



*Proteção de dados e ética*

**OPEN SCIENCE TOOLKIT**



# Open Science Toolkit

## Data Protection and Ethics

This course helps you to get to grips with responsible data sharing.



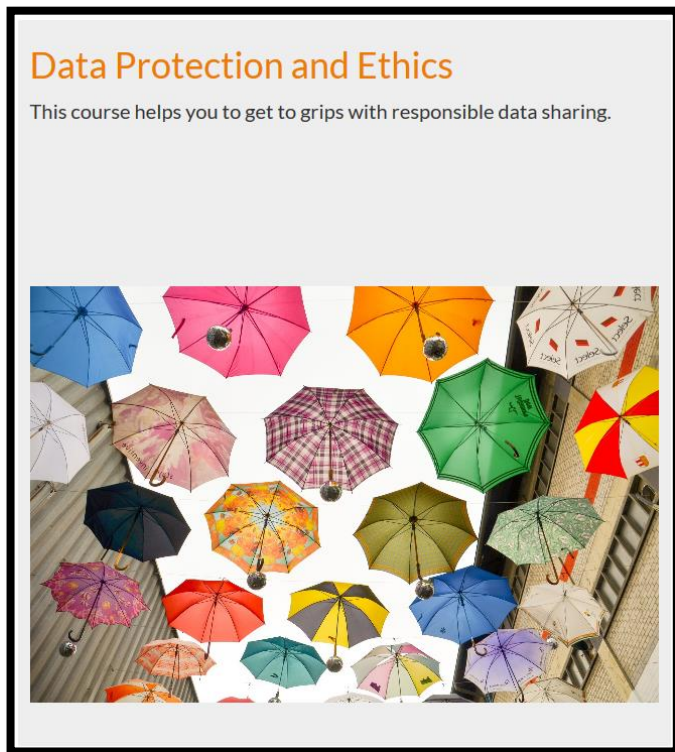
# Módulo: Proteção de dados e ética

## OBJETIVOS DE APRENDIZAGEM

Este modulo abrange as questões da proteção de dados em particular, e de forma mais genérica as questões éticas. Pretende apoiar a compreensão dos princípios básicos da proteção de dados introduzindo as técnicas de implementação da proteção de dados no processo de investigação.

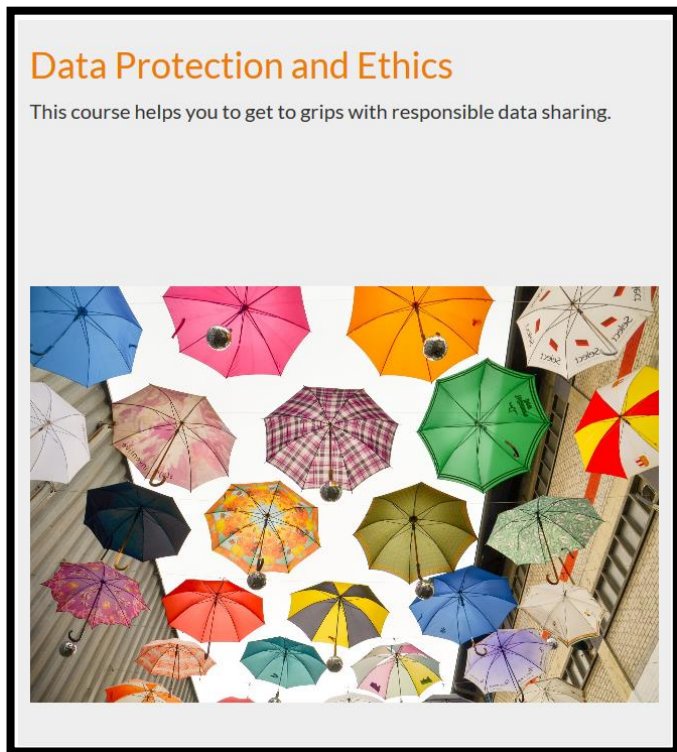
- ✓ Identificar **dados pessoais** e como podem ser **protegidos**;
- ✓ Saber o que considerar na elaboração de ***consent forms***
- ✓ Conhecer como **armazenar dados em segurança**
- ✓ Compreender como **anonimizar** conjuntos de dados

# Visão geral dos conteúdos do módulo



- Proteção de dados
  - O que são dados pessoais
  - Proteger dados pessoais
- Requisitos legais
  - GDPR
- Proteção de dados e partilha de dados de investigação
- Proteção no planeamento da investigação
- Consentimento informado
- Gerir dados em segurança
  - Armazenamento de dados em segurança
- Estratégias de anonimização

# Visão geral dos conteúdos do módulo



## FOCADO NA PRÁTICA

### Consentimento informado

- Processos práticos para obter consentimento
- *Consent forms*, declarações

### Gerir dados em segurança

- Armazenamento de dados em segurança

### Estratégias de anonimização

- Casos de estudos e ferramentas de anonimização

# Alguns aspetos relevantes (1/5)

Informação  
sistematizada

## What are personal data?

- + What are personal data?
- + Protecting personal data
- + Legal requirements - EU General Data Protection Regulation (GDPR)
- + Legal requirements - GDPR research exemptions

### — What are personal data?

The term personal data refers to any information that can be used to identify living (and sometimes even deceased) individuals. For example, their name or date of birth. Some personal information - such as a person's origin, political opinion, religious beliefs, health, trade union membership, or sexual orientation - is classified as sensitive personal data. If you are handling personal and/or sensitive personal data as part of your research, you need to make sure it is protected.

If you aren't sure whether your research involves collecting personal data, it is best to speak with someone in your ethics team or data protection office to get advice as early as possible so that you can build in the appropriate data protection measures.

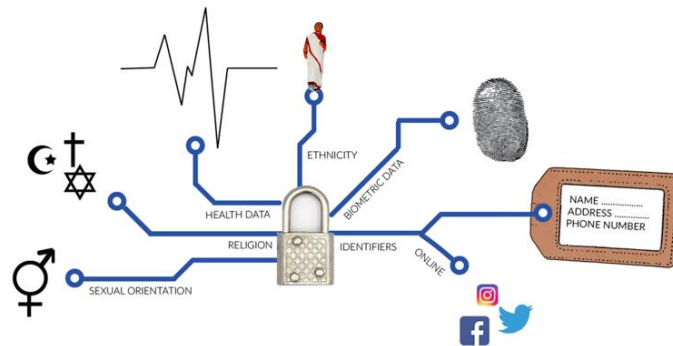


Image reused from CESSDA ERIC Expert Tour Guide on Data Management which is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License



# Alguns aspetos relevantes (2/5)

Dicas e questões  
práticas

## Planning your research

Consider what personal data you'll be working with and any measures you'll need to put in place from the earliest stages of your project. Think about:



Do you really need to collect personal data?

Always take the time to re-consider if there are ways that you can collect data anonymously. This could save you a lot of headaches later. So only commit to collecting personal data if it is essential to your research.



Data collection



# Alguns aspetos relevantes (3/5)


Ajudas práticas

FOSTER

## Sample consent forms

If you collect personal data in your research, you are required to obtain informed consent for data collection, processing, sharing, and preservation. The [CESSDA Expert Tour Guide on Data Management](#) includes useful guidance on developing consent forms. Check out the sample forms they provide - like the one below - which can help you create your own consent forms.

For more information on ethical aspects of research including the creation of consent forms, check out this [presentation](#) by Libby Bishop, UK Data Service in the FOSTER portal.



### Consent Form for [name of project]

Please tick the appropriate boxes

	Yes	No
<b>Taking Part</b>		
I have read and understood the project information sheet dated DDMMYYYY.	<input type="checkbox"/>	<input type="checkbox"/>
I have been given the opportunity to ask questions about the project.	<input type="checkbox"/>	<input type="checkbox"/>
I agree to take part in the project. Taking part in the project will include being interviewed and recorded (audio or video).	<input type="checkbox"/>	<input type="checkbox"/>
I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.	<input type="checkbox"/>	<input type="checkbox"/>
<b>Use of the information I provide for this project only</b>		
I understand my personal details such as phone number and address will not be revealed to people outside the project.	<input type="checkbox"/>	<input type="checkbox"/>
I understand that my words may be quoted in publications, reports, web pages, and other research outputs.	<input type="checkbox"/>	<input type="checkbox"/>
<b>Please choose one of the following two options:</b>		
I would like my real name used in the above	<input type="checkbox"/>	
I would not like my real name to be used in the above.		<input type="checkbox"/>
<b>Use of the information I provide beyond this project</b>		
I agree for the data I provide to be archived at the UK Data Archive. <sup>1</sup>	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>
I understand that other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	<input type="checkbox"/>	<input type="checkbox"/>
<b>So we can use the information you provide legally</b>		
I agree to assign the copyright I hold in any materials related to this project to [name of researcher].	<input type="checkbox"/>	<input type="checkbox"/>
Name of participant [printed] _____	Signature _____	Date _____
Researcher [printed] _____	Signature _____	Date _____
Project contact details for further information: Name, phone, email addresses, etc.		
<b>Notes:</b>		
1. Other forms of participation can be listed.		
2. More detail can be provided here so that decisions can be made separately about audio, video, transcripts, etc.		
<small>This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.0 UK, England &amp; Wales License. To view a copy of this license, visit <a href="http://creativecommons.org/licenses/by-nc-sa/2.0/uk/">http://creativecommons.org/licenses/by-nc-sa/2.0/uk/</a></small>		

UK Data Archive sample consent form (as presented in CESSDA Expert Tour Guide on Data Management)



# Alguns aspetos relevantes (4/5)

Boas práticas

## European Genome-Phenome Archive

This short video by ELIXIR EXCELERATE provides a concise overview of the European Genome-Phenome Archive and the use of BEACON technology.

ELIXIR-EXCELERATE - Human genomics and translational data



# Alguns aspetos relevantes (5/5)

## Sugestão de ferramentas

### Case study - anonymisation tools



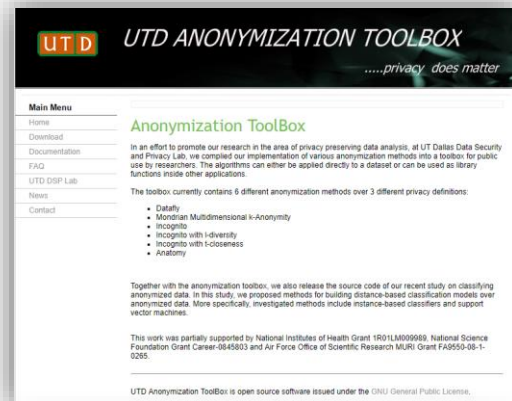
The screenshot shows the Amnesia website. At the top, there is a navigation bar with links: Home, Get Amnesia!, What is Amnesia?, Documentation, On-line version, and About. The main heading is "Amnesia". Below it, a paragraph describes the tool: "Amnesia is a data anonymization tool, that allows to remove identifying information from data. Amnesia not only removes direct identifiers like names, IDs etc but also transforms secondary identifiers like birth date and zip code so that individuals cannot be identified in the data. Amnesia supports k-Anonymity and l-Diversity." Below the text is a logo consisting of a stylized person icon and the word "amnesia". To the right of the logo is a large orange arrow pointing right. At the bottom left, the word "Amnesia" is written again. At the bottom right, there are three colored circles (orange, grey, grey).

#### What?

Amnesia is a freely available anonymisation tool that helps researchers to automatically remove direct identifiers and to transform indirect identifiers in their research data. Amnesia is available both as an [online service](#) and as a [local application](#).

Originally developed through support from the Athena Research Center and the MEDA project, the latest version has been provided through support from the OpenAIRE2020 project and the H2020 project My Health My Data.

A webinar explaining how the tool works was aired in April 2018. The slides and recordings are available from the [OpenAire](#) website.



The screenshot shows the UTD Anonymization Toolbox website. At the top, there is a header with the UTD logo and the text "UTD ANONYMIZATION TOOLBOX" and "....privacy does matter". Below the header is a "Main Menu" with links: Home, Download, Documentation, FAQ, UTD DSP Lab, News, and Contact. The main heading is "Anonymization ToolBox". Below it, a paragraph describes the toolbox: "In an effort to promote our research in the area of privacy preserving data analysis, at UT Dallas Data Security and Privacy Lab, we compiled our implementation of various anonymization methods into a toolbox for public use by researchers. The algorithms can either be applied directly to a dataset or can be used as library functions inside other applications." Below the text is a list of methods: "The toolbox currently contains 6 different anonymization methods over 3 different privacy definitions: Differential, Mondrian Multidimensional k-Anonymity, Incognito, Incognito with l-diversity, Incognito with l-closeness, and Anonymity." Below the list is a paragraph: "Together with the anonymization toolbox, we also release the source code of our recent study on classifying anonymized data. In this study, we proposed methods for building distance-based classification models over anonymized data. More specifically, investigated methods include instance-based classifiers and support vector machines." Below the paragraph is a paragraph: "This work was partially supported by National Institutes of Health Grant R01LM006669, National Science Foundation Grant CAREER-0845003 and Air Force Office of Scientific Research MURI Grant FA9550-06-1-0265." At the bottom, there is a footer: "UTD Anonymization ToolBox is open source software issued under the GNU General Public License."



The screenshot shows the edding 3000 website. At the top, there is a header with the text "Hi, I am" and "My social security n". Below the header is a large image of an edding 3000 marker. At the bottom, there is a small illustration of a person standing next to a ladder.



Obrigado! Questões?

[www.fosteropenscience.eu](http://www.fosteropenscience.eu)

Twitter: @fosterscience

Facebook: fosteropenscience