



Data Sharing & Reuse

Daniel Faria, Jorge Oliveira, Gil Poiares-Oliveira







Data Sharing & Reuse

# 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 up in Request	Users approved by the data owner	???



# **Selecting a Repository**

- o 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	<b>/</b>	×	<b>V</b>	<b>V</b>	<b>✓</b>
CC BY	<b>/</b>	<b>V</b>	<b>V</b>	<b>V</b>	<b>✓</b>
CC BY-SA	<b>/</b>	<b>V</b>	<b>V</b>	<b>V</b>	×
CC BY-ND	<b>/</b>	<b>V</b>	<b>V</b>	×	<b>✓</b>
CC BY-NC	<b>/</b>	<b>V</b>	×	<b>V</b>	<b>✓</b>
CC BY-NC-SA	<b>/</b>	<b>V</b>	×	<b>V</b>	×
CC BY-NC-ND	<b>V</b>	<b>V</b>	×	×	<b>V</b>



# **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





Data Sharing & Reuse

# 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 <u>dmportal.biodata.pt</u>



- 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

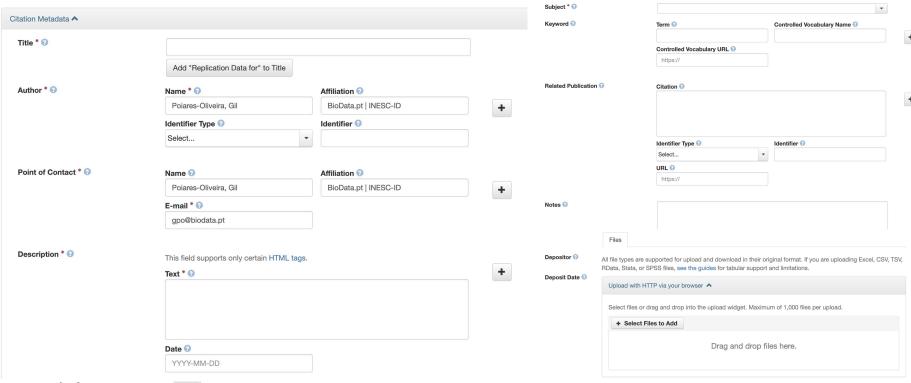


# **Group Exercise**

- Access our training instance of DataVerse at dev.dmportal.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 <u>mock project</u> 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**







# **Group Exercise**

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



Data Sharing & Reuse

# **Thank You!**

# **Questions?**



