



Università  
Ca' Foscari  
Venezia

e-infrastructures  
austria



ELSEVIER

# A publisher point of view on open research data

Wouter Haak – VP Research Data Management

Nov, 2016


## New roles in Open Science and Data Stewardship

**As a researcher, when I wake up in the morning the first thing I think of is**



A) "I just love to write more papers about my research"

OR



B) "I just thought of a new way to filter my data that might actually lead to a result that makes more sense"

**So let's talk a bit about data**



## What are we really after: malaria



**WWARN**

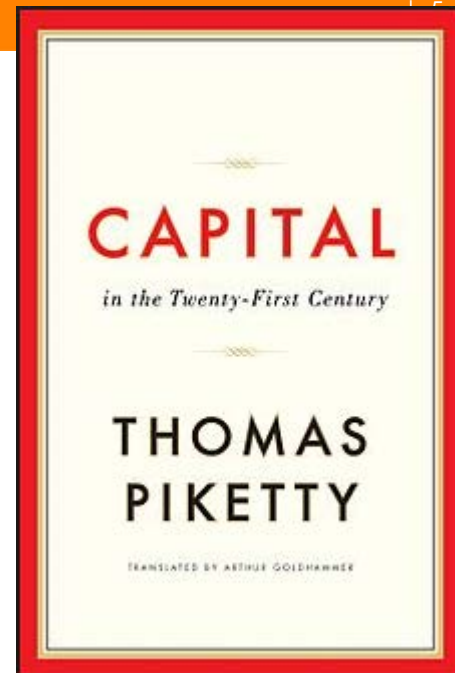
WWARN, the first malaria data sharing network, has used pooled analysis of shared data to provide evidence to **help improve dosing regimens** of malaria treatments

- [260 institutions globally](#) have worked with WWARN, and over 120,000 individual patient records have been contributed
- Based on the results, the World Health Organization has revised the recommended dose of DP, a commonly used antimalarial for young children

# What are we really after: social sciences

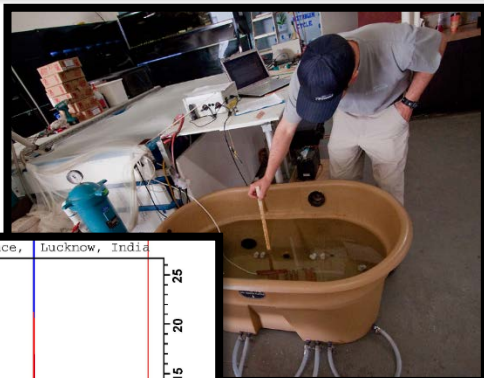
**Capital in the Twenty-First Century** is a 2013 book by French economist Thomas Piketty.

- It focuses on wealth and income inequality in Europe and the United States since the 18th century
- Central thesis is that when the rate of return on capital ( $r$ ) is greater than the rate of economic growth ( $g$ ) over the long term, the result is concentration of wealth, and this unequal distribution of wealth causes social and economic instability
- All raw data, normalized data, the analysis, and methods have all been made publicly available on a dedicated website <https://www.quandl.com/data/PIKETTY>

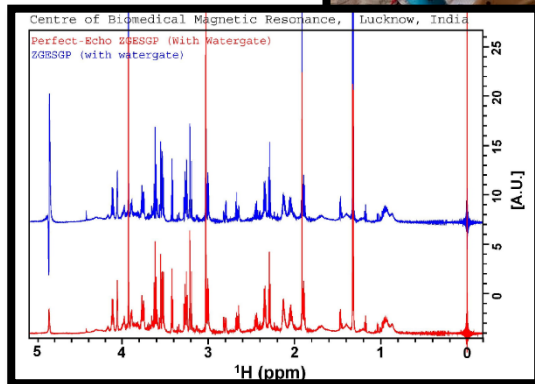
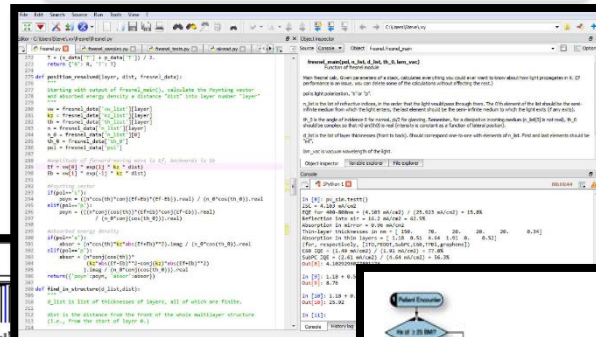


# When we talk about data, we really talk about the following:

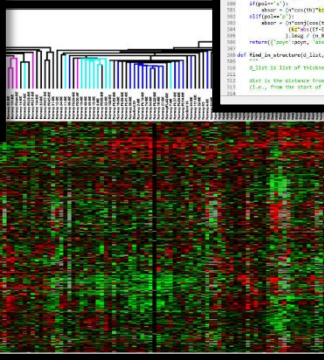
Machine & environment settings



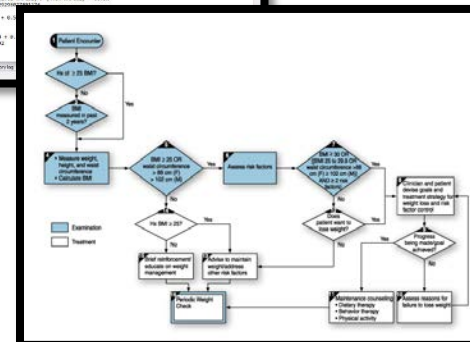
Scripts, analyses, algorithms



Raw data



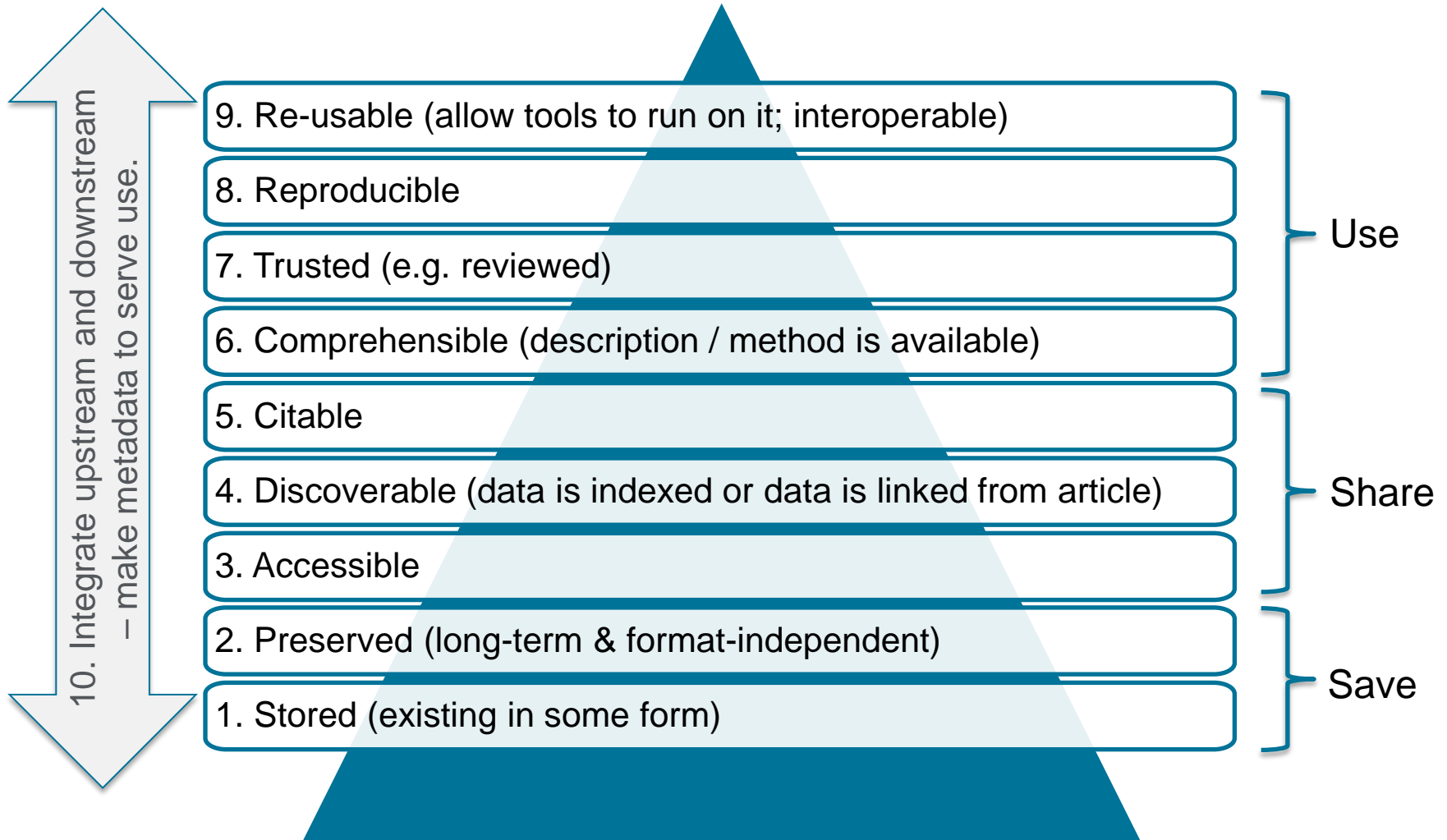
Processed data



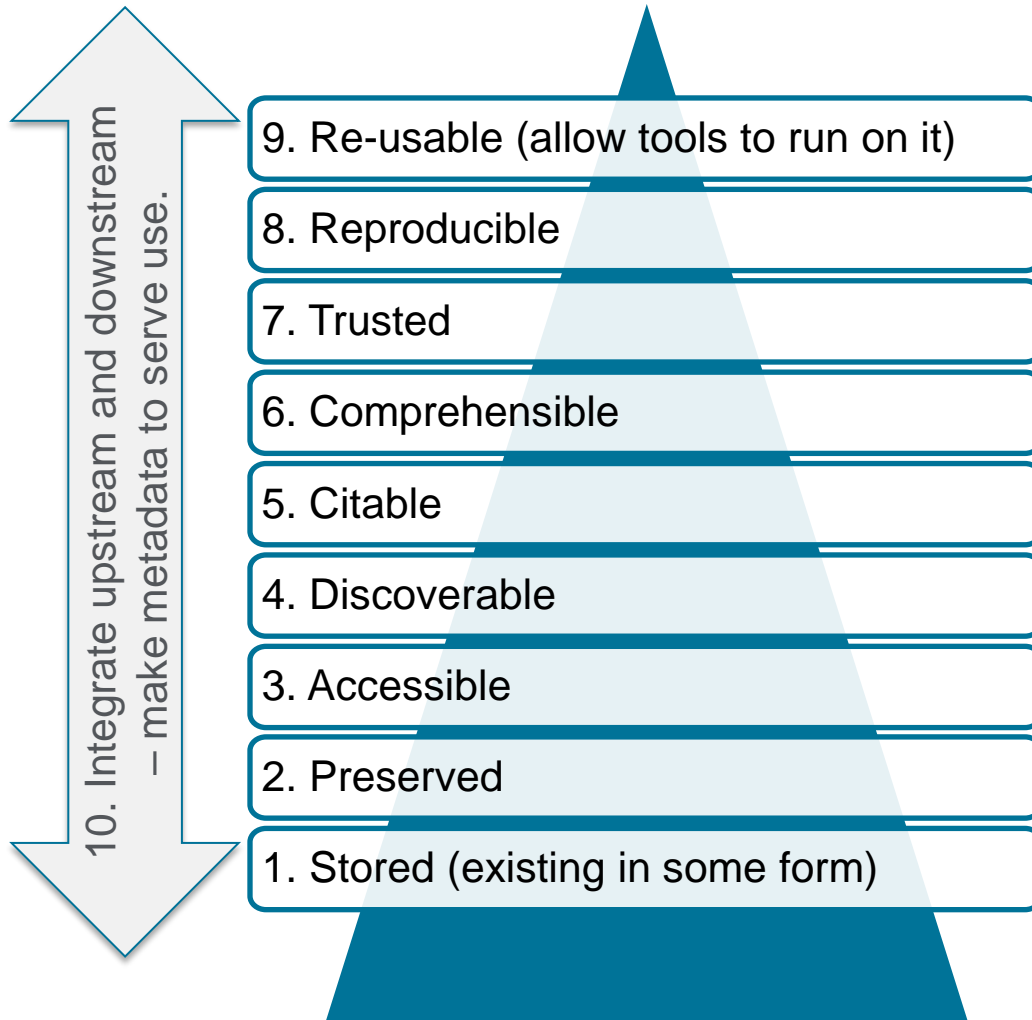
Protocols, methods, workflows

Full provenance needed

# The 10 components of effective research data



## Just another way of saying: “Data needs to be FAIR”



**F**indable  
**A**ccessible  
**I**nteroperable  
**R**eusable



# Compliance increasingly important for research data

## Some funder examples

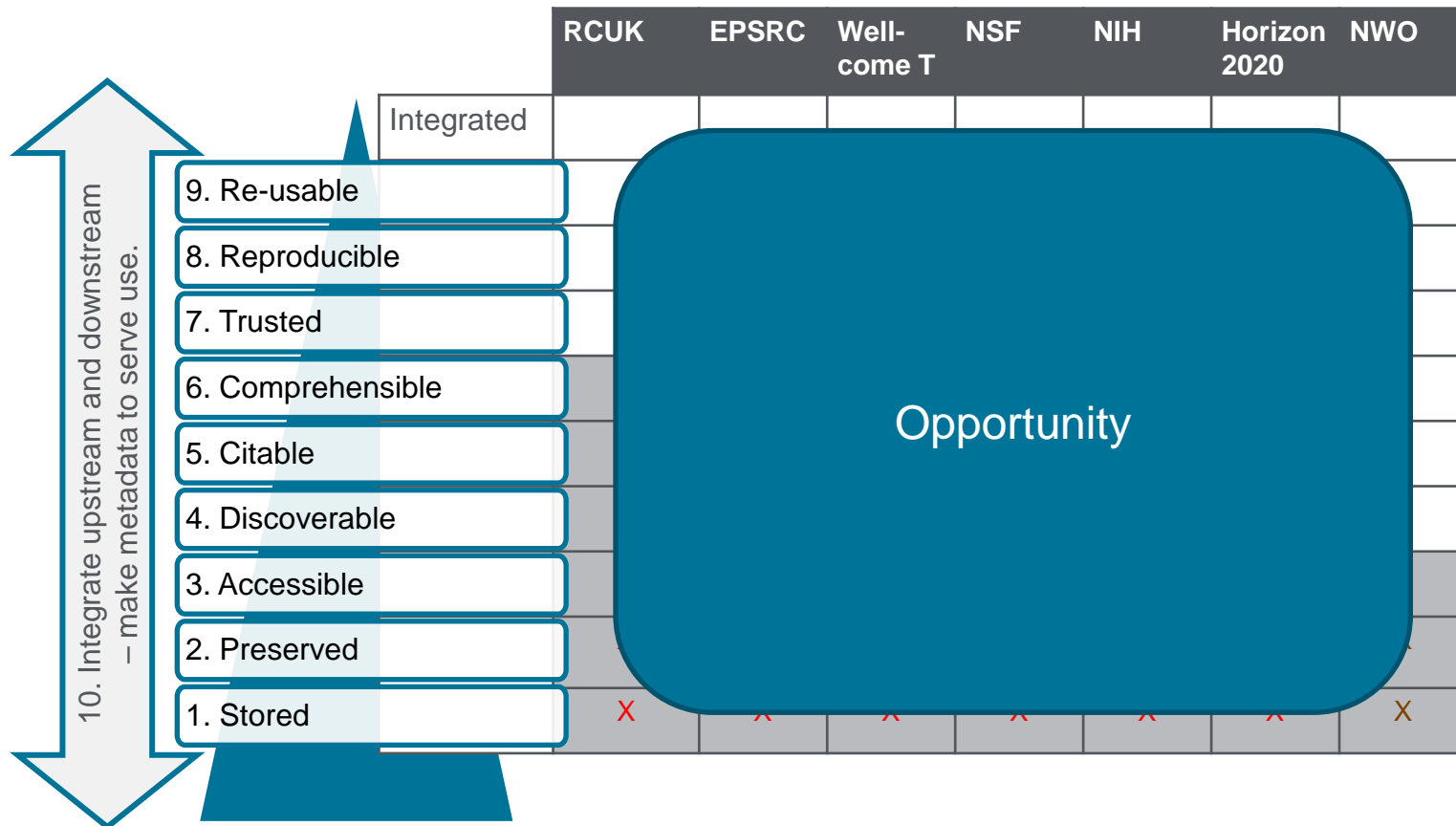
X) Required  
 X) Recommended

	RCUK	EPSRC	Well-come T	NSF	NIH	Horizon 2020	NWO
Integrated							
9. Re-usable							
8. Reproducible							
7. Trusted							
6. Comprehensible	X	X			X	X	
5. Citable	X	X			X		
4. Discoverable	X	X			X	X	
3. Accessible	X	X	X	X	X	X	X
2. Preserved	X	10 Years Minimum	10 Years Minimum	Duration depends	Duration depends	X	X
1. Stored	X	X	X	X	X	X	X

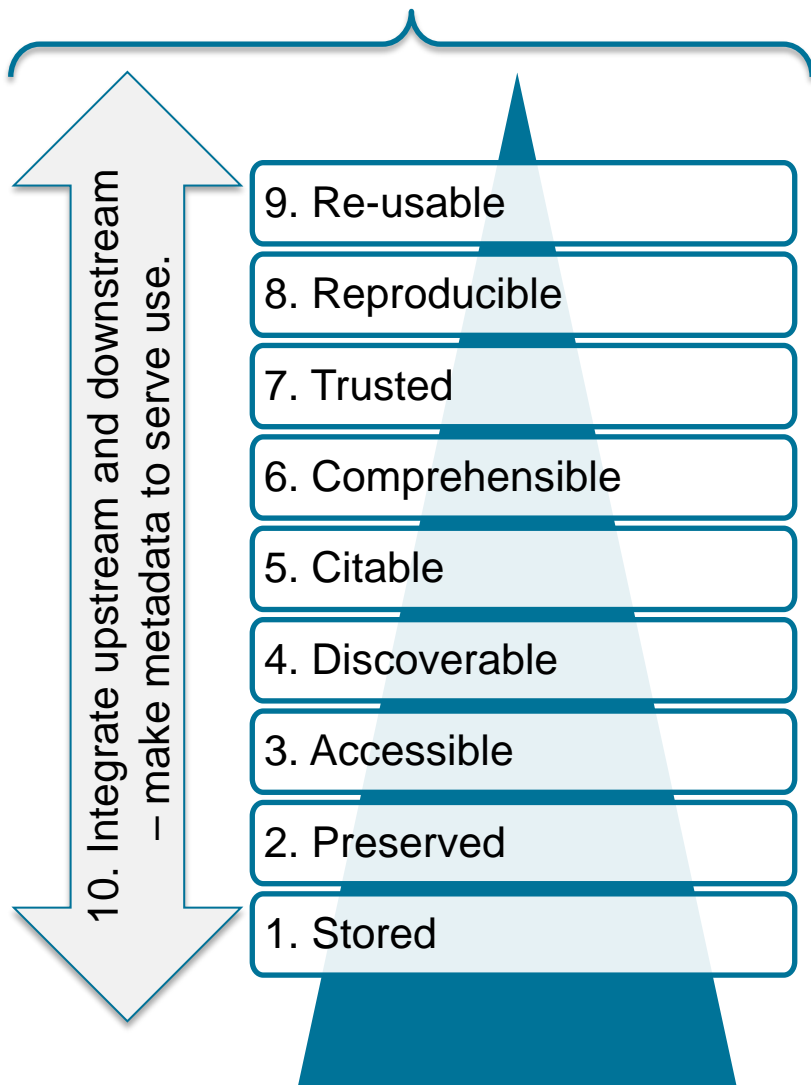
10. Integrate upstream and downstream  
 - make metadata to serve use.

Based on analysis of policies; august 2016  
[https://docs.google.com/document/d/1ODK1D4qs5KI59o-ADo-Ps5\\_zhzdOlc\\_E0ur8EEB1hIY/](https://docs.google.com/document/d/1ODK1D4qs5KI59o-ADo-Ps5_zhzdOlc_E0ur8EEB1hIY/)

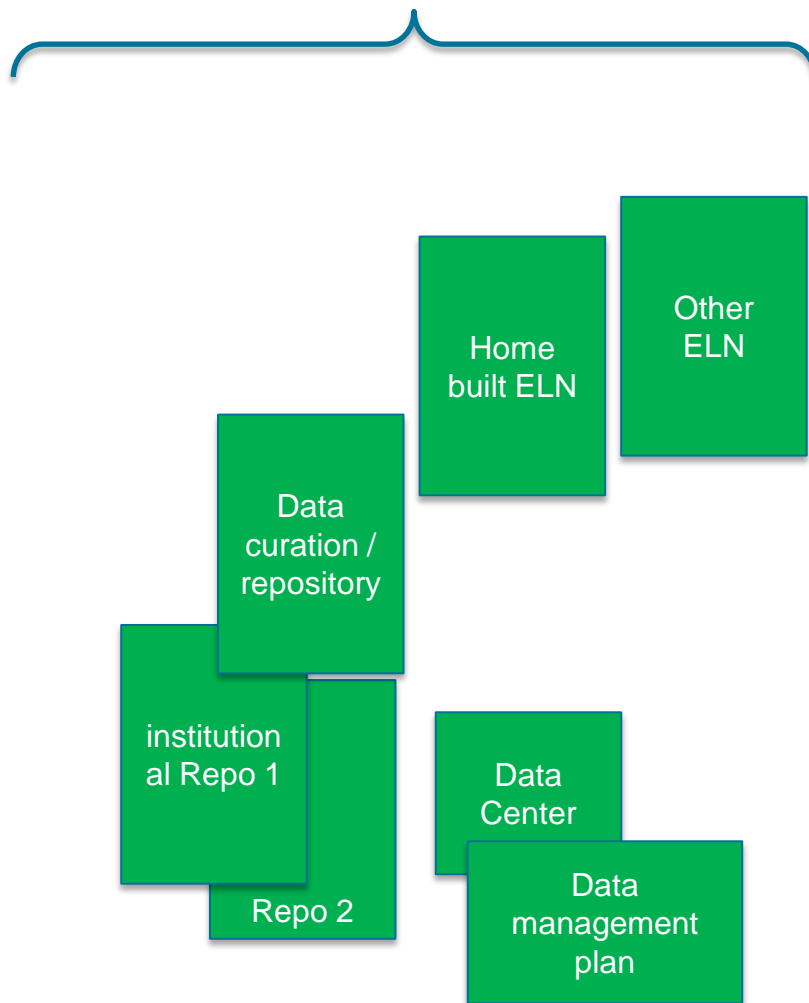
# Opportunity to move beyond compliance!



## The 10 components for effective research data




## Most researchers already work with several data tools



**What can Elsevier do?**


# Linking papers to data: phase 1

## Elsevier Database linking program – example Pangaea





### Marine Geology

Volume 204, Issues 1–2, 28 February 2004, Pages 43–57



---

#### Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of *Globigerina bulloides*. A proxy to determine deep-water circulation patterns?


A.N.A Volbers  , R Henrich

University of Bremen, Faculty of Geosciences, Department of Paleoceanography and Sedimentology, P.O.Box 330440, D-28334 Bremen, Germany


---


**Abstract**

The modern Atlantic Ocean, dominated by the interactions of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), plays a key role in redistributing heat from the Southern to the Northern


[http://dx.doi.org/10.1016/S0025-3227\(03\)00372-4](http://dx.doi.org/10.1016/S0025-3227(03)00372-4) 

[Get rights and content](#)

Bibliographic information 


Citing and recommended articles 


Applications and tools



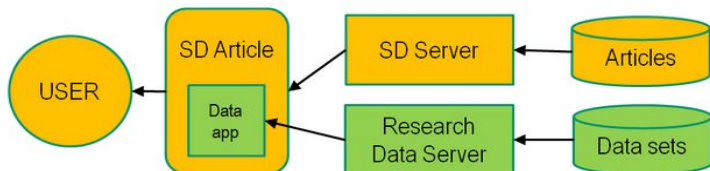
**PANGAEA® – Related Data**

Dissolution index of *Globigerina bulloides* in recent and Last Glacial M...



Workspace 

- Supplementary data at PANGAEA
- Bidirectional links between PANGAEA & ScienceDirect
- Data visualized next to the article

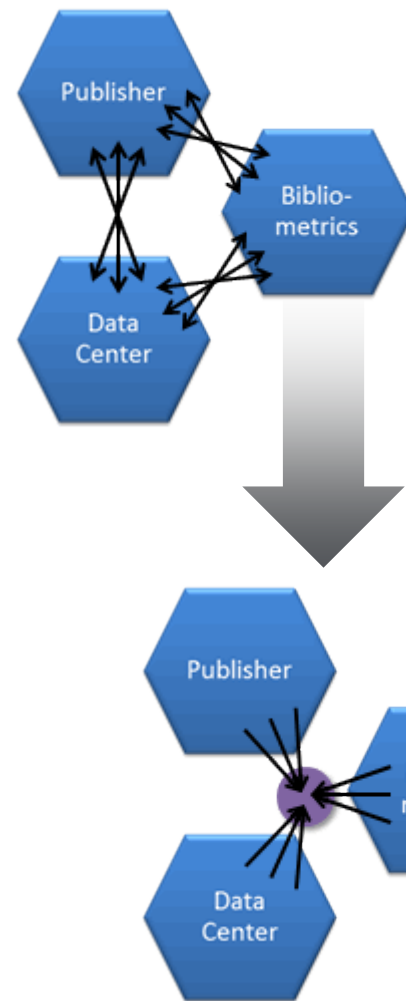


## Linking papers to data: phase 2

### [www.Scholix.org](http://www.Scholix.org)

# SCHOLIX

- ICSU/WDS/RDA [Publishing Data Service Working group](#)
- Creating linked-data model for exposing DOI to DOI links outside publisher's firewall
- Collaboration between CrossRef, DataCite, Europe PubMed Central, ANDS, Thompson Reuters, Elsevier, OpenAire



Objective: move from

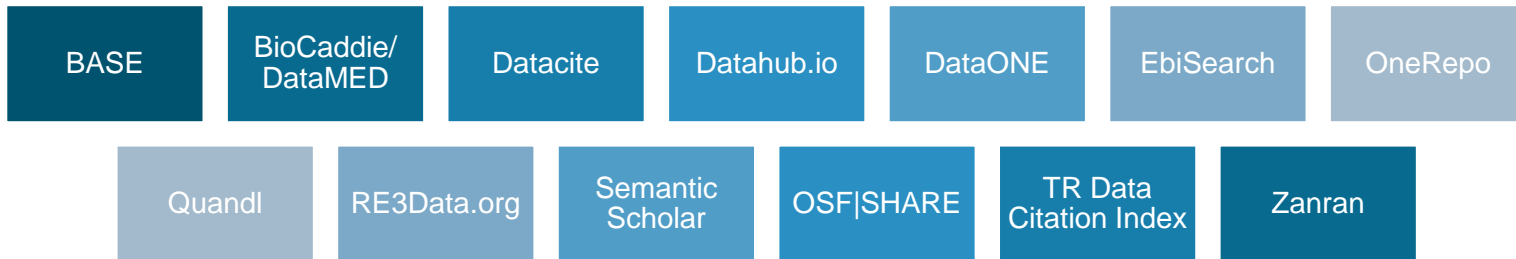
a plethora of (mostly) bilateral arrangements between the different players...

.. to ..

.. **a one-for-all cross-referencing service** for articles and data

## Phase 3: we need a research datasearch engine!

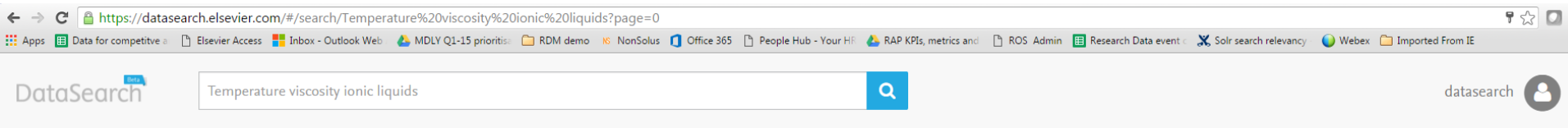
- Many (broad) datasearch examples already available



- Some common themes:
  - search of metadata only (i.e. ranking based on metadata only)
  - And/or federated search (i.e. no ranking)
  - And/or focused on giving credit (citation) rather than on discoverability
- Uncommon (because difficult):
  - Deep indexing of datasets (so real ranking and filtering)
  - Search engine really focused on data discovery

# Phase 3: Elsevier Data Search

## E.g. search for "Temperature viscosity ionic liquids"



**Filter Results** reset

1620284 results for *Temperature viscosity ionic liquids*

**Types**

- Image (1415019)
- Tabular Data (872183)
- Document (55138)
- File Set (2985)
- Raw Data (1331)
- Video (1174)
- Slides (825)
- Software (8)
- Statistical Data (7)

**Sources** 1

- ScienceDirect (1453411)
- arXiv (144281)
- PubMed Central (12729)
- ThermoML at NIST TRC (7770)
- NeuroElectro (1361)
- Dryad (432)
- PetDB (215)
- ICPSR (49)
- Harvard Dataverse (33)
- Mendelev Data (3)

2

**Solvent Properties of Functionalized Ionic Liquids for CO<sub>2</sub> Absorption**  
*L.M. Galán Sánchez, G.W. Meindersma & A.B. de Haan - 2006-10-24*  
 Ionic liquids can be used as solvents for gas absorption operations in order to improve the process economy and general efficiency of gas separations. This work investigates solvent properties of ionic liquids and compares them to amine solutions used for absorption of carbon dioxide (CO<sub>2</sub>). The CO<sub>2</sub> solubility into six different room temperature...  
 IMAGE 2 TABULAR DATA 3

**Ion conductive characteristics of ionic liquids prepared by neutralization of alkyimidazoles**  
*Hiroyuki Ohno & Masahiro Yoshizawa - 2002-03-20*  
 A wide variety of ionic liquids was prepared by the neutralization of five kinds of imidazole derivatives and nine kinds of acids. Their physical and chemical properties such as melting points, glass transition temperature, viscosity, and ionic conductivity were studied. Among these, imidazoles neutralized with imide-type acids were revealed to have...  
 IMAGE 3 TABULAR DATA 4

**CO<sub>2</sub> removal with 'switchable' versus 'classical' ionic liquids**  
*E. Privalova, M. Nurmi, M.S. Marañón, E.V. Murzina, P. Maki-Arvela, K. Eränen, D.Yu. Murzin & J.-P. Mikkola*  
 ► Comparison of ionic liquid systems acting as chemical vs. physical solvents in practical terms. ► Recycling and reuse issues of ionic liquids in carbon dioxide capture. ► Introduction of new types of switchable ionic liquids. ► Studying the capture-release cycle behavior of the aforementioned ionic liquids. ► In essence, we have focused on engineerin...  
 IMAGE 3 TABULAR DATA 3

**DataSearch.Elsevier.com**

1. Across repositories
2. (Deep) indexing of data, so not just metadata
3. Data preview

**Temperature dependence of viscosity for room temperature ionic liquids**

● Description

- Table 4
- Table 8
- Fig. 5
- Table 3
- Table 2

RTIL <span>3</span>	T/ °C									
	10	15	20	25	30	35	40	50	60	70
[Triethylsulfonium][Imide]	56	46	39	33 (30) [49]	28	24	21	17	13	11
[1-Butyl-3-methylimidazolium][Imide]	84	69	58 (52) [36]	47	41	34	29	22	17	14
[1-Butyl-1-methylpyrrolidinium][Imide]	128	106	89	76 (70) [50]	64	54	46	34	26	20
[1-Butyl-3-methylimidazolium][CF <sub>3</sub> SO <sub>3</sub> ]	131	107	90 (90) [36]	74	61	52	44	33	25	19

**Table 4**  
 TABULAR DATA  
 Absolute viscosities (in mPas) of air/moisture-tolerant RTILs at 10 temperatures



# Hivebench Electronic Lab Notebook

[www.hivebench.com](http://www.hivebench.com)

hivebench

Search for results

Notebook

Protocol

Inventory

Data



## Notebook

Team Notebook 15

experiments 13  
schizophrenia risk genes

My Notebook 1

Lab notebook

helena 05/21/2016

Schizophrenia risk gene x

helena 04/15/2016

schizophrenia risk gene mir137

helena 03/07/2016

ZNF804A expression

helena 02/05/2016

expression of tcf4

helena 01/21/2016

expression of znf804a

helena 01/12/2016

fetal brain expression

helena 12/31/2015

expression ZNF804A in human

Experiments

### schizophrenia risk gene mir137

helena created on April 15th 2016



-Wash in PBS with 0.1%Tween20, 3 X, and 10 min of each time

-Incubate with secondary antibody (peroxidase-conjugated goat anti-mouse IgG, etc.) following the manufacturer's instruction, usually 1:200 to 1:2000; 1 hour at room temperature

-Wash in PBS with 0.1%Tween20, 3 X, and 10 min of each time

5. Developing: (Use of chromomeric substrate- 3,3'-diaminobenzidine, DAB)

-Transfer the membrane to a shallow tray.

-Add 10ul H2O2 (30%) to 10 ml of 0.05%DAB in PBS, mix well immediately;

-Pour the DAB to the membrane, incubate at room temperature with gentle shaking in the dark if possible.

-Monitor the progress of the reaction carefully. When the bands are of the desired intensity (2-5 min), wash the filter briefly in water, and in PBS.

-Dry the membrane and photograph it to provide a permanent record of the experiment.



Protocols and results

# Manage, Store: Mendeley Data repository

**MENDELEY DATA** Beta Browse My datasets New dataset Log in [Create account](#)

## Reproducible experiments on dynamic resource allocation in cloud data centers

DOI: 10.17632/xz6gv65m6d.6  
Contributor(s): Andreas Wolke

**Description of this data**  
In Wolke et al. we compare the efficiency of different resource allocation strategies experimentally. We focused on dynamic environments where virtual machines need to be allocated and deallocated to servers over time. In this companion paper, we describe the simulation framework and how to run simulations to replicate experiments or run new experiments within the framework.

**Experiment data files** [Download all files \(6\)](#)

- Results.zip** (63 KB)
 

CSV files with simulation and experimentation results.
- github.paper.IS2015-master.zip** (8 MB)
 

<https://github.com/jacksonson/paper.IS2015/tree/7165452f4e9c540f98e1e57058de06f9fb192e8f>
- github.workload-master.zip** (222 MB)
 

<https://github.com/jacksonson/workload/tree/713dc5382b82e4ec1e1b6a998c80af3f7c08219f>
- Dockerfile** (1 KB)
 

Used to create the Docker container provided in IS2015.tar.gz
- IS2015.tar.gz** (1.3 GB)
 

Docker container file with installed simulation framework. Run simulations: (cd /root/work/paper.IS2015/control/Control && ./startsim\_reprozip) Run analysis: (cd /root/work/paper.IS2015/analysis && ./startanalysis-sim)
- reprozip.rpz** (160 MB)
 

ReproZip package of the simulation executed in the Docker container.

**Version 6** | Published: 13 Dec 2015

This data is associated with the following peer reviewed publication:

Reproducible experiments on dynamic resource allocation in cloud data centers

[Cite this article](#)

Published in:  
Information Systems

**latest version**

---

**Version 6** 2015-12-13  
Published: 2015-12-13  
DOI: 10.17632/xz6gv65m6d.6

[Cite this dataset](#)

Wolke, Andreas (2015), "Reproducible experiments on dynamic resource allocation in cloud data centers", Mendeley Data, v6  
<http://dx.doi.org/10.17632/xz6gv65m6d.6>

**Previous versions**

Version 5	2015-11-21
Version 4	2015-11-16
Version 3	2015-11-14
Version 2	2015-11-14
Version 1	2015-10-12

**Version comparison**

Version 5

Linked to published papers – or not

Researcher in control: how/when to share

Versioning and provenance


# Publish data and software: e.g. SoftwareX and DataInBrief have top 25% most read OA articles on ScienceDirect

Home > Books & Journals > SoftwareX

## SoftwareX

Editors-in-Chief: Dr. Kate Keahey, Dr. Frank Seinstra, Dr. David Wallom  
View full editorial board

[Open Access](#)



ISSN: 2352-7110



### AWARD FOR INNOVATION IN JOURNAL PUBLISHING



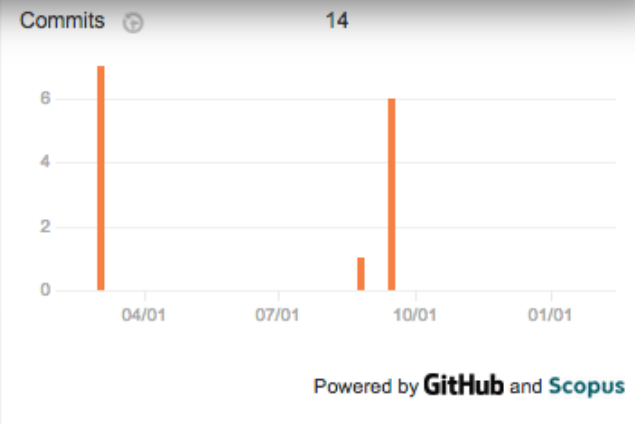
**Elsevier**  
*SoftwareX*

Editors Dr. Kate Keahey, Dr. Frank Seinstra and Professor David Wallom

## Code metadata

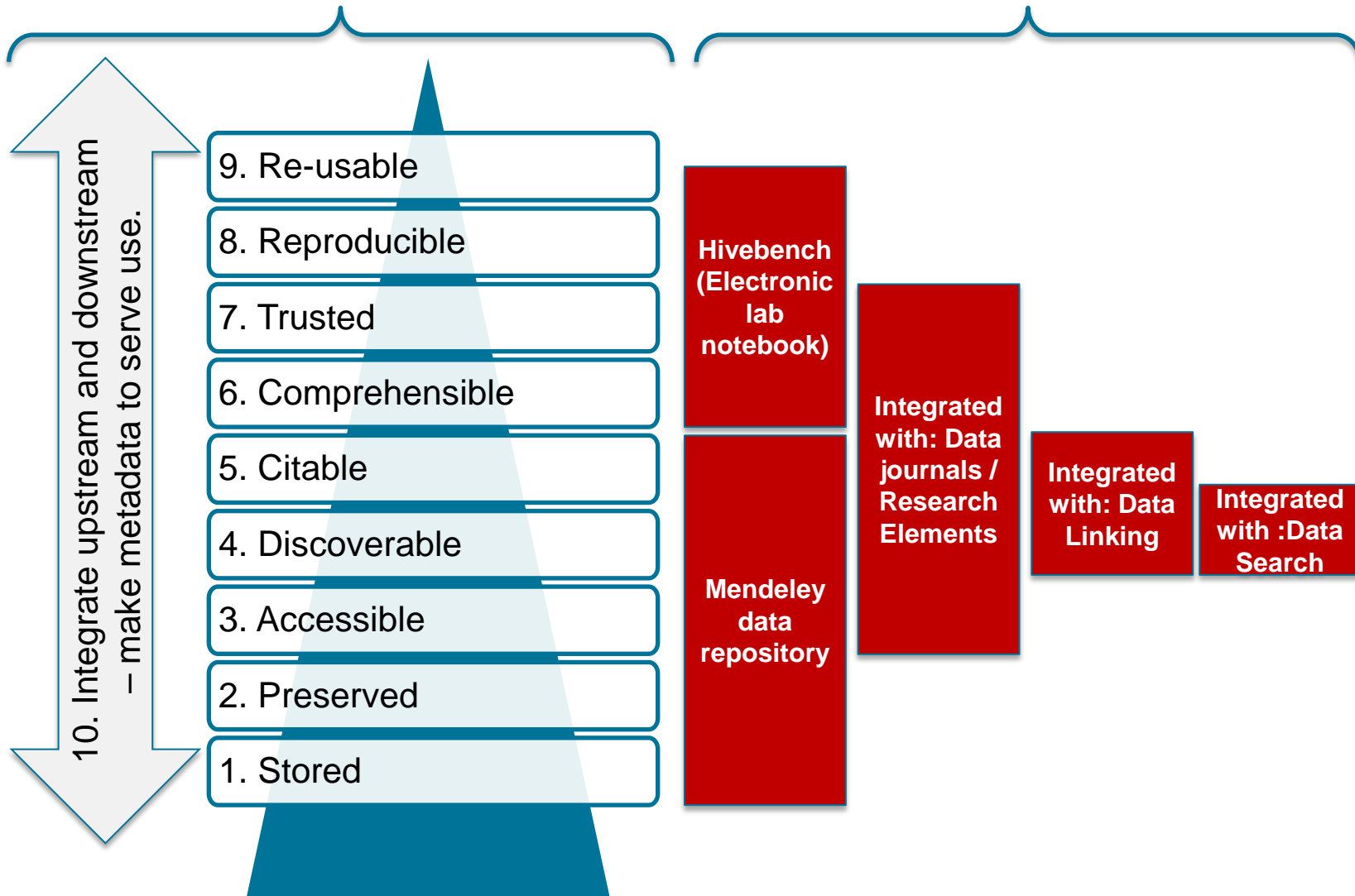
Current code version	v0.6
Permanent link to code/repository used of this code version	<a href="https://github.com/ElsevierSoftwareX/SOFTX-D-15-00005">https://github.com/ElsevierSoftwareX/SOFTX-D-15-00005</a>
Legal Code License	<i>NCSA open source license</i>
Code versioning system used	<i>git</i>
Software code languages, tools, and services used	<i>C, C++, Python, Bash; MPI, OpenMP, CUDA</i>
Compilation requirements, operating environments & dependencies	<i>Compilers: GNU/Intel/Cray; OS: Linux (RedHat, Debian, Ubuntu, CentOS, SUSE); Dependencies: GDAL, GEOS, PROJ4, SPRNG, PySAL, OpenGeoDa, etc.</i>
If available Link to developer documentation/manual	<a href="https://github.com/cybergis/cybergis-toolkit">https://github.com/cybergis/cybergis-toolkit</a>  <a href="http://cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php/ct">http://cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php/ct</a>
Support email for questions	CyberGIS Helpdesk ( <a href="mailto:help@cybergis.org">help@cybergis.org</a> )

Table options ▼

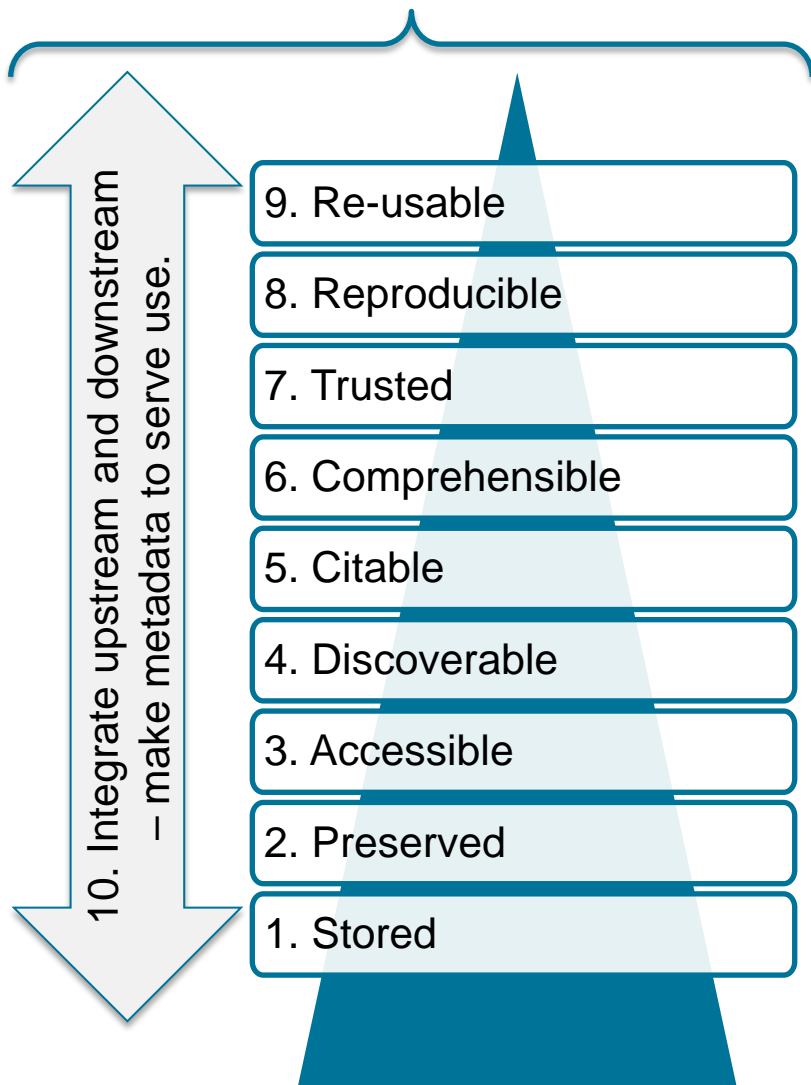


## The 10 components for effective research data

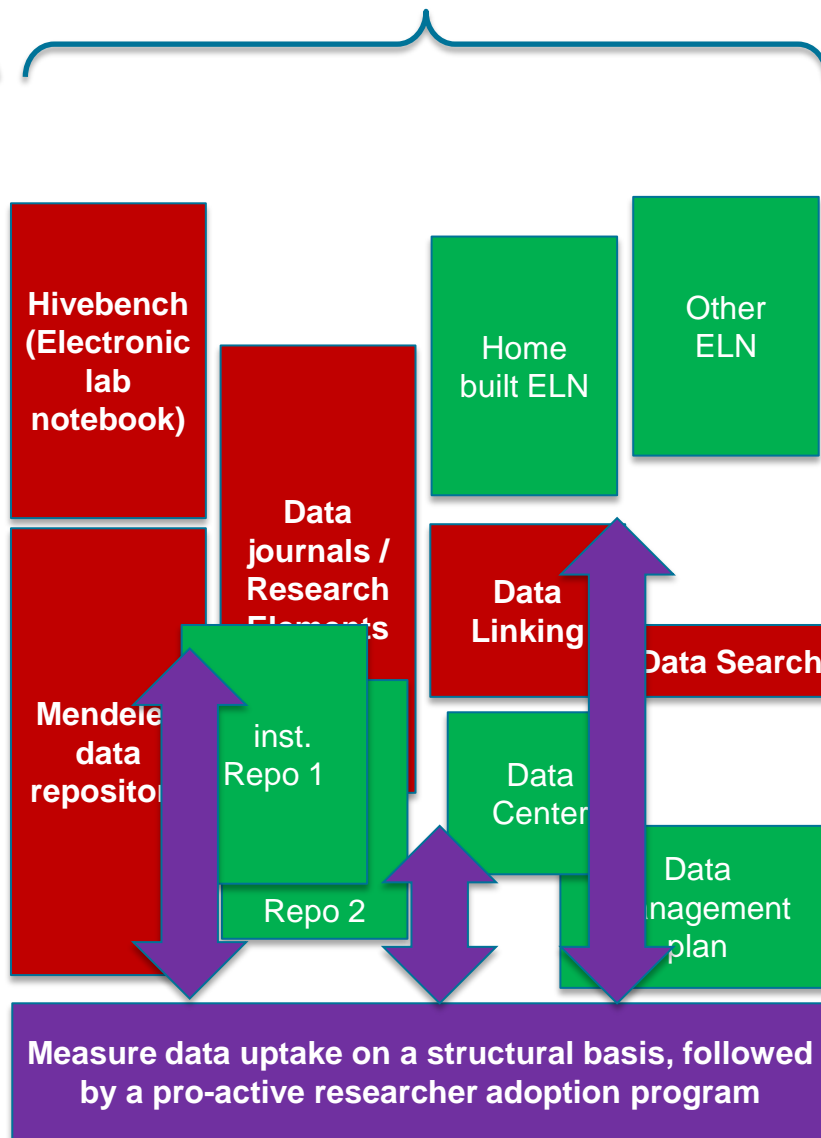
## Summary of initiatives discussed



## The 10 components for effective research data



## Moving towards an integrated ecosystem to help the researcher and the institution



**Grazie mille!**

