday March 10th
www.dataone.org/webinars

#DWS2015
@DataONEorg