



# Research data management for data on renewable materials and products

Dr. Ana Slavec

Consulting statistician at the InnoRenew CoE

RDA Europe Ambassador for Engineering/Renewable materials

Faculty of Mathematics, Natural Sciences and Information Technologies

University of Primorska

Koper, 9. 1. 2019

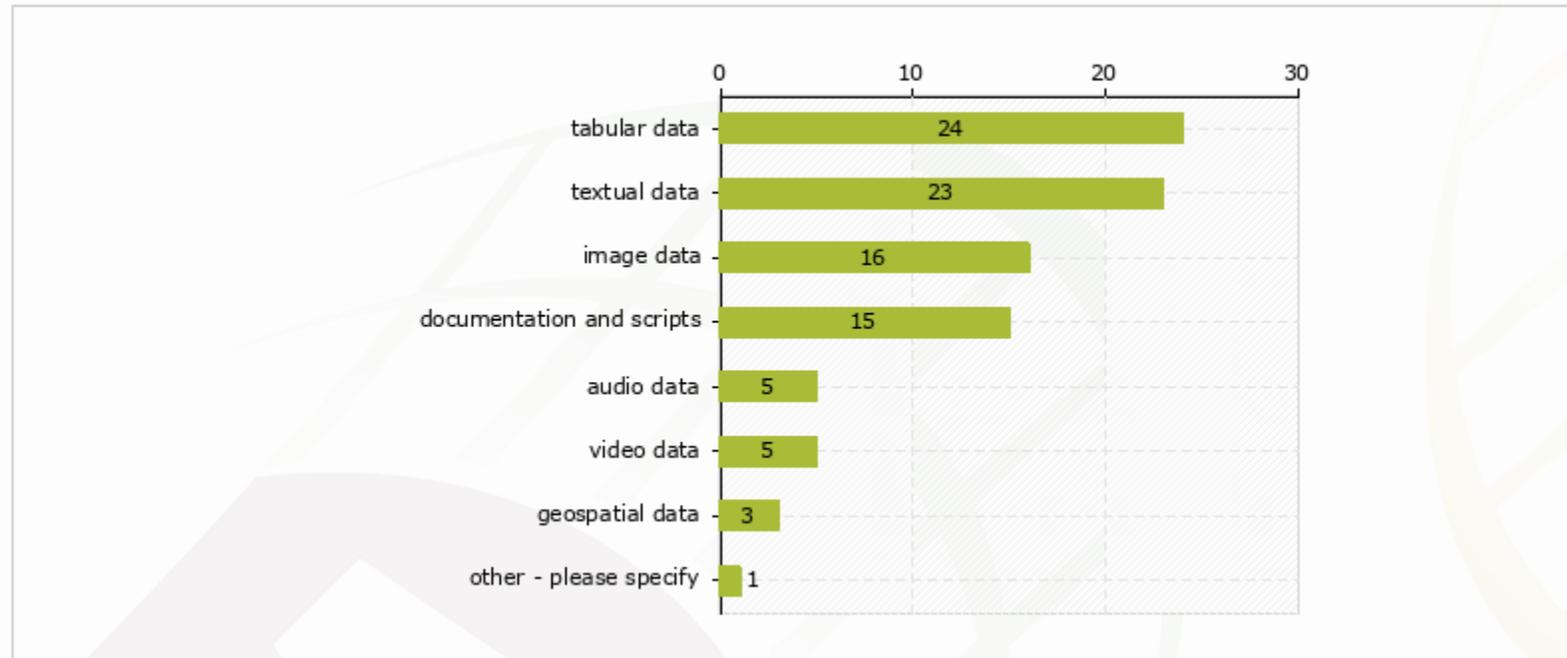


## Workshop on Research Data Management for Renewable Materials and Products

- 10.00 – 11.00 Introduction to research data management and presentation of RDA
- 11.00 – 12.00 Practical session 1
- 13.00 – 14.00 Best practices for using spreadsheets in RDM
- 14.00 – 15.00 Practical session 2
- 15.00 – 16.00 Presentation of practical session results

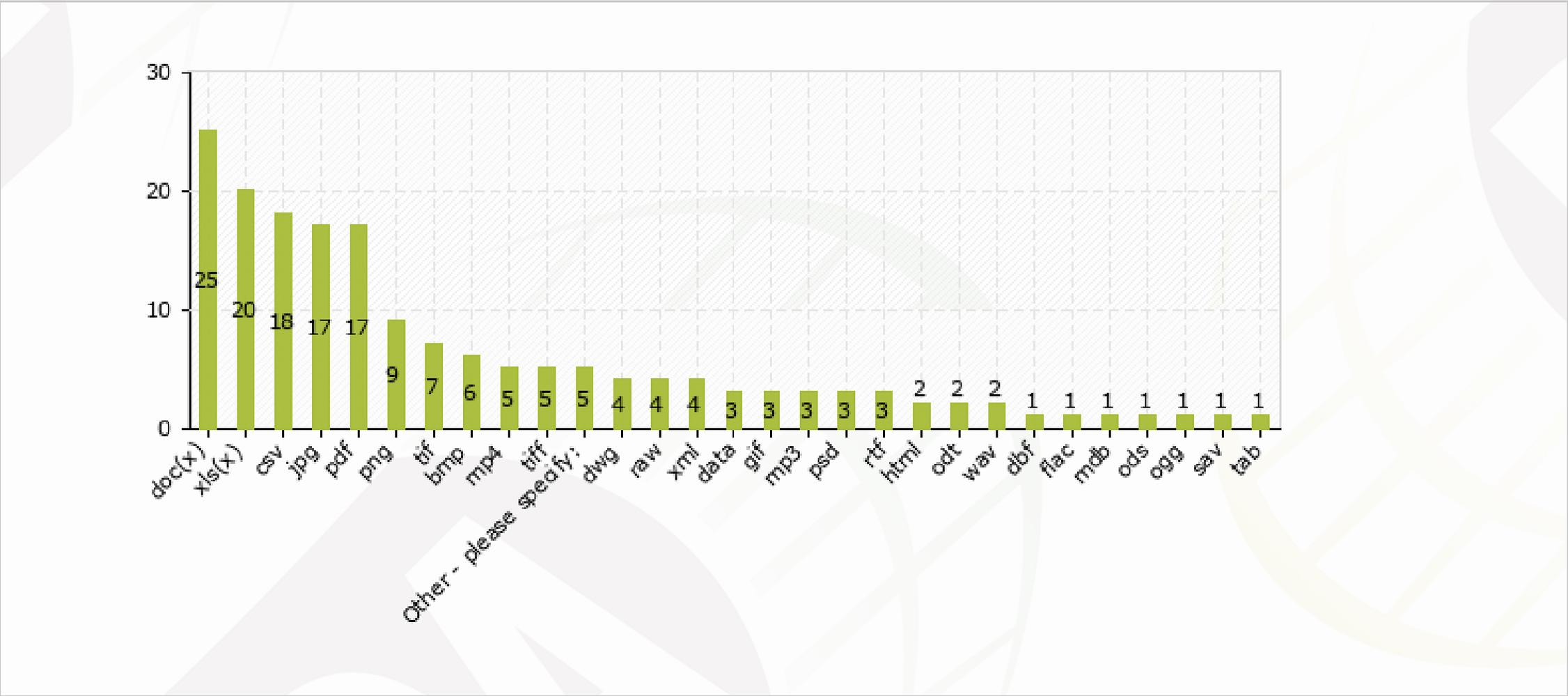


# What type of digital content do you generate in your research? (n=28)



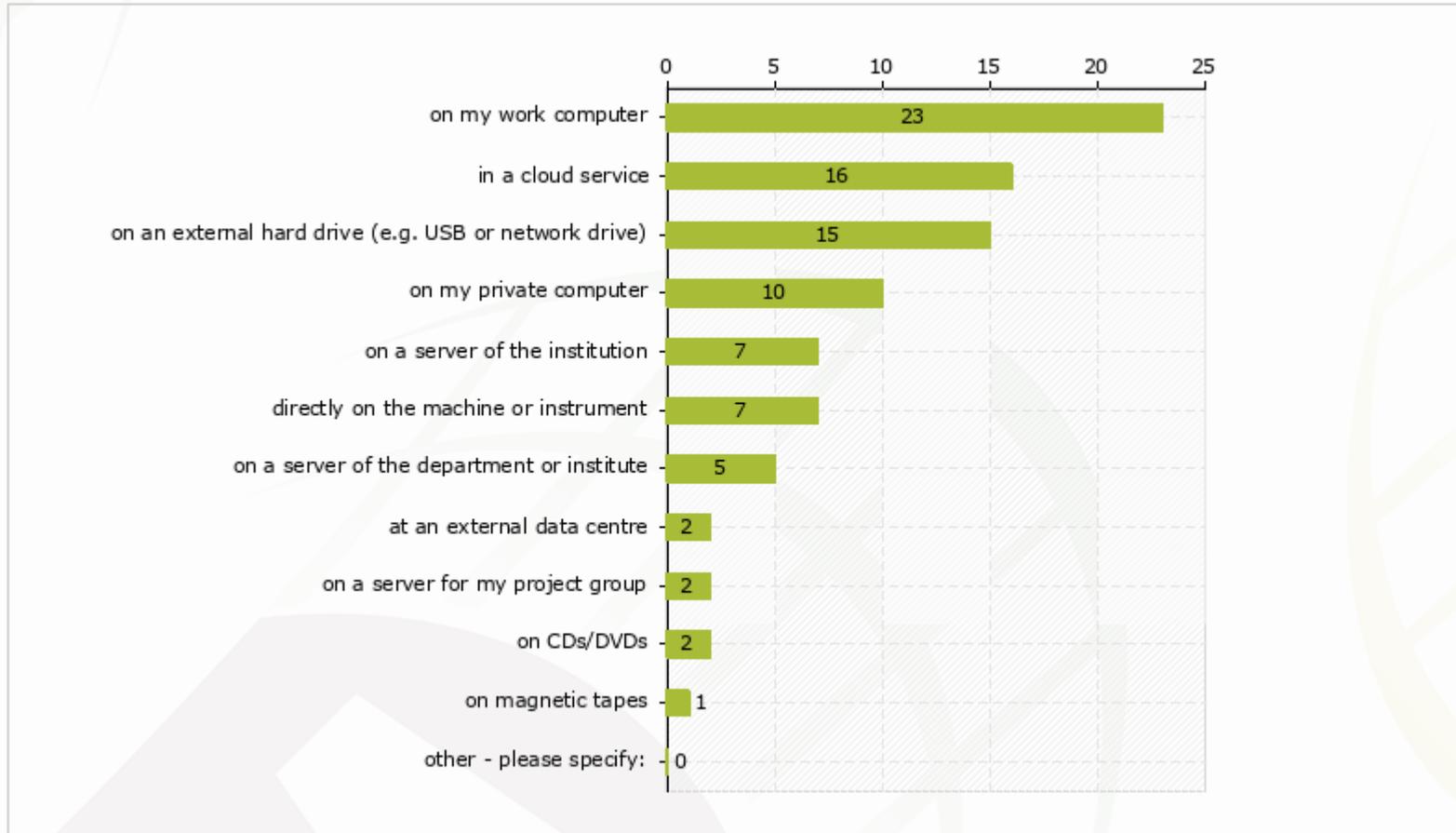


# What are the formats that you use to save your data? (n=28)





# Where do you store your research data? (n=28)





# Outline

## Introduction to Research Data Management

- What it is and why should that interest us?
- What are the RDM challenges in engineering?
- How to prepare a Research Data management plan?
- What are the recommended file formats?
- What are the FAIR principles?
- How to select the most appropriate repository?

## Research Data Alliance

- RDA Recommendations and Outputs
- How to get involved?

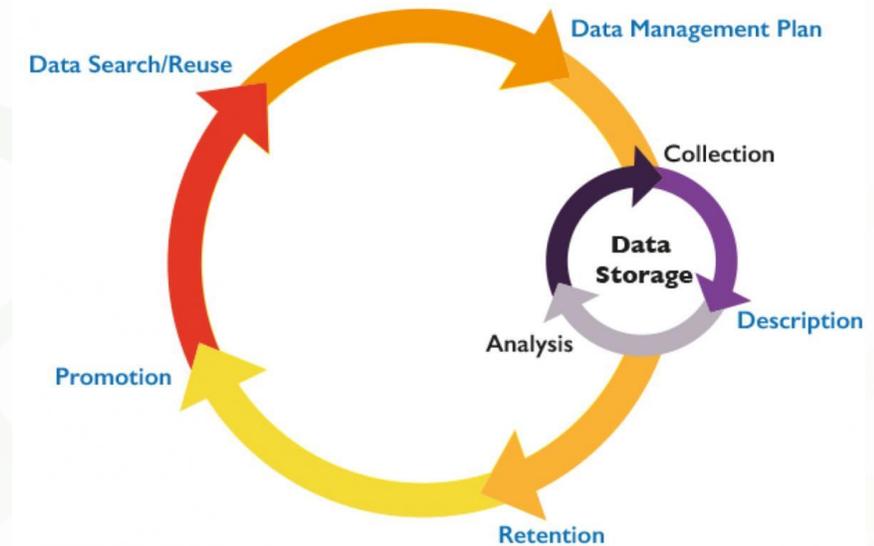
## Practical exercise



# What is Research Data Management?

- RDM describes the organisation, storage, preservation, and sharing of data collected and used in a research project.
- It involves decision about how data will be preserved and shared after the project is completed.

## The Research Data Management Lifecycle





# Why should this interest us?

## 1. Ethics

- RDM is part of the responsible conduct of research, i.e. the practice of scientific investigation with integrity
- Reproducibility crisis in science (well-managed and accessible data allows others to validate and replicate findings)

## 2. Requirements imposed by funders and publishers

- Open Science will be one of the pillars of the Horizon Europe, the next EU framework programme for research and innovation

## 3. Saves times and resources (in the long run)



# Key data challenges in engineering

- 🌐 A lot of research is done in collaboration with industry which is reluctant to share data
- 🌐 Fear of losing competitive advantage
- 🌐 Fear of data being misused and misinterpreted
- 🌐 Lack of metadata standards and ontologies (for certain engineering fields)
- 🌐 Lack of domain-specific repositories (for certain engineering fields)



# Data management plan (DMP)

- A formal document that outlines how data are to be handled during a research project, and after the project is completed
- What should a DMP include?
  - Description of data to be collected/created
  - Standards/methodologies for data collection and management
  - Ethics and intellectual property right
  - Plans for data sharing and access
  - Strategy for long-term preservation
- Useful resources: <http://www.dcc.ac.uk/resources/data-management-plans>



# DMP template – questions to respond

1. What data will you collect or created? How?
2. What documentation and metadata will accompany the data?
3. How will you manage any ethical and legal issues?
4. How will the data be stored and backed up during research?
5. How will you manage access and security?
6. How will you share the data? Are any restrictions required?
7. Who will be responsible for data management? What resources will you require to deliver your plan?



# 1. Data collection

- Type, format and volume of data
- Formats and software
- Reuse of existing data (secondary data)
- Standards and methodologies
- Structure and name of files
- Versioning
- Quality assurance processes



# Recommended file formats

Type of data	Recommended formats	Acceptable formats
Tablular data	csv, .tab, .por, .xml	.txt, xls, .dbf, .ods, .sav, .dta, .mdb
Geospatial data	.shp, .shx, .dbf, .prj, .sbx, .sbn, .tif, .tfw, .dwg, .gml	.mdb, .mif, .kml, .ai, dxf, .svg
Textual data	.rtf, .txt, .xml	.html, .doc
Image data	.tif	.jpg, .gif, .tif, .tiff, .raw, .psd, .bmp, .png, .pdf
Audio data	.flac	.mp3, .aif, .wav
Video data	.mp4, .ogv, .ogg, .mj2	.avchd
Documentation and scripts	.rtf, .pdf, .xhtml, .htm, .odt	.txt, .doc, .xls, .xml

Source: <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>



## 2. Documentation and metadata

- Documentation to help secondary users to understand and reuse data
- Metadata is „data about data“ (Examples: persistent identifier such as DOI, publication date, title, authors, description, keywords, licence, funding, related identifiers, etc.)
- Documentation may also include details on the methodology used, analytical and procedural information, definition of variables, vocabularies, units of measurement, assumptions made, and the format and file type of the data
- Existing community metadata standards: General (e.g. Dublin Core) or discipline specific (e.g. DDI)
- Metadata directory: <http://rd-alliance.github.io/metadata-directory/>

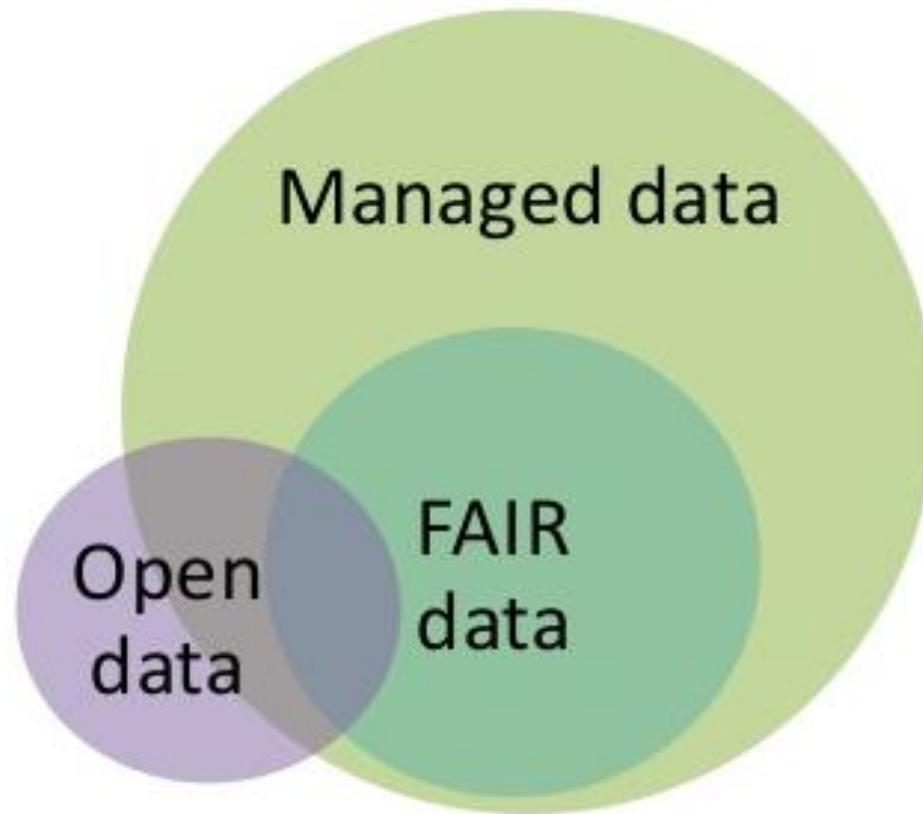


# The FAIR data principles

- 🌐 **Findable:** metadata and data should be easy to find for both humans and computers
- 🌐 **Accessible:** once the users finds the required data, they need to know how can they be accessed, possibly including authentication and authorisation
- 🌐 **Interoperable:** data usually need to be integrated with other data be interoperable with applications or workflows for analysis, storage, and processing
- 🌐 **Reusable:** metadata and data should be well-described so that they can be used in different settings



# Open data vs. FAIR data



Jones, S. 2018. Open data, FAIR data and RDM: the ugly duckling. Available at: <https://zenodo.org/record/1196631#.XhbcdUdKh3g>



## 3. Ethics and legal compliance

- Review of research plans involving sensitive research on human subjects are submitted to the ethical committee in charge of the corresponding research area
- Risk of disclosure assessed before, during, and after data collection:
  - Informed consent sought for data collection, processing and long-term preservation
  - Identification of direct and indirect identifiers in data files (removal, aggregation, pseudoanonymisation, or anonymisation of variables)
  - Restriction of access to the data in cases, when anonymisation would hinder the reusability of data
- Compliance with General Data Protection Regulation (GDPR)
- Confidential information and trade secrets



## 4. Storage and backup

- Copyright and intellectual property rights (Consortium agreements)
- Licences for reuse (e.g. [Creative commons](#))
- Restriction on reuse of third-party data
- Data sharing restrictions (embargo periods)
- Security measures and standards for confidential data:
  - Safe transfer from data collection to secured systems
  - Management of risks to data security
  - Control of access to data
  - Secure access for collaborators



## 5. Selection and preservation

- What data must be retained/destroyed for contractual, legal, or regulatory purposes
- Foreseeable research uses for the data (validation of research findings, conduct of new studies, teaching)
- Length of retention and preservation
- Repository or archive for data to be held
- Costs (repository charges, time and effort to prepare the data for sharing/preservation)



# Research data repositories

- General purpose repositories (e.g. Zenodo, Figshare)
  - Institutional data repositories
  - Domain specific repositories
- Registry of research data repositories: <https://www.re3data.org/>





# Example of institutional repository

The screenshot shows the institutional repository interface for the University of Primorska. The header includes the university logo, the language 'SLO', and a 'Login' link. Navigation links for 'Search', 'Browsing', 'Upload document', and 'Statistics' are present. A search bar is set to 'RUP'. The main content area is titled 'Search the repository' and includes a search form with the following fields:

- Query: [ ] search in Title
- AND [ ] search in Author
- AND [ ] search in Abstract
- AND [ ] search in Year of publishing
- Work type: Research Data or Corpuses (data) \* old and bolonia study programme
- Language: All languages
- Search in: RUP
- Options:  Show only hits with full text

Buttons for 'Search', 'Simple search', and 'Reset' are located below the form. A note states: 'Search without a query returns a maximum of 100 hits!'. The search results list two items:

- Nasilje v družini**  
 Mateja Sedmak, Ana Kralj, Zorana Medarić, Blaž Simčič, 2009, complete scientific database or corpus  
 Abstract: Osrednji namen predstavljenega projekta je analiza družinskega nasilja v Sloveniji s posebnim poudarkom na stopnji družbene tolerance v odnosu do nasilja, stopnji ozaveščenosti in senzibilnosti za omenjeno problematiko ter analiza medijskega diskurza družinskega nasilja. Konkretni cilji projekta so bili: 1) Statistična analiza družinskega nasilja v Sloveniji v zadnjih dvajsetih letih in analiza trendov gibanja; 2) Analiza obče družbene klime v Sloveniji v odnosu do družinskega nasilja; 3) Analiza medijskega diskurza in obravnave družinskega nasilja; 4) Oblikovanje strokovnih iztočnic in strategij ukrepov (obrnave) ter preventive nasilja v družinskih okoljih (poudarek na ozaveščanju in senzibilizaciji javnosti)  
 Keywords: nasilje, družina, nasilje v družini, psihično nasilje, fizično nasilje, otroci, starostniki, ozaveščanje, preventivni ukrepi, zakonodaja, žrtve  
 Published: 15.10.2013; Views: 2408; Downloads: 64  
 Full text (0,00 KB)
- Migrations, integration and multiculturality**  
 Simona Zavratnik Zimic, Ana Kralj, Zorana Medarić, Blaž Simčič, 2009, complete scientific database or corpus  
 Abstract: Goal of the project is to analyse current Slovenian Integration Policy (which includes fields of economic, cultural, political and



# Example of general-purpose repository

The screenshot shows the Zenodo website interface. At the top, there is a blue navigation bar with the Zenodo logo, a search bar containing 'InnoRenew CoE a', and links for 'Upload' and 'Communities'. A user profile for 'ana.slavec@gmail.com' is visible on the right. Below the navigation bar, the page title is 'InnoRenew CoE and Project research'. The main content area displays search results for 'All versions'. On the left, there are three filter panels: 'Access Right' with 'Open (106)', 'File Type' with options like Pdf (102), Pptx (5), Rmd (2), Csv (1), Docx (1), Mp4 (1), and Xlsx (1), and 'Keywords' with options like Wood (8), Wood Modification (7), Modified Wood (5), and Beech (3). The search results show 'Found 3 results.' with a pagination control showing page 1. The first result is dated 'May 9, 2019 (v1)', is a 'Dataset', and is 'Open Access'. The title is 'The Effects of Cycle and Treadmill Desks on Work Performance and Cognitive Function in Sedentary Workers: data repository of a review and meta-analysis.' The authors are Podrekar Nastja, Kozinc Žiga, and Šarabon Nejc. The description states: 'This repository contains additional files related to the review and meta-analysis. The first dataset contains list of search terms. The second dataset contains list of studies included in the meta-analysis. The third dataset contains study evaluation using PEDro scale tool. The fourth dataset contain'. It was uploaded on May 9, 2019. The second result is dated 'December 17, 2018 (v1)', is a 'Dataset', and is 'Open Access'. The title is 'Using shear-wave elastography in skeletal muscle: data repository of a reliability study'. The authors are Nejc Šarabon, Kozinc Žiga, and Podrekar Nastja. The description states: 'This repository contains raw data from experiments investigating the reliability of shear-wave elastography (SWE) to assess muscle stiffness. The first dataset contains meat specimen experiments (using porcine meat specimens). The second and third datasets contain the measurements of human subjects,'. It was uploaded on December 17, 2018. The third result is dated 'February 22, 2018 (v2)', is a 'Dataset', and is 'Open Access'. The title is 'Wood preservatives utilizing low-value olive oil production by-products: Analysis'. The authors are Burnard, Michael; Schwarzkopf, Matthew; Tverezovskiy, Viacheslav; Treu, Andreas; Humar, Miha; and Kutnar, Andreja. The description states: 'The objective of this study was to develop and assess the efficacy of two experimental methodologies for the maleinisation of lampante oil to be used for wood protection. Two maleinisation techniques were used to chemically modify low value lampante oil in an attempt to limit'.



# Example of domain-specific repository



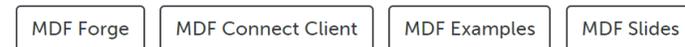
## The Materials Data Facility (MDF)

A simple way to publish, discover, and access materials datasets

### Actions



### Quick Links



## Publish Data



### Sign up and Join

Sign up for free Globus account using existing credentials. Join group for access to data publication capabilities.

- ▶ Create a Free Account
- ▶ Join this Group



### Collect

Collect the data into your preferred file structure, preferably in openly accessible formats. Feel free to nest files as deeply as necessary for your use case, our indexers will find them!

- ▶ Set up an endpoint



### Publish

Follow instructions in the form to publish your dataset.

- ▶ Publish Your Data
- ▶ Python Tools



## 6. Data sharing

- Where, how, and to whom?
- Conditions for data sharing (data sharing agreements)
- When will you make data available
- Persistent identifiers
- Outline of expected restrictions due to confidentiality, lack of consent agreements or intellectual property rights
- Non-disclosure agreements for protection of confidential data



## 7. Responsibilities and resources

- Roles and responsibilities for data capture, metadata production, data quality, storage and back-up, data archiving and data sharing.
- Determine if specialist expertise is required (data stewards)
- Determine if additional hardware or software is needed



# DMP examples

## Material science:

Hong, X. 2016. Effects of mine waste materials in the nort-central Mojave dessert. <https://dmptool.org/plans/23144/export.pdf>

Sergan, V. 2014. Orientational order induced by a polymer network in the isotropic phase of liquid crystal.

<https://dmptool.org/plans/12137/export.pdf>

## Social science:

Slavec, A. 2019. Using questionnaires to measure attitudes and behaviours of bulding users: Data Management plan.

<https://zenodo.org/record/3592299#.XhZutkdKguU>



# Example of dataset deposited on Zenodo

- Burnard, Michael, Schwarzkopf, Matthew, Tverezovskiy, Viacheslav, Treu, Andreas, Humar, Miha, & Kutnar, Andreja. (2018). Wood preservatives utilizing low-value olive oil production by-products: Analysis [Data set]. Zenodo.

<http://doi.org/10.5281/zenodo.1248463>



February 22, 2018

Dataset Open Access

# Wood preservatives utilizing low-value olive oil production by-products: Analysis

Burnard, Michael; Schwarzkopf, Matthew; Tverezovskiy, Viacheslav; Treu, Andreas; Humar, Miha; Kutnar, Andreja

The objective of this study was to develop and assess the efficacy of two experimental methodologies for the maleinisation of lampante oil to be used for wood protection.

Two maleinisation techniques were used to chemically modify low-value lampante oil in an attempt to limit leaching, increase hydrophobicity, and impart some level of antimicrobial performance when impregnated in wood. Pine and beech wood specimens were treated with the modified oils and underwent leaching, accelerated weathering, and fungi tests. The following analysis assessed the efficacy of the modified oil treatments in improving these characteristics.

Preview

SpecimenID	Species	Treatment	Number	StartDate	EndDate	Duration	Inoculant	InitialDryW
B-DM-MB-1	B	DM	1	13/07/2016	02/11/2016	16	pl	1.6776
B-DM-MB-10	B	DM	10	13/07/2016	02/11/2016	16	tv	1.6966
B-DM-MB-11	B	DM	11	13/07/2016	02/11/2016	16	tv	1.6703
B-DM-MB-12	B	DM	12	13/07/2016	02/11/2016	16	tv	1.6715
B-DM-MB-13	B	DM	13	13/07/2016	02/11/2016	16	tv	1.6945
B-DM-MB-14	B	DM	14	13/07/2016	02/11/2016	16	tv	1.6578
B-DM-MB-15	B	DM	15	13/07/2016	02/11/2016	16	tv	1.6721

Files (578.8 kB)

Name	Size
OOTFungi1MS_11112017_v1_r0.csv	50.8 kB
md5:9027d6f469bb7744830a66104efc25f	
OOTLeaching1MS-150417_v1_r0.csv	36.8 kB
md5:46f077ebdad2bd837f401eab2d6252a3	
OOTSupplementalAnalysis.pdf	439.0 kB
md5:b2e3e138460981d776e130d14faad23c	
OOTSupplementalAnalysis.Rmd	52.4 kB
md5:a3bfec64c23f9164a36bdf1268c3c562	

107 views 112 downloads

See more details...

Indexed in

**Publication date:** February 22, 2018

**DOI:** [10.5281/zenodo.1248463](https://doi.org/10.5281/zenodo.1248463)

**Grants:** [European Commission](#)

- InnoRenew CoE - Renewable materials and healthy environments research and innovation centre of excellence (739574)

**Communities:** [InnoRenew CoE and Project research](#)

**License (for files):** [Creative Commons Attribution 4.0 International](#)

**Versions**

Version 2 Feb 22, 2018  
10.5281/zenodo.1248463

Version 1 Feb 22, 2018  
10.5281/zenodo.1182868

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.1182867](https://doi.org/10.5281/zenodo.1182867). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

**Share**

**Cite as**

Burnard, Michael, Schwarzkopf, Matthew,



# Research Data Alliance (RDA)



# What is RDA?

RDA is an international **member based organization** focused on the development of infrastructure and community activities that reduce barriers to data sharing and exchange, and the acceleration of data driven innovation worldwide.

With more than 8,800 members globally representing 137 countries, RDA includes **researchers, scientists and data science professionals** working in multiple disciplines, domains and thematic fields and from different types of organisations across the globe.

*RDA is building the social and technical bridges that enable open sharing of data to achieve its vision of researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.*



# What does RDA do?

*Members come together through self-formed, volunteer, focussed Working Groups, exploratory Interest Groups to exchange knowledge, share discoveries, discuss barriers and potential solutions, explore and define policies and test as well as harmonise standards to enhance and facilitate global data sharing & re-use.*

RDA members collaborate together across the globe to tackle numerous infrastructure & data sharing challenges related to:

- 🌐 Reproducibility
- 🌐 Data preservation
- 🌐 Best practices for domain repositories
- 🌐 Legal interoperability
- 🌐 Data citation
- 🌐 Data type registries
- 🌐 Metadata
- 🌐 and so many more!



DATA SHARING



# Who Can Join RDA?

Any individual or organization, regardless of profession or discipline, with an interest in **reducing the barriers to data sharing and re-use** and who agrees to RDA's guiding principles of:

- Openness
- Consensus
- Balance
- Harmonization
- Community-driven
- Non-profit and technology-neutral

Individual Membership is free at <https://www.rd-alliance.org/user/register>



# Why Join RDA as an Individual Member?

## Individual Member Benefits

- **Contribute** to acceleration of data infrastructure development
- Work and **share experiences** with collaborators throughout the world
- **Access** to extraordinary network of colleagues with various levels of experience, perspectives and practices
- Gain greater **expertise** in data science regardless of whether one is a student, early or seasoned career professional
- **Enhance** the quality and effectiveness of personal work and activities
- **Improve** one's competitive advantage professionally and positioning oneself for leadership within the broader research community

Individual RDA Members 8,810



# RDA IGs and WGs

	Interest groups	Working groups
Number (Aug 19)	61	33
Members	Experts from the community	Experts from the community
Purpose	Platform for exchange on various topics	Focus on specific goal (concrete output)
Outputs	Supporting outputs	Recommendations and supporting outputs
Duration	Active over longer periods	Fixed period 12-18 months
Groups relevant for engineers	<ul style="list-style-type: none"><li>• <a href="#">Research Data Management in Engineering IG</a></li><li>• <a href="#">RDA/CODATA Materials Data, Infrastructure &amp; Interoperability IG</a></li><li>• <a href="#">Physical Samples and Collections in the Research Data Ecosystem IG</a></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">International Materials Resource Registries WG</a></li><li>• <a href="#">Persistent identification of instruments WG</a></li><li>• <a href="#">International Materials Data Registries WG</a></li></ul>



# 15th Plenary Meeting

# 15th Research Data Alliance Plenary Meeting

18-20  
MARCH  
2020

MELBOURNE  
AUSTRALIA



[www.rd-alliance.org](http://www.rd-alliance.org)





# Call for Early Careers

- Currently engaged in Bachelor, Masters, PhD or Postdoc studies (i.e. enrolled in a higher education study course or within maximum 5 years after completing a PhD);
- Study is conducted in a higher education or research institution based in one of the EU Member states or Associated countries
- Studies are of relevance to RDA Recommendations or Outputs and cover at least one of the Working or Interest Groups

15th Research Data Alliance  
Plenary Meeting  
Melbourne Convention and Exhibition Centre, MCEC  
Data for Real-World  
Impact  
18-20  
MARCH  
2020  
MELBOURNE  
AUSTRALIA

RDA Europe  
Early Career & Expert  
Travel grants

Deadline: 3<sup>rd</sup> February 2020

[→ Read blog post about my experience at the 12th RDA plenary as ERC](#)



# RDA in Slovenia

Website:

<https://www.rd-alliance.org/groups/rda-slovenia>

Contact persons: **Janez Štebe**,  
**Irena Vipavc Brvar** and **Maja Dolinar** (Slovenian Social Science Data Archive)



RESEARCH DATA ALLIANCE  
**EUROPE**





# Acknowledgments and resources

- 🌐 Jeuse, A. 2019. Adopting RDA Recommendations and Outputs Across the research data lifecycle: <https://www.rd-alliance.org/rda-outputs-overview-presentation>
- 🌐 RDA in a Nutshell (August 2019). <https://www.rd-alliance.org/sites/default/files/attachment/RDA-in-a-nutshell-August-2019.pptx>



This work has been produced with the support of the RDA Europe Ambassador programme. It has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme via the RDA EU 4.0 project (ref. GA no. 777388).



# Let's stay in contact



Ana Slavec

InnoRenew CoE

[ana.slavec@innorenew.eu](mailto:ana.slavec@innorenew.eu)

@aslavec

RDA Europe Ambassador for Engineering/Renewable materials: <https://www.rd-alliance.org/rda-disciplines/rda-europe-ambassadors>



This work has been produced with the support of the RDA Europe Ambassador programme. It has received funding from the European Union's Horizon 2020 (H2020) research and innovation programme via the RDA EU 4.0 project (ref. GA no. 777388).



# Hands-on exercise

**A) If you don't have data yet prepare a data management plan for your project using the [DMPonline tool](#):**

- 🌐 Description of data to be collected/created
- 🌐 Standards/methodologies for data collection and management
- 🌐 Ethics and intellectual property
- 🌐 Plans for data sharing and access
- 🌐 Strategy for long-term preservation

**B) If you already have data, prepare them to be deposited in a repository:**

- 🌐 Find a suitable domain-specific repository for your field using the [registry of research data repositories](#).
- 🌐 Select an appropriate format and prepare your dataset for deposit in a domain-specific repository or general-purpose repository (Zenodo)
- 🌐 Prepare a readme file that explains the content of the data file