



Ready for
BioData.pt
Management?



Intensive Course

Data Sharing & Reuse

Daniel Faria, Jorge Oliveira, Gil Poiares-Oliveira



I – Challenges



Learning Outcomes:

- Tackle the RDM challenges that arise in data sharing and reuse

Data Sharing & Reuse

- **Data Sharing** is the stage at which the only concern is making data available to others in a manner amenable to **Reuse**
 - We can (and often do) share data earlier in the life cycle with collaborators
 - The data is not final, but the challenges are identical



Data Sharing & Reuse

- **Data Reuse** is not really a stage
 - Making our data reusable should be a concern across the whole life cycle
 - Reusing existing data is a concern during the planning and collection stages
 - The challenges of these scenarios are two sides of the same coin



Data Sharing & Reuse

- **Challenges:**
 - Rights, obligations and restrictions
 - Accessibility
 - Selecting a repository
 - Licensing
 - Security
 - Documentation



Rights, Obligations & Restrictions

- **Rights:** are you the data owner or do you have consent from the data owner to share the data?
- **Obligations:** does your funder or institution require you to share the data?
- **Restrictions:** are there ethical, legal, contractual or IP reasons not to share the data



Accessibility

Type of Access	Who can access the data?	Metadata
Open Access	Anyone	Public
Registered Access	Authenticated users	Public
Controlled Access	Users approved by a Data Access Committee	Public
Access upon Request	Users approved by the data owner	???



Selecting a Repository

- Is there a discipline-specific repository available for your data?
- If not, is there an institutional or general purpose repository you could use?
- Does the repository support your desired type of access?
- Does it support the metadata standard for your domain?
 - Does it allow you to customize the metadata?



Licensing

- If you deposit your data in a public repository (open or registered access), you should add a license to it if able
 - Some repositories don't allow you to choose (data deposited there becomes public domain by default)
- For data under controlled access, a license usually does not make sense, as users will access the data under a contract



Licensing

Licence	Can I copy & redistribute the work?	Is it required to attribute the author?	Can I use the work commercially?	Am I allowed to adapt the work?	Can I change the licence when redistributing?
CC0	✓	✗	✓	✓	✓
CC BY	✓	✓	✓	✓	✓
CC BY-SA	✓	✓	✓	✓	✗
CC BY-ND	✓	✓	✓	✗	✓
CC BY-NC	✓	✓	✗	✓	✓
CC BY-NC-SA	✓	✓	✗	✓	✗
CC BY-NC-ND	✓	✓	✗	✗	✓



Security

- Are the security practices of a data repository adequate to your needs?
 - Does it implement an established authentication and authorization protocol?
 - Does it encrypt the data?
 - Does it have a reasonable backup policy? (this can be hard to check)



Documentation

- Documentation is again one of the most critical concerns when sharing data, as reusability hinges almost entirely on the quality of your documentation
- Fortunately, if we followed good documentation practices throughout the data life cycle, we don't need to do much at this stage
 - We can export and publish our experiments from the ELN
 - We can publish our data processing and analysis workflow



II – Hands-On



Learning Outcomes:

- Share data on a data repository
- Assess the reusability of a files

- A framework for data sharing on the web
 - Allows private sharing between collaborators
 - Organization consists of:
 - “Dataverses” (essentially folders) that can be unlimitedly nested and be public or private
 - Datasets which can be published or unpublished
- Has generic plus some domain-specific metadata



- It is a self-hostable framework, with several instances available throughout the work, mainly in universities
- BioData.pt has a public instance available at dmportal.biodata.pt
- For this course, we will be using a training instance available at dev.dmportal.biodata.pt
 - Again, do not use this instance beyond the course



Group Exercise

1. Access our training instance of DataVerse at dev.dmpportal.biodata.pt
2. Create a new account and fill in your details
3. Access the COVID-2.0 BioDataVerse
4. Choose one of the datasets listed on the [mock project](#) and draft a dataset for it on DataVerse, inside the COVID-2.0 BioDataVerse (you do not need to upload a data file, just fill-in the metadata describing the dataset)

Group Exercise

Citation Metadata ^

Title * ?

Add "Replication Data for" to Title

Author * ?

Name * ?

Poiares-Oliveira, Gil

Affiliation ?

BioData.pt | INESC-ID

+

Identifier Type ?

Select...

Identifier ?

Point of Contact * ?

Name ?

Poiares-Oliveira, Gil

Affiliation ?

BioData.pt | INESC-ID

+

E-mail * ?

gpo@biodata.pt

Description * ?

This field supports only certain [HTML tags](#).

Text * ?

+

Date ?

YYYY-MM-DD

Subject * ?

Keyword ?

Related Publication ?

Notes ?

Depositor ?

All file types are supported for upload and download in their original format. If you are uploading Excel, CSV, TSV, RData, Stata, or SPSS files, [see the guides](#) for tabular support and limitations.

Deposit Date ?

Controlled Vocabulary Name ?

+

Controlled Vocabulary URL ?

https://

Citation ?

+

Identifier Type ?

Select...

Identifier ?

URL ?

https://

Files

Upload with HTTP via your browser ^

Select files or drag and drop into the upload widget. Maximum of 1,000 files per upload.

+ Select Files to Add

Drag and drop files here.



Group Exercise

5. Access a dataset entered by one of your colleagues and assess its reusability



Thank You!

Questions?