

# ERA-CAPS Expert Working Group on data standards

Christian Bachem

Ruth Bastow

George Gkoutos

Wilhelm Gruissem

José Pereira Leal

Hadi Quesneville

Ulrich Schurr

Björn Usadel

Theo Saat

Paul Wiley

Uli Schurr

Forschungszentrum Jülich

Germany

[u.schurr@fz-juelich.de](mailto:u.schurr@fz-juelich.de)

# Common data sharing policy of ERA-CAPS

- Adopted March 2014

Publicly-funded research data (following OECD 2007)

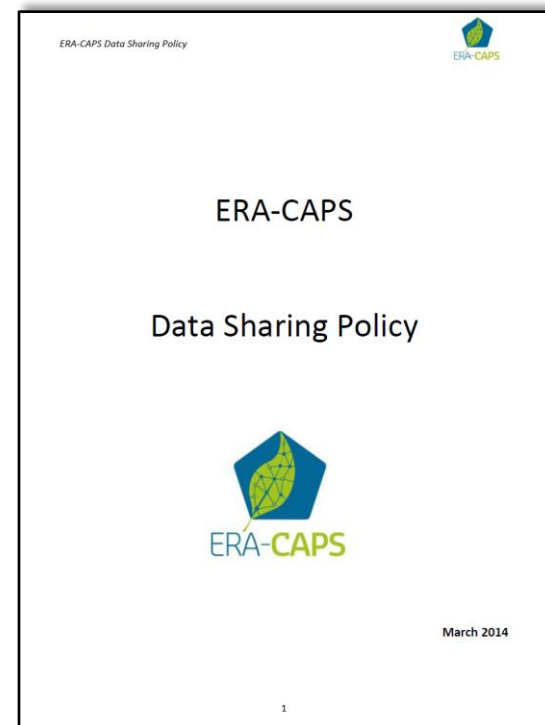
- are a public good, produced in the public interest;
- should be openly available to the maximum extent possible.

Requirements for projects

- Data management plan
- Data officer

Data management plan includes

- Volumes, data types, ..
- Standards, ....
- Data infrastructure
- Ethics and IP
- Data sharing and secondary use
- Timeframes, responsibilities, budgets, ...
- Provisions on data management after the project



# Standards and Science – Standards and Creativity

Standard



Creativity (Chaos)



a contradiction?

## Standards and Science – Standards and Creativity



# Why standards?

## Implicit expectation

- access of data beyond the initial generator(s) of the data will allow wider utilization
- Enhance reproducibility
- Allow
  - integration with other data
  - (automated) extraction from data collections
- Increase efficiency of funding

Standards go beyond Good Scientific Practice, which “only” requires adequate description of data acquisition

# What is the problem ?

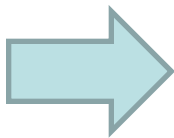
- Everybody likes to use standards, if he/ she has a benefit from it (incentive system)
- Standardization is seen (is) as a risk for creative developments and progress “beyond standard”
- Standard require effort
  - Need to be developed
  - Need to be implemented (even, if the individual researcher has no advantage from it)...for the benefit of the community and not for the sake of standardization

# Background and scope of the EWG



Original scope (according to proposal text)

- “To maximise the use, re-use, re-purposing and sharing of the outputs from projects funded through ERA-CAPS, a set of community-led standards will be developed and agreed upon through a series of workshops.”
- “These standards will build upon existing best practices within the bioscience research community”
  - Analyse „standard in plant sciences“
  - Describe benefits and risks
  - Propose ways forward
  - Recommend options for ERA-CAPS

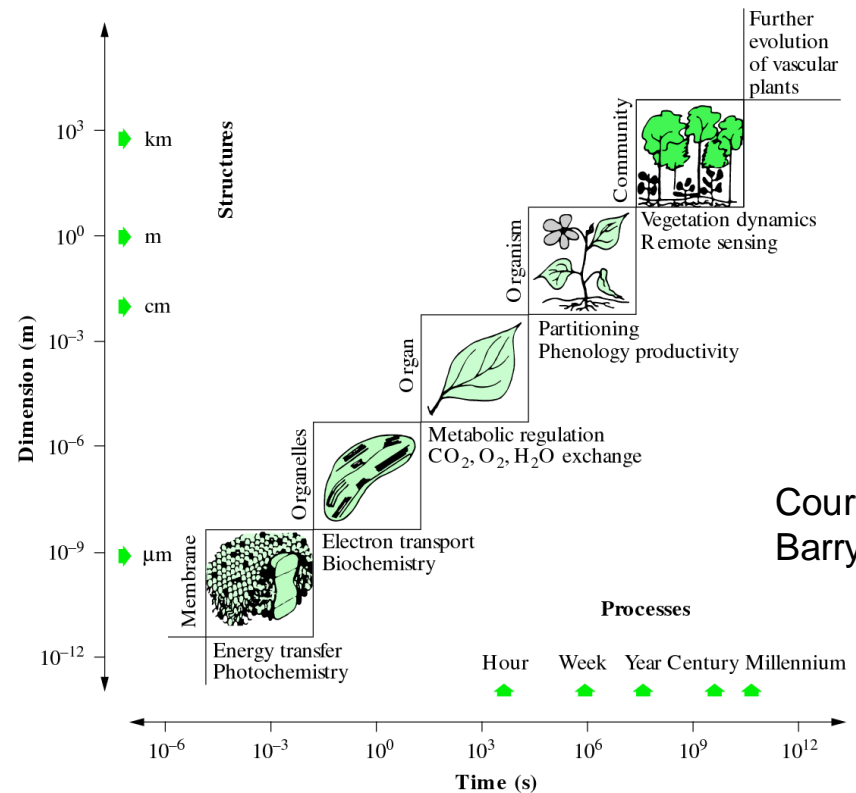


# What are standards ?

Standards in plant sciences enable

“knowledge generation based on coupling data and information about all relevant domains of plant systems”

- Genome
- Epigenetics
- Transcriptome
- Metabolome
- Physiology
- Phenotype
- Environment



Courtesy to  
Barry Osmond



# What are standards ?

Standards in plant sciences enable

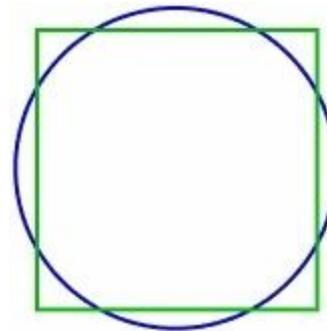
“knowledge generation based on coupling data and information about all relevant domains of plant systems”

- Genome
- Epigenetics
- Transcriptome
- Metabolome
- Physiology
- Phenotype
- Environment
- Specialised communities with conceptual differences
- Difference in stability/dynamics of methods/ technology
- Integration requires communication between them

## Standards shall

- be based on agreement within an expert community
- be communicated and applicable to the wider science community
- NOT hinder progress within a specialized community nor in the wider science community

Squaring the  
circle ??



## Squaring the circle

Standardization is a **process** mediating between creativity and standard with the aim to provide wider accessibility



## Standards and stability

Standards must

- be adapted to progress in a specialist community
- be stable enough to not confuse the wider community

Thus,

- a standard should not be released before it is stable and validated by the community and
- needs to be precisely communicated
- backward compatibility should be pursued

# Standards and stability



	CD	DVD	HD-DVD	BlueRay
Jahr	1982	1996	2006	2006
Kapazität [GB]	0,7	4,7 / 8,5	15 / 30	25 / 50 (200)
Wellenlänge [nm]	780 (Infrarot)	650 (Rot)	405 (Violett)	405 (Violett)
Schichten	1	2	2	2 (6)
Datenrate [MB/s]	8	22	36	36 (288)

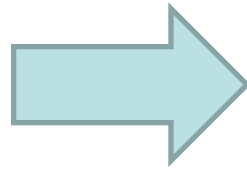
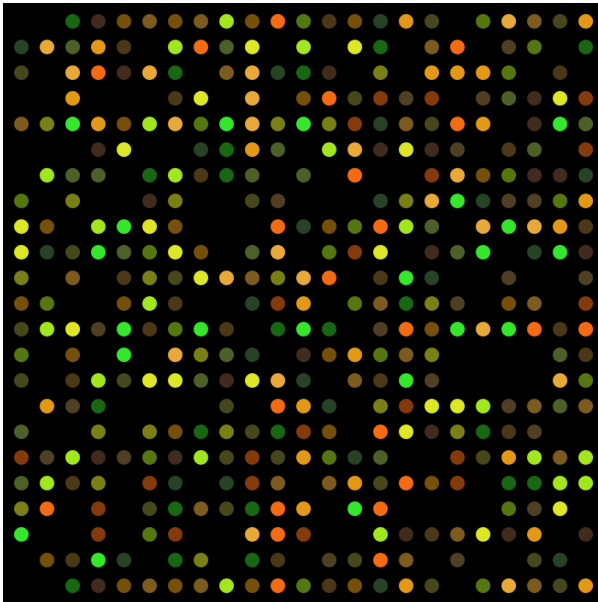
A change in standard

- needs to bring a significant benefit for the users
- is often associated with technology/ concept development

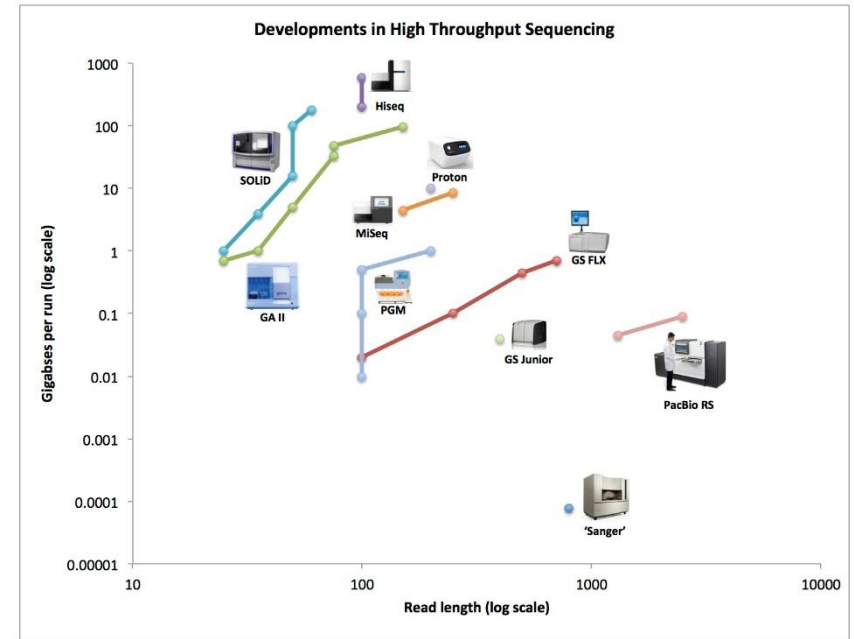


# Standards and stability

Microarray



NGS

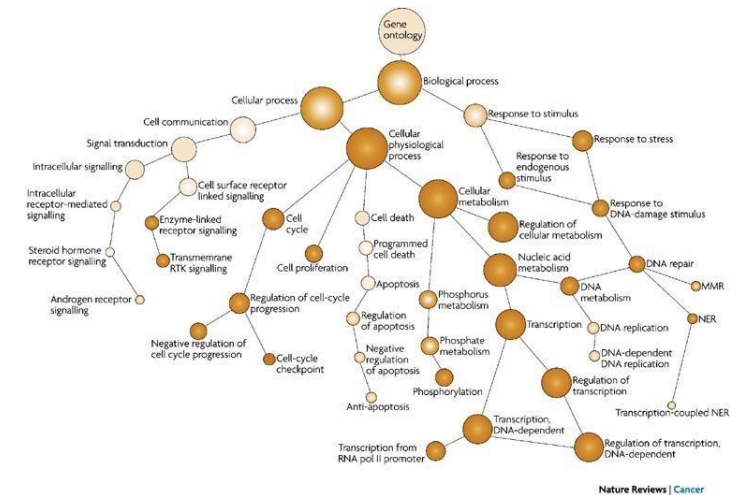


A change in standard

- needs to bring a significant benefit for the users
- is often associated with technology/ concept development

# Interoperability and ontologies

- Definition of terminology (short/ stunted/ ...)
- Categorized systems
- Definition of relationships relative to each other (siblings, ...)



## Example Gene Ontology (GO)

- cellular components
- biological processes
- molecular function



# Implementing a standardization process

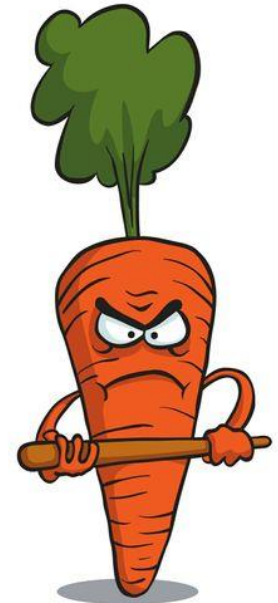
- Learning and integrating
  - non-plant communities
  - building on and integrating existing plant science – related standardisation efforts (projects, infrastructures, databases, ...)
- Integrate globally
- Generate a **common portal/ mechanisms for standardisation** addressing the diverse layers of standards
  - Collect SOPs
  - Community standards
  - Interoperability
  - Requires long-term commitment
  - Reference point for dissemination





# A standard is only delivering, if it is used

- Difference between interest of a single project and the community
  - Develop best practice examples showing the benefit of standards
- Provide resources to facilitate the „extra“ beyond the immediate interests of the data generator
- Appreciate and give incentives to groups developing, hosting and implementing standards
- Enforce standards e.g. through funding agencies and publishers based on global portal and expert panels
- Communicate and teach standards and methods





## Recommendations

- Providing a good standards process requires long-term commitment and additional effort
  - Framework of „institutions/ mechanisms“ (e.g. global portal) developing standards with a long-term perspective for plant sciences
  - Organise meetings to link existing projects/ institutions working on standards
  - Continue, as a consortium of funders like ERA-CAPS, to value and support standard development and implementation by (flagship) projects and enforcement

# Standards and standardisation

- Are integral part of the science process
- Are undervalued, because the additional effort to achieve interoperability is not valued
- Must be flexible enough to allow progress and rigid enough to allow reliability and exchange of (quality) data

From

**either/ or**

to

**both**

where appropriate

