

# MetaData can be easy

Rainer Krug

Department of Evolutionary Biology and Environmental Studies, UZH

2019/09/11 (updated: 2019-09-03)

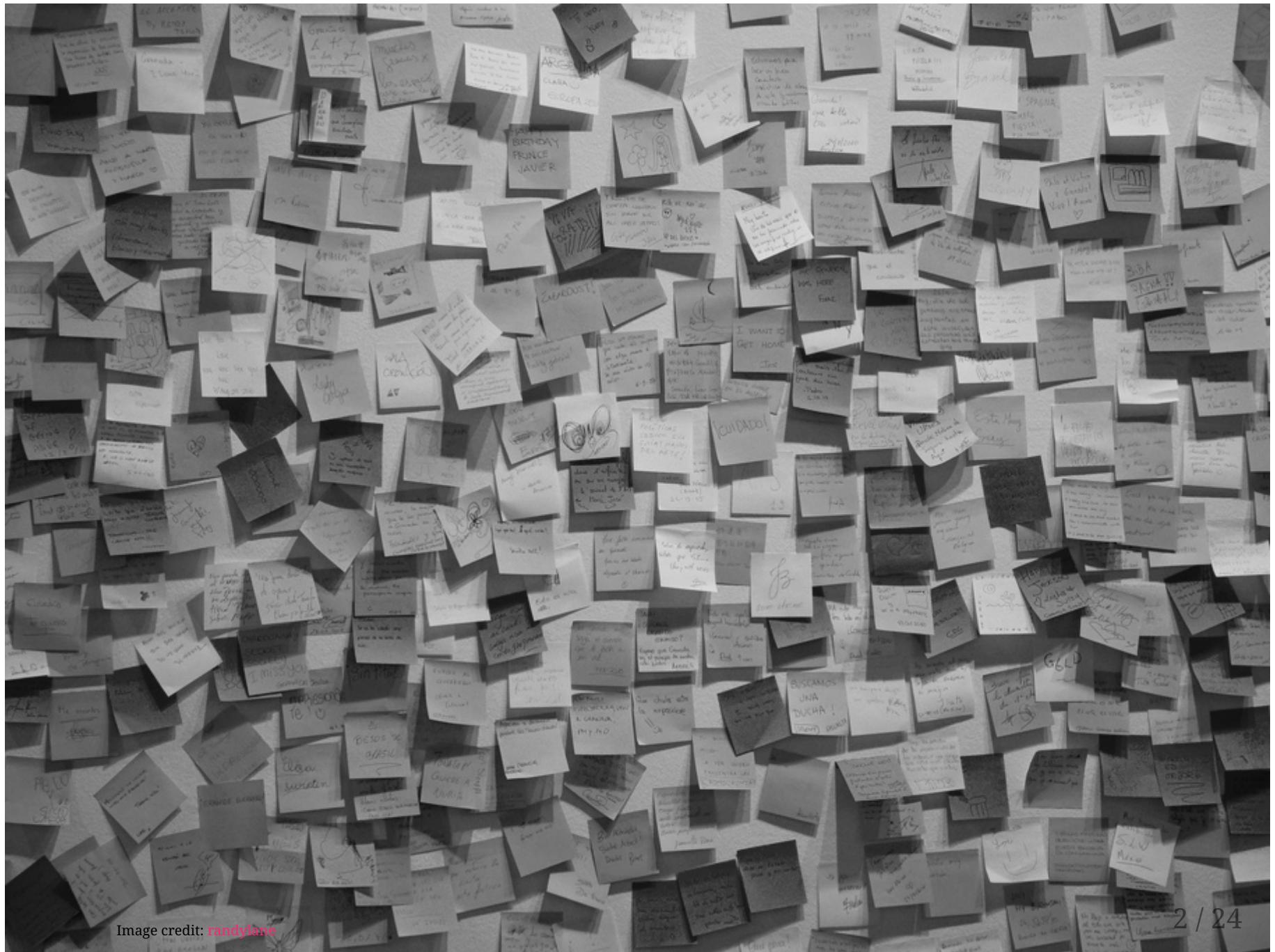


Image credit: [randylare](#)

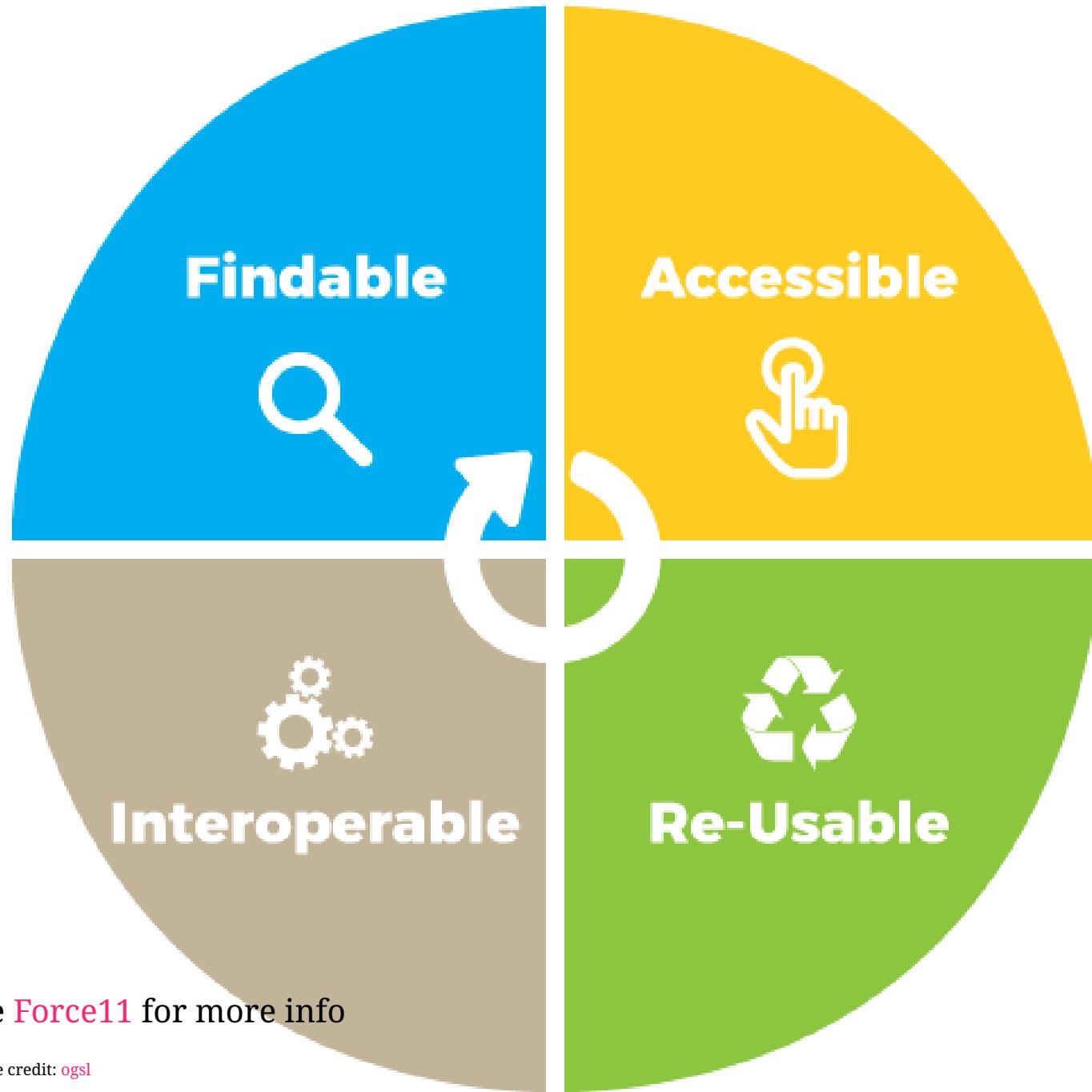


Image credit: [ogsl](#)

# Why do we need this?

- Whose MSc / PhD / first research project was more than 3 years ago?
- Who remembers
  - where the data is and if it is still readable?
  - what the column names mean?
  - which units were used?
  - how the analysis was done (script)?

How can you re-use your data?

How can anybody else re-use your data?

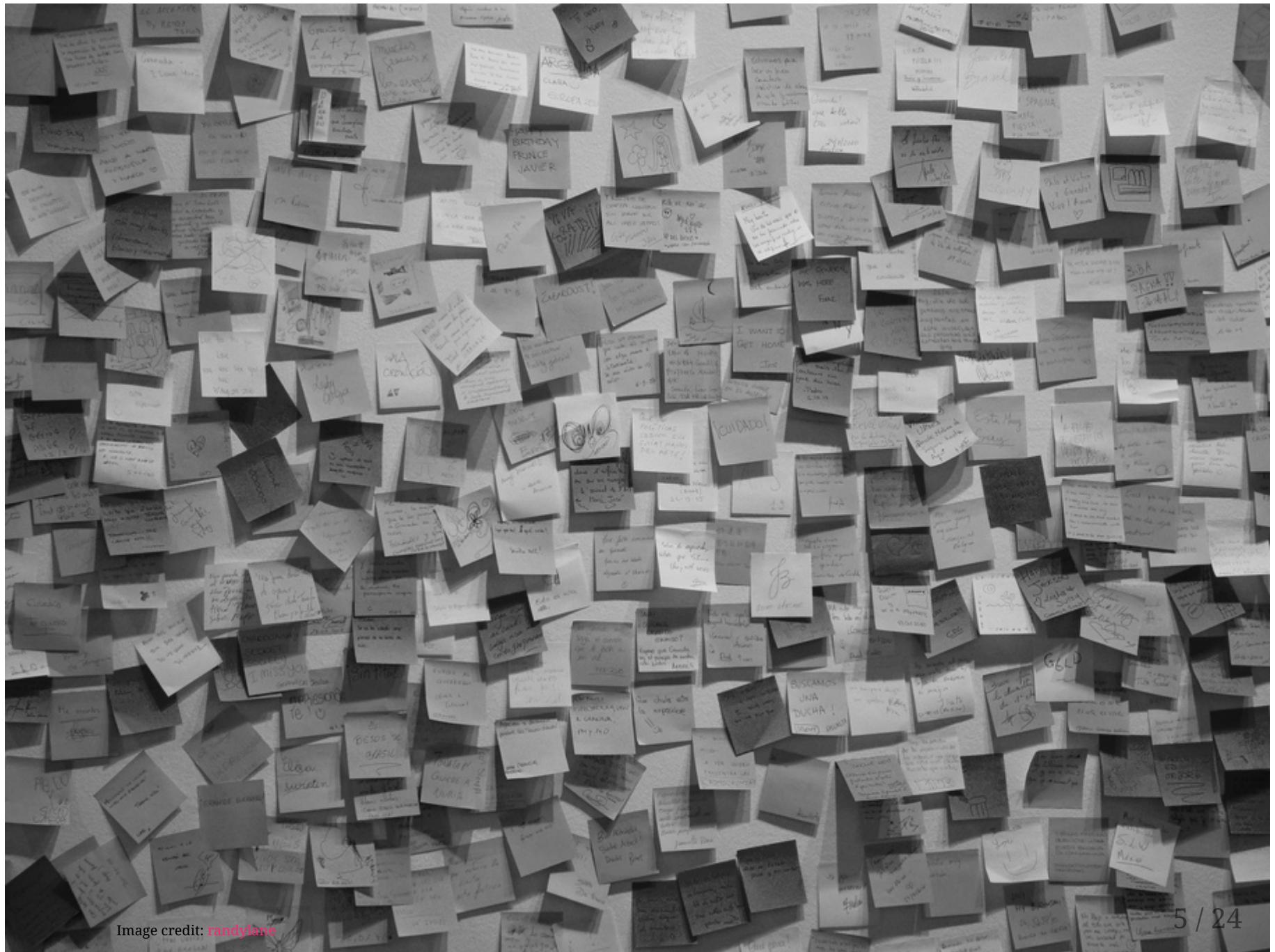


Image credit: [randylare](#)



Image credit: randylane



Image credit: [Pooktree](#)

7 / 24

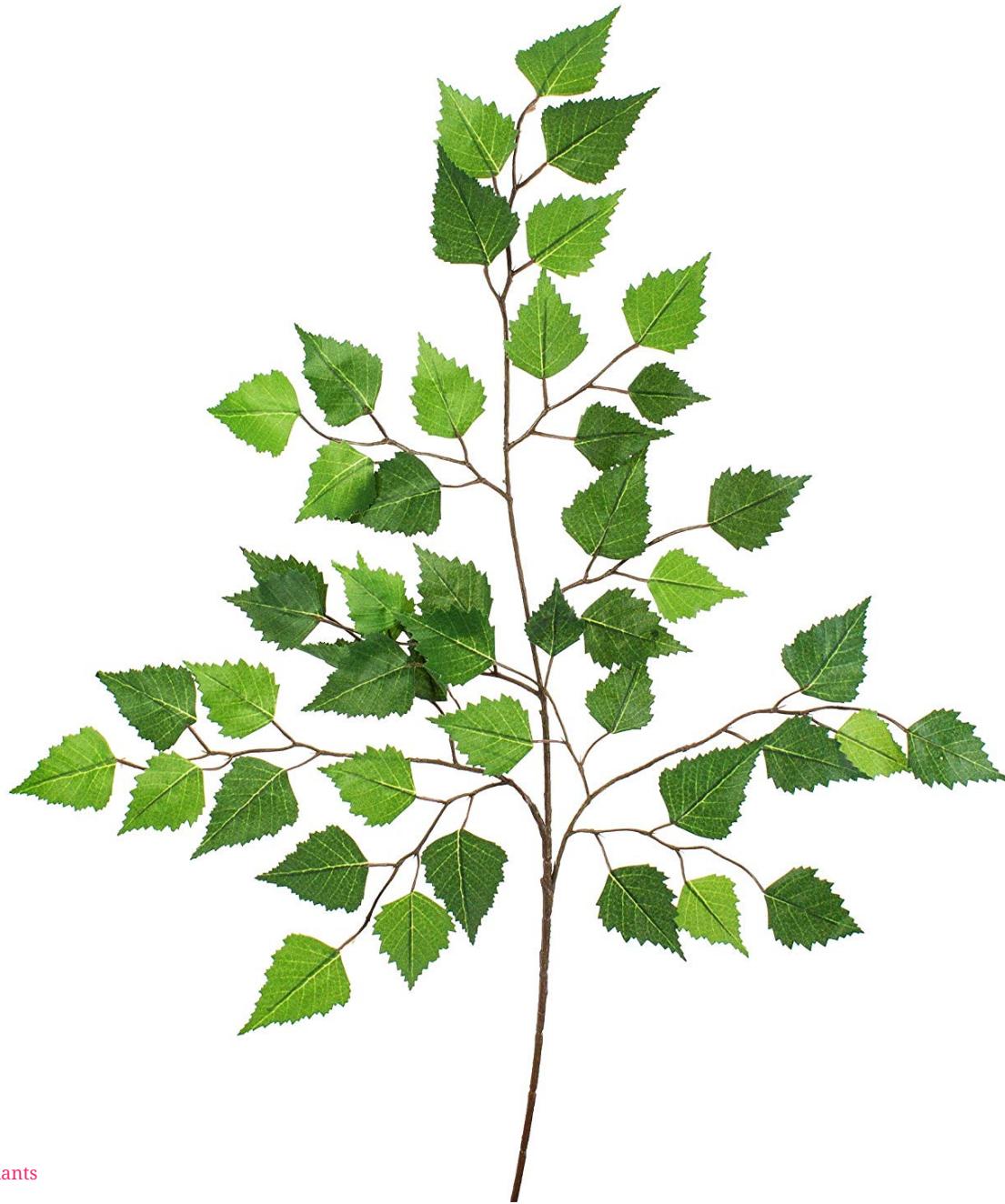


Image credit: [Artplants](#)

8 / 24

Access provided by University of Zurich ×

FORUM | VOLUME 34, ISSUE 6, P494-496, JUNE 01, 2019

## Ecological Data Should Not Be So Hard to Find and Reuse

Timothée Poisot • Anne Bruneau • Andrew Gonzalez • Dominique Gravel • Pedro Peres-Neto • Show footnotes

Published: May 02, 2019 • DOI: <https://doi.org/10.1016/j.tree.2019.04.005> •



Drawing upon the data deposited in publicly shared archives has the potential to transform the way we conduct ecological research. For this transformation to happen, we argue that data need to be more interoperable and easier to discover. One way to achieve these goals is to adopt domain-specific data representations.

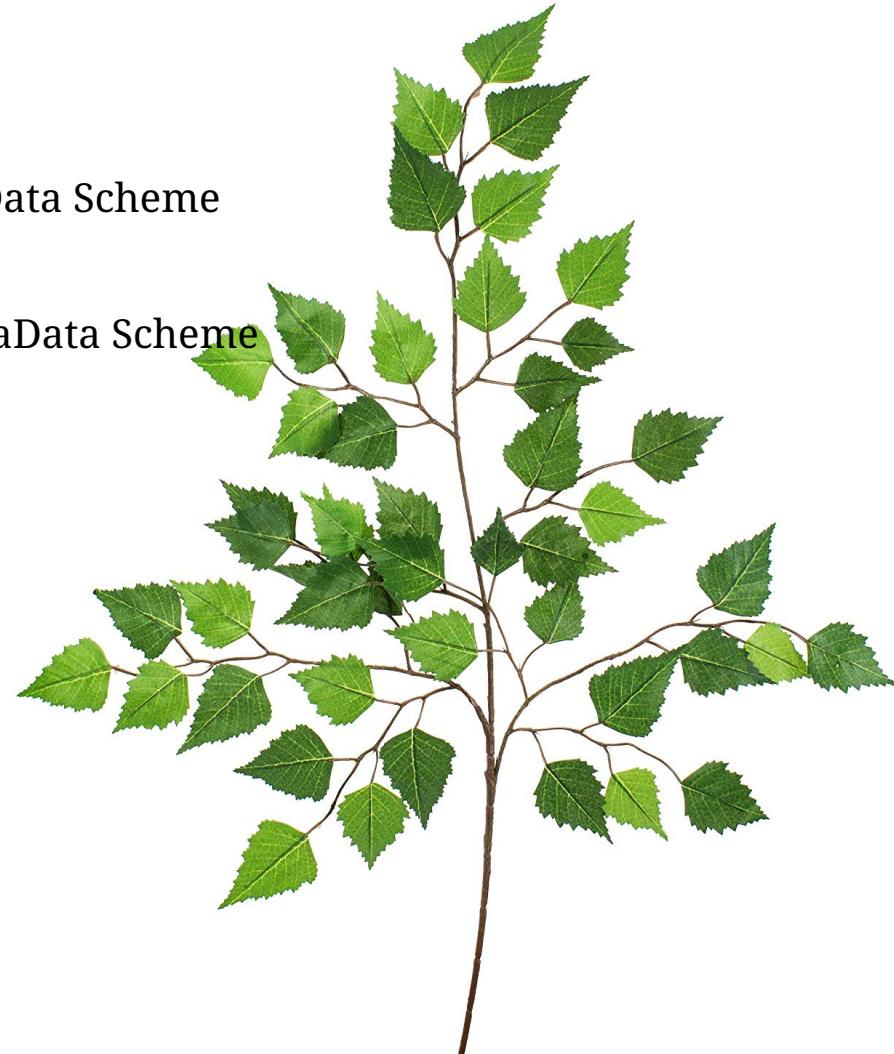
### Keywords

ecological synthesis • open data • metadata • data standardization

Poisot et al. (2019)

# Two aspects

1. **Developing**  
a new Domain Specific MetaData Scheme
2. **Using**  
the new Domain Specific MetaData Scheme



# Tool which will help to

**develop a new Domain Specific MetaData Scheme [developer & researcher: R & Spreadsheet]**

- together with scientists
- encourage buy in & "ownership" of scheme

**enter MetaData [user: browser & Spreadsheet]**

- easy, fun & transparent

**validate MetaData [user: browser]**

- report of errors and inconsistencies

**export MetaData [user: browser]**

- formats suitable for repository submission & indexing



# Use programming language R

## Two packages

### 1. scheme generator - developer - `dmdScheme`

- template scheme,
- base functionality for
  - entering,
  - validating and
  - exporting the metadata and
- tools to develop new schemes and packages

### 2. actual scheme - user - e.g. `emeScheme`

- a specific scheme and can contain
- additional functionality for
  - entering,
  - validating and
  - exporting the metadata and
- additional specific tools

emeScheme.dff22850b8ff

Search Sheet

Home Insert Draw Page Layout Formulas Data Review View

Normal Page Break Preview Page Layout Custom Views Ruler Formula Bar Zoom 200% Zoom to Selection New Window Arrange All Freeze Panes Freeze Top Row Freeze First Column Split Hide Switch Windows View Macros Record Use Relative References

A1 propertySet

	A	B	C	D	E	F	G
1	propertySet	valueProperty	unit	type	suggestedValues	Description	DATA emeScheme v0.9.5
2	Experiment	name		character		The name of the experiment.	
3		temperature		character	treatment, in degrees celsius, measurement	Temperature used for all treatments. If different between treatments, use "treatment" and specify in the Treatment sheet.	
4		light		character	treatment, light, dark, cycle , e.g. 16:8 LD	Light used for all treatments. If different between treatments, use "treatment" and specify in the Treatment sheet.	
5		humidity		character	treatment, relative humidity in %	Humidity used for all treatments. If different between treatments, use "treatment" and specify in the Treatment sheet.	
6		incubator		character	none, bench	What type of incubator is used.	
7		container		character		What type of container is used.	
8		microcosmVolume	ml	numeric		Volume of the microcosm container. Not the volume of the culture medium!	
9		mediaType		character			
10		mediaConcentration	g/l	numeric			
11		cultureConditions		character	axenic, dirty, clean	Conditions of the cultures for all treatments.	
12		communityType		character	treatment, single trophic level, multiple trophic level	Characterisation of the microbe community.	
13		mediaAdditions		character			
14		duration	days	integer		Length of the experiment in days. This should only include the time in which the measurements were taken!	
15		comment		character		Additional features of the Experiment you want to provide	
16							
17	DOCUMENTATION	Experiment	Species	Treatment	Measurement	DataExtraction	DataFileMetaData

emeScheme.df22850b8ff

Share | Comments

Home Insert Draw Page Layout Formulas Data Review View

Ruler Formula Bar Zoom 150% Zoom to Selection New Window Arrange All Freeze Panes Freeze Top Row Freeze First Column Hide Split Unhide Switch Windows View Macros Record Use Relative References

A1 fx propertySet

	A	B	C	D	E	F	G	H
1	propertySet	Species						
2	valueProperty	speciesID	name	strain	source	density	functionalGroup	comment
3	unit					cells / ml		
4	type	character	character	character	character	character	character	character
5	suggestedValues					treatment	bacteria, bacterivore, predator, phototroph	
6	Description	Id of the species and strain. Each speciesId has to be unique.	Scientific name of the species or unknown.		Where the species was obtained from.	Initial density used for all treatments. If different between treatments, use "treatment" and specify in the Treatment sheet.	Functional group of the species.	
7	DATA							
8	MULTIPLE ROWS							
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								
41								
42								
43								

DOCUMENTATION Experiment Species Treatment Measurement DataExtraction DataFileMetaData +

14 / 24

Screenshot of a Microsoft Excel spreadsheet titled "emeScheme.df22850b8ff". The spreadsheet is in "View" mode with "Gridlines" and "Headings" checked. The zoom level is set to 150%. The formula bar shows "fx propertySet".

The table has columns A through H. Rows 1 through 5 define properties for "DataFileMetaData". Row 6 provides a detailed description of the "dataFileName" column.

	A	B	C	D	E	F	G	H
1	propertySet	DataFileMetaData						
2	valueProperty	dataFileName	columnName	columnData	mappingColumn	type	description	comment
3	unit							
4	type	character	character	character	character	character	character	character
5	allowedValues			ID, Treatment, Measurement, Species, other		integer, numeric, character, logical, datetime, date, time		
6	Description	the name of the data set.	Name of column in the data file. Each column in the data file needs to be documented! or NA	The type of the data in the column.  ID: ID field (unique ID of unit of replication);  Treatment: specifies treatment;  Measurement: contains measurements;  Species: contains species;  other: other type of data	columnData = Treatment: treatmentID as in the Treatment tab;  columnData = Species: treatmentID referring to species composition as in the Treatment tab  columnData = Measurement: measurementID as in the Measurement tab;  otherwise: NA	Type of the column.	if column contains measurement: General description. If type is datetime, date, or time, give the order of year month day hour minute second as e.g. ymdhms, ymd, or hms. (Do not give any other information, e.g. give nothing about how months are entered (e.g. number or name), or how years, months, day, etc are separated.	
7	DATA							
8	MULTIPLE ROWS							
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								

Validation of data against dmdS... +

Petchey group me... UZH - Departmen... Group Data Policy IPBES Chapter 4 Home Server and... Keyboard Maestro Car Play Radio Conferences »

## 2.2 Overall MetaData - error

The details contain the different validations of the metadata as a hierarchical list. errors propagate towards the root, i.e., if the 'worst' is a 'warning' in a validation in `details` the error here will be a 'warning' as well.

x  
NA

### 2.2.1 DataFileMetaData - error

The details are a table with one row per unique validation. The column `Module` contains the name of the validation, The column `error` contains the actual error of the validation. The following values are possible for the column `isTRUE`:

TRUE : If the validation was `OK`.  
FALSE: If the validation was an `error`, `warning` or `note`.  
NA : If at least one validation resulted in `NA`

One or more FALSE or missing values values will result in an ERROR.

Module	errorCode	isOk
types	OK	TRUE
allowedValues	OK	TRUE
dataFilesExist	error	FALSE
datetimeFormatSpecified	OK	TRUE
mappingColumnInNameOrParameter	error	FALSE
columnNameInDataFileColumn	error	FALSE
dataFileColumnInColumnNames	OK	TRUE

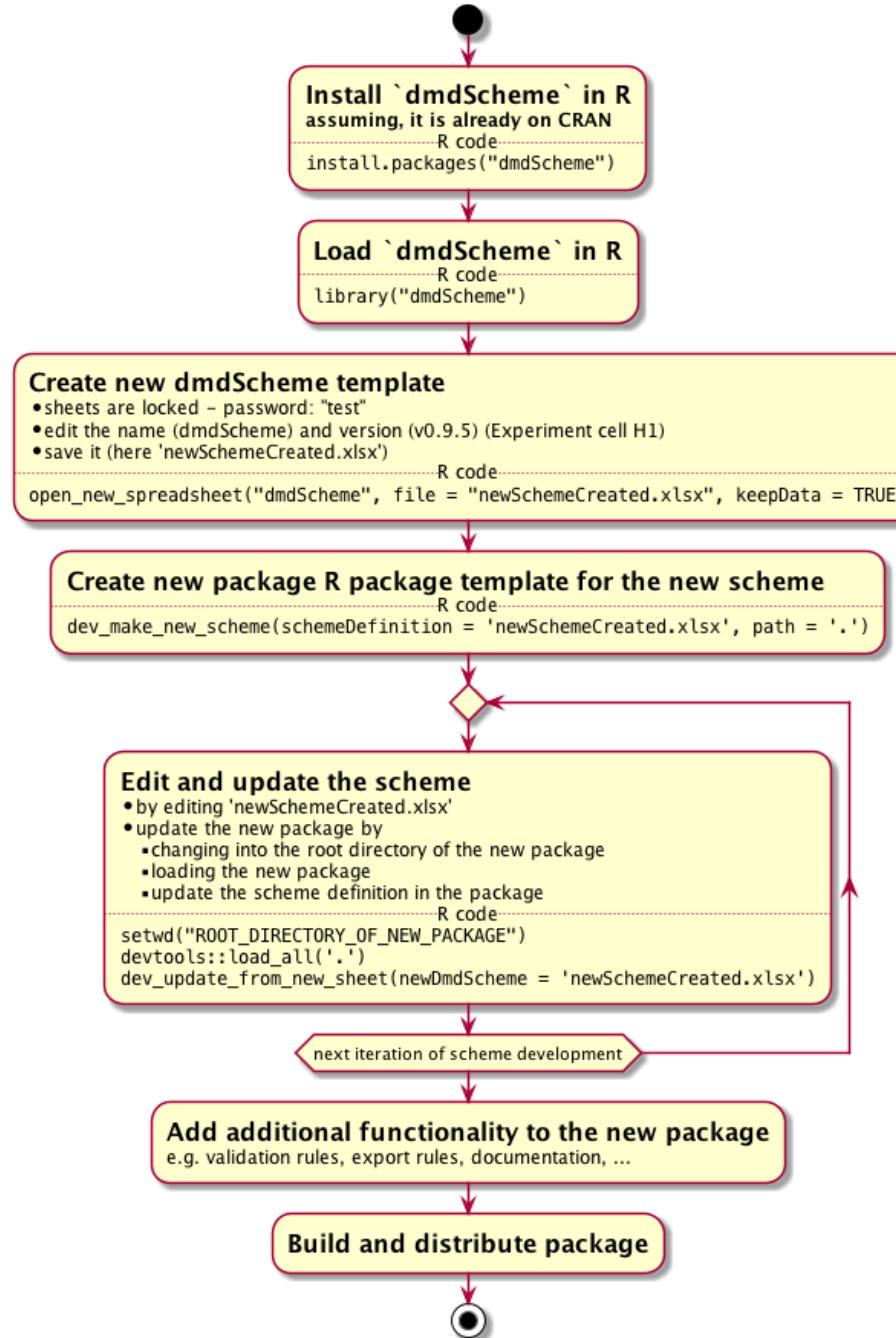
#### 2.2.1.1 datafile exists in path - error

The details are a table with one row per unique `variable`. The following values are possible for the column `isTRUE`:

TRUE : If `dataFile` exist in the given `path`  
FALSE: If `dataFile` does not exist in the given `path`  
NA : empty cell

# How to develop a new Domain Specific MetaData Scheme

(basic) R knowledge necessary



## Live Demo

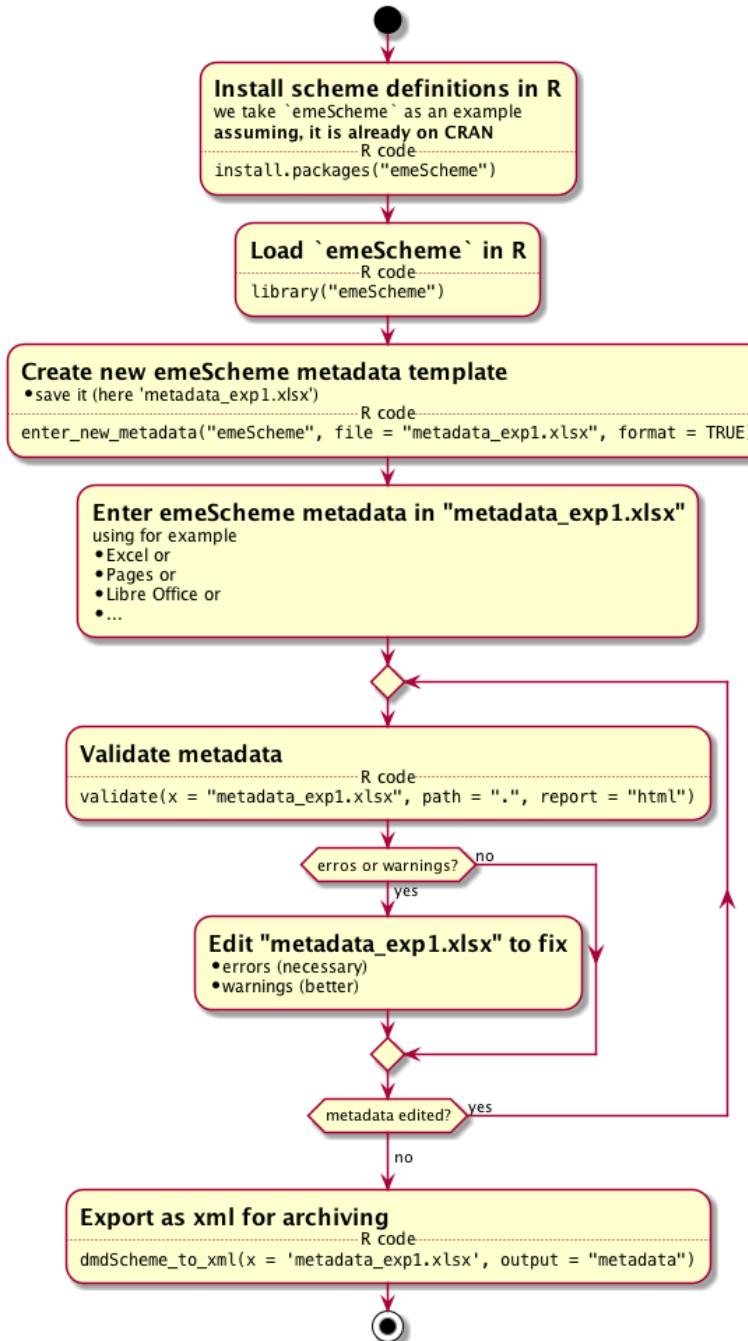
How to Develop a New Domain Specific MetaData Scheme

# How to use the new Domain Specific MetaData Scheme

from R  
base package  
also possible from R

The screenshot shows a web application window titled "Scheme App" running on a local server at 127.0.0.1:3369. The interface is divided into several sections:

- Schemes to be loaded**: A section listing installed dmdSchemes:
  - emeScheme 0.9.4.1
  - theNewScheme 1.0A message "[1] "Packages Loaded:" is displayed below the list.
- New Empty Scheme**: A section with a button labeled "Open empty scheme in Spreadsheet".
- Upload Spreadsheet containing Metadata**: A section with a "Select Spreadsheet" input field containing "Browse..." and "No file selected".
- Validate Uploaded Metadata**: A section with a button labeled "Validate metadata from spreadsheet".
- Export Uploaded Spreadsheet to xml**: A section with a button labeled "Export metadata from spreadsheet to xml".



## Live Demo

How to Enter New MetaData



Talk

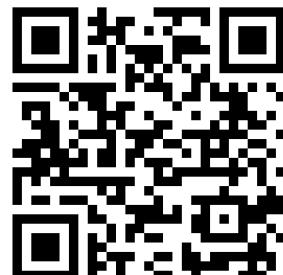


Image credit: [eater.com](#)

dmdScheme



24 / 24