



@MarjanGrootveld  
@openaire\_eu

# Questions about Research Data Management - and some answers



# Where do we go?

- What's OpenAIRE?
- EC's Open Research Data Policy
- Questions and answers about RDM and DMPlanning
- And apologies for the unbalanced number of Dutch examples

# WHAT IS OPENAIRE?



Fosters the **social** and **technical** links

OpenAIRE that enable Open Science in Europe and beyond

**Human  
Network**



**Digital  
Network**

A “dual core” eInfrastructure  
for Open Scholarship



# OpenAIRE

- **50 Partners** from every EU country, and beyond
- In **24/7** operation since **2010**
- 4 project phases to date
- Outlook: OpenAIRE Advance starts in 2018



# OpenAIRE support materials

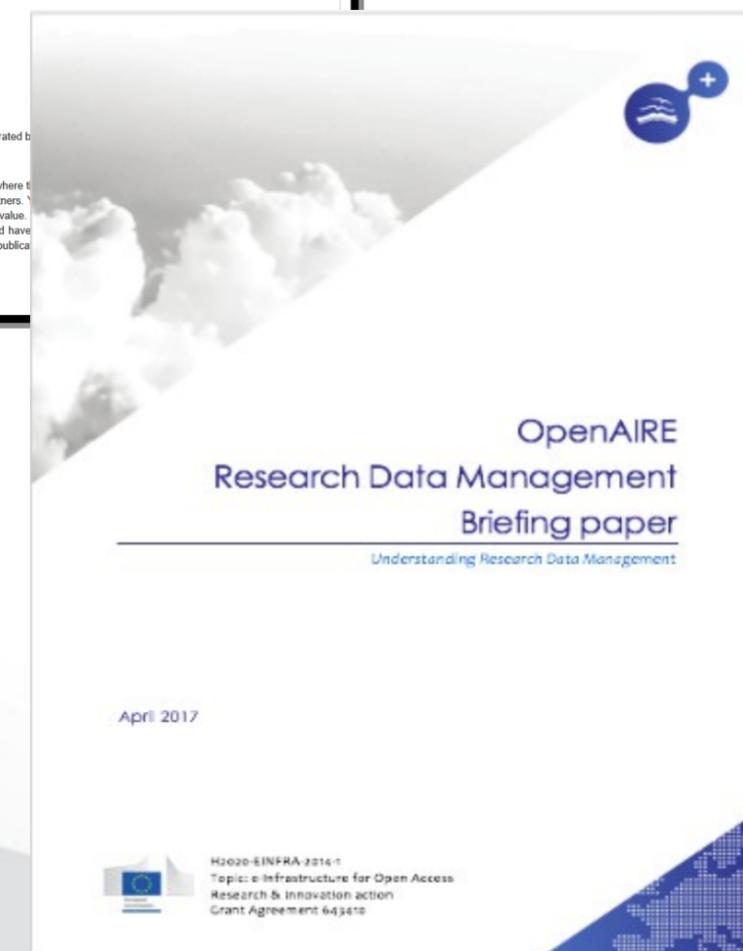
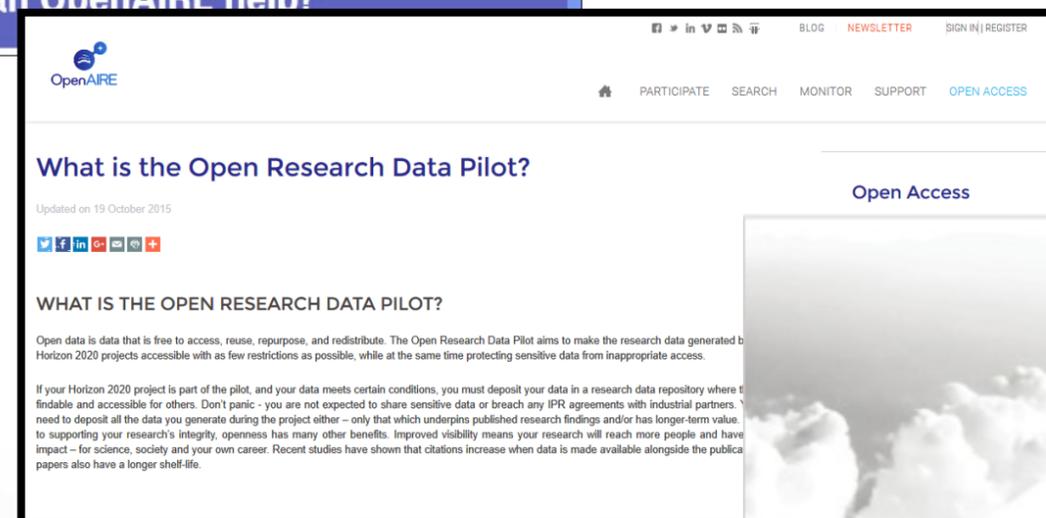
- Briefing papers, factsheets, webinars, workshops, FAQs
- Information on:
  - Open Research Data Pilot
  - Creating a data management plan
  - Selecting a data repository

<https://www.openaire.eu/what-is-the-open-research-data-pilot>

<https://www.openaire.eu/support> (factsheets)

<https://www.openaire.eu/webinars/>

<https://www.openaire.eu/briefpaper-rdm-infonoads>





[PARTICIPATE](#)

[SEARCH](#)

[MONITOR](#)

[SUPPORT](#)

[OPEN ACCESS](#)



Search in 22,036,305 publications 571,373 datasets from 2,711 repositories and OA journals

G7 Science Ministers  
endorse Open Science



# EC'S OPEN RESEARCH DATA POLICY

## Open Research data

### Open Research Data Pilot (2014-2016)

**As of the Work Programme 2017: extended to cover all thematic areas of Horizon 2020 ('open by default')**

- Projects may still opt-out at any stage (IPR, personal data protection, national security, other reasons)
- Mainly concerns data underlying publications (other data optional)
- Data Management Plan obligatory by M6 (not part of project evaluation)

**Approach: as open as possible, as closed as necessary**

## Results from the Open Research Data Pilot (Finalised data 2014-2016)

Year	Core areas (CA)		Non-core areas (NCA)	Sample Size (signed GAs)
	Stay in	Opt-out	Opt-in	
2014-2015	67,44%	32,56%	2,5%	390 (CA) 3244 (NCA)
2015-2016	68,13%	31,87%	15,25%	935 SGA (CA) 6250 (NCA)
<b>Average</b>	<b>67,78%</b>	<b>32,21%</b>	<b>8,87%</b>	

Voluntary opt-ins:  
81 + 953 projects

## Clarifying terminology...



In the past our policy mainly addressed the 'accessibility' part of FAIR.

- Started off with 'open access to research data'
- Moved towards open (research) data with the ORD pilot (which also covered further aspects)
- We are now seeing openness as one component of FAIR data and aim to address all of the FAIR aspects in Horizon 2020

# RESEARCH DATA – OPEN BY DEFAULT



EC in the Guidelines: “This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept (...) without suggesting any specific technology, standard, or implementation solution”

FAIR Data Management



# The EC Open Research Data policy

## Key sources of information

- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- Guidelines on FAIR Data Management in Horizon 2020  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)  
-
- Infographic summarising key policy points  
[http://ec.europa.eu/research/press/2016/pdf/opendata-infographic\\_072016.pdf](http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf)
- Open Access and Data Management  
[http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)  
-

# SUBMITTED QUESTIONS

# Many thanks for submitting your questions!

- “Funder processes”
  - What is required when we opt out (on a H2020 project starting in 2017)?
  - How qualified are <DMP> reviewers?
  - Do funders check that action points in plans are carried out? (DMP compliance)
- Support for DM Planning
  - How are librarians involved in producing/ collaborating on DMPs? And in getting good funding?
  - How can we calculate the impact of RDM on institutional budgets?
- Storing and sharing data
  - How can we be sure that data we collect is real, and how to process data without altering it?
  - How can we manage research data in repositories?
- Legal and ethical questions
  - How can we deal with IP issues, especially in the case of project partners from the private sector?
  - GDPR: should agreement on long-term preservation and sharing of data be built into the consent process?
  - How is anonymisation implemented in text analysis and in large surveys in social sciences? (education researcher dealing with datasets containing students’ opinions and personal information.)
- Which online course will teach my team to manage the project’s data coherently, reproducible and open access?

# 1. Funder processes – opting out

- Opting out is possible at all stages: grant application, project start, during project.
- Consider if a *partial* opt-out is possible: it's preferable to full opt-out.
- “During the lifetime of a project, a total opt-out is possible for any of the reasons highlighted above. In this case, Article 29.3 is removed from the Grant Agreement via an amendment.” (1)
- Justify opting-out in the grant proposal or the DMP.
- Inform your project officer.
- Adapt your DMP.
- “A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt out are still encouraged to submit a DMP on a voluntary basis.” (2)

(1)

[http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm)

(2)

[http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

# Reasons for opting out

- Participation is incompatible with the Horizon 2020 obligation to protect results that can reasonably be expected to be **commercially or industrially exploited**;
- Participation is incompatible with the need for confidentiality in connection with **security issues**;
- Participation is incompatible with rules on **protecting personal data**;
- The project will **not generate / collect any research data**; or
- There are **other legitimate reasons** not to take part in the Pilot.

*Note that **partial opt out is possible** – and preferable to full opt out!*

# Funder processes – reviewers

“How qualified are reviewers?” Maybe not formally qualified, but

- The EC contracts data management experts to review DMPs and provide recommendations
- Project officers receive training
- The Commission installed some Expert Groups, e.g. on the FAIR data principles



European Commission > Register of Commission expert groups and other similar entities > Group Details

REGISTER OF COMMISSION EXPERT GROUPS and Other Similar Entities

Group Details - Commission Expert Group

Name: **Horizon 2020 Commission expert group on Turning FAIR data into reality (E03464)** Active

Type A - Individual expert appointed in his/her personal capacity

Name	Nationality	Professional Title	Membership Status
<a href="#">Collins Sandra</a>	Ireland		Member
<a href="#">Genova Françoise</a>	France		Member
<a href="#">Hodson Simon</a>	United Kingdom		Member
<a href="#">Jones Sarah</a>	United Kingdom		Member
<a href="#">Laaksonen Leif</a>	Finland		Member
<a href="#">Mietchen Daniel</a>	Denmark		Member
<a href="#">Patrauskaite Ruta</a>	Lithuania		Member
<a href="#">Wittenburg Peter</a>	Germany		Member

# What makes a good DMP?

Has the researcher taken time to reflect on what to do?

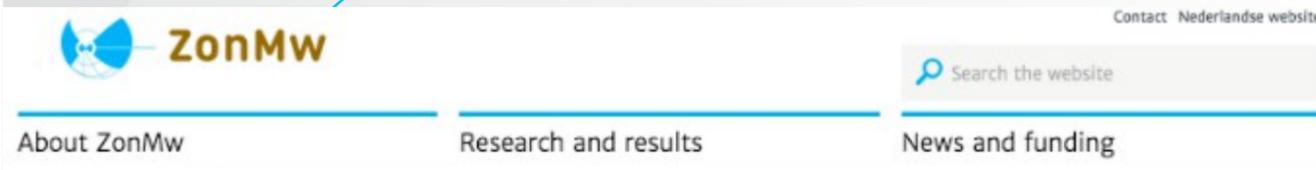
- The reviewer wants to be reassured that due consideration has been given to data and that the approach seems reasonable.
  - Focus on the data – don't describe how you'll deliver your publications.
  - Make plausible that all project partners and work packages commit to the DMP.
- Is the plan appropriate?
  - Adopt relevant standards and practices that are in line with norms for your field.
  - Show that you know about support services e.g. university storage, subject repositories...
- Does the plan show proper engagement with the issues?
  - Be specific.
  - Justify decisions and any restrictions.
  - Plan when you evaluate and update your DMP.



# Funder processes – DMP Compliance

“Do funders check that action points in plans are carried out?”

- When you announce an update, make sure to deliver it.
- Machine-actionable DMPs will support compliance checking.



When a project has ended, ZonMw checks whether the researcher has properly carried out the DMP. This is done on the basis of a few key items, which are made public.

**There are 5 key items:**

1. The link to the repository's search engine, or to the online (metadata) catalogue or web portal listing the data collection
2. The DOI code (persistent identifier) providing a permanent link to the data collection
3. The link (or a persistent identifier) to the data collection's terms of use (this key item can be dropped if the collection is open access)
4. The metadata standard allowing the data collection to be linked with other collections
5. The link to the archive or trusted digital repository for long-term archival

## In the spotlight

### Subjects

- [Data management in research](#)
- [Grant term data files](#)
- [Data management plan DMP](#)
- [Data management in ZonMw-procedures](#)
- [FAIR data](#)

### How to create a Data management plan



<https://www.zonmw.nl/en/research-and-results/access-to-data/>

# 2. Support for Research Data Management

“How are librarians involved in producing/ collaborating on DMPs?”

Grootveld (2017)

**Activities by research cycle phase**  
 The list below was compiled by collecting the activities mentioned on university library websites as part of their data management support services. We have sorted them into categories based on the different phases of the research data life cycle.

**'Creating data' phase:**

- Providing assistance in writing data management plans and project proposals / grant submissions
- Providing templates for data management plans
- Providing information and advice on grant provisions and data-related demands set by funders (NWO, ZonMw, Horizon2020)
- Providing information and advice on institutional data policy and codes of conduct
- Providing support in searching for existing data, or referring to systems and search engines (such as NARCIS, DataCite, and B2Find)
- Providing support and advice on legal and ethical matters (such as WBP, or the risk qualification of data) and data anonymisation
- Providing advice on the costs of data management, and how to budget for it in a project proposal

**'Processing data' phase:**

- Providing (data) lab environments and Virtual Research Environments (VREs)
- Providing local storage or repository systems to save and share data during research (such as DataverseNL, Beehub, SURFdrive, Figshare, or SharePoint)
- Providing software management systems
- Providing advice on (technical) data security
- Referring to discipline-specific storage solutions
- Providing persistent identifier services (such as Handle or DOI)
- Providing advice on temporary storage
  - Providing advice on metadating and standards
  - Providing assistance in documenting data

**'Analysing data' phase:**

- Providing visualisation tools and infrastructure
- Providing computational services (on campus or at SURFsara, for instance)

Van Berchum &

- Providing or listing relevant software (such as lab journals)
- Providing assistance in documenting data

**'Preserving data' phase:**

- Providing advice on sustainable archiving
- Providing a Trustworthy Digital Repository (currently just 4TU.ResearchData)
- Referring or linking to a national or international TDR
- Providing persistent identifier services (such as Handle or DOI)
- Archiving non-digital data

**'Giving access to data' phase:**

- Providing advice on access, both Open and Restricted, and embargo periods, if applicable
- Registering data sets in a local CRIS system
- Providing information and advice on funders' and publishers' demands
- Providing information and advice on licences
- Referring to relevant data journals

**'Reusing data' phase:**

- Providing search engines (locally, NARCIS, etc.)
- Providing information and advice on data citation
- Offering course modules to research master and PhD students

# Support for RDM – budgeting

“How can we calculate the impact of RDM on institutional



The screenshot shows the Utrecht University Research Data Management website. The header includes the Utrecht University logo, a search bar, and a language selector for 'Nederlands'. The main heading is 'Research Data Management'. Below this is a navigation bar with links: 'Your data management guide', 'Data Management Glossary', 'FAQ', 'RDM Support', and 'Contact us'. The breadcrumb trail reads 'UU.nl / Research / Research Data Management / RDM Support / Data Management Plan'. The main content area features a blue background with laboratory glassware. A yellow banner reads 'Data Management Plan'. Below this, there is a section titled 'Research Data Management' with a sub-section 'Your data management guide'. The main text describes a Data Management Plan (DMP) as a document outlining key aspects of collecting, storing, and managing research data. It states: 'A data management plan (DMP) outlines all key aspects of collecting, storing and managing research data during and after a project. Investing time and attention in DMP creation'. A quote follows: 'We offer help to formulate a sound and effective data management plan.'



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/). **Please cite as: A. Westerhof, T.E. Pronk, A. van der Kuil, A. Mordant (2016) Data Management Cost Guide.**

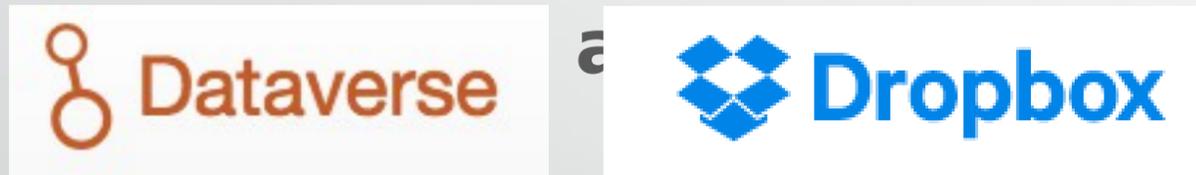
[https://www1.edugroepen.nl/sites/RDM\\_platform/Financieel1/Data%20Management%20Costs.aspx](https://www1.edugroepen.nl/sites/RDM_platform/Financieel1/Data%20Management%20Costs.aspx)

**When citing, please also cite: UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex.**



# 3. (Storing and) sharing data

With collaborators while research is active



(Open) data sharing



Data Repository

Data are stable, searchable, citable, clearly licensed

# Storing and sharing data – data integrity

“How can we be sure that data we collect is real, and how to process data without altering it?”

Hm. “real” is hard to guarantee, but transparency helps:

- Be explicit about your method for collecting/generating and processing data.
- Always store your raw data and document your changes.
- Commit to a good data organisation: folder and file naming, versioning agreements etc.
- Have a clear procedure for authorising people to access the data.
- Find a trustworthy repository for long-term access: they have processes for data integrity and authenticity and access regimes.

CoreTrustSeal

Data Seal of Approval

ICSU World Data System

nestor seal

ISO 16363



# Storing or archiving? Both!

**Storing and backing up files while research is active**

**Archiving or preserving data in the long-term**



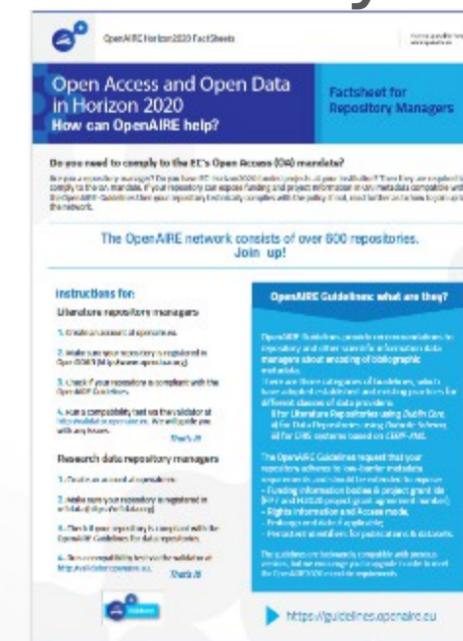
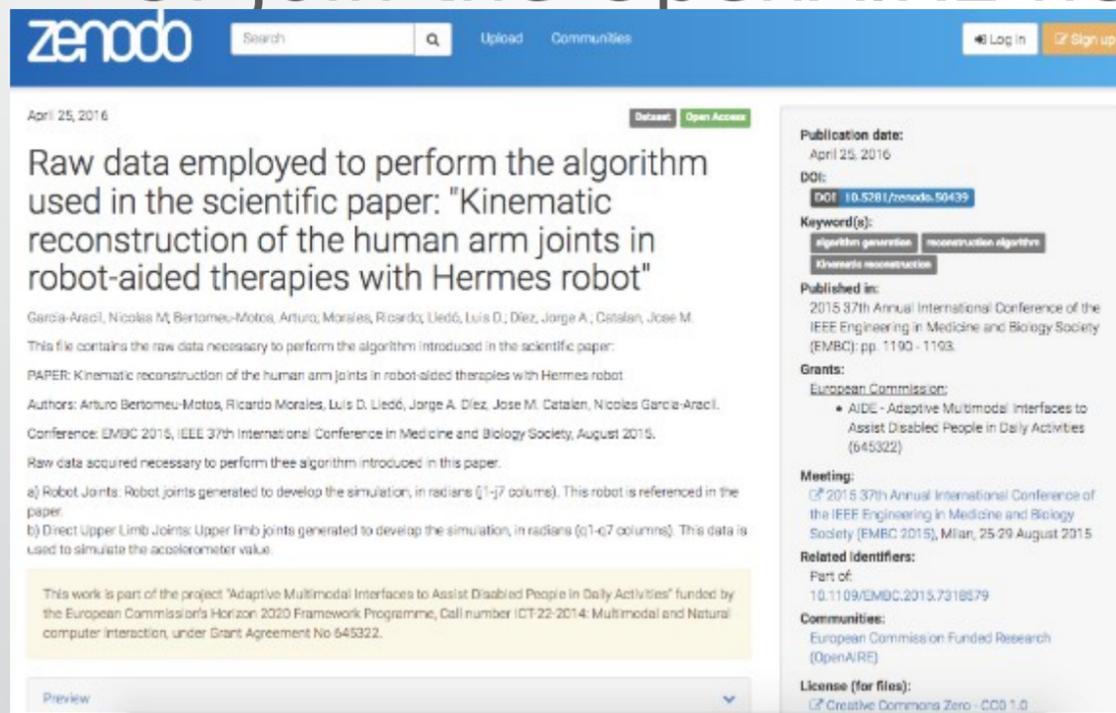
**Easy to change or delete**

**Safeguarded and preserved**

# Storing and sharing data

“How can we manage research data in repositories?” Hm.

- Rather, you outsource data management (stewardship) to the repository.
  - TDRs may seem picky, but that’s necessary for preservation.
- Or join the OpenAIRE network when you want your content



<https://zenodo.org/>

<https://www.openaire.eu/openaire-factsheet-for-repository-managers>

# 4. Legal and ethical – IPR

“How can we deal with IPR issues, especially in the case of project partners from private sector?”

Good practice: H2020 project i-Media-Cities addresses IPR in – public! – DMP.

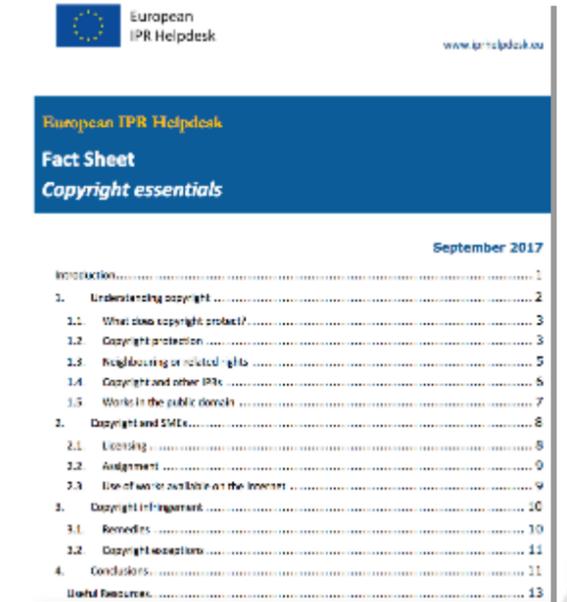


Innovative e-environments for Research on Cities and the Media

## 2. Data of I-Media-Cities

When speaking of the data used, collected, processed and generated within I-Media-Cities, a distinction has to be made between the data and metadata existing before the project (*Background*) and the data and metadata created within the project (*Foreground*), as defined in D 1.1.

I-Media-Cities will use, collect, process and generate both types of data and metadata, but the IPR-related solutions for these two types differ greatly:



Guidance  
How to draw up your consortium agreement

Version 2.0  
23 February 2015

- <https://www.iprhelphdesk.eu/>
- [http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/h2020-guide-cons-a\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/other/gm/h2020-guide-cons-a_en.pdf)
- <https://www.iprhelphdesk.eu/Fact-Sheet-Copyright-Essentials>
- <https://imediacities.eu/results/>

Disclaimer: OpenAIRE is no legal advisor



# Legal and ethical – GDPR

“Once GDPR – General Data Protection Regulation – comes into effect next year, what issues are likely to emerge for RDM where human subjects are involved (e.g. social science datasets) - should agreement on the long-term preservation and sharing of those data be built into the consent process?”

- GDPR makes explicit **exemptions for the processing of personal data for research** as well as for archiving personal data “in the public interest” (including long-term preservation).
- No clear picture yet on **complementary national laws**.
- Highly recommended to include in **informed consent forms** that data might or will be stored for the long term, and that these datasets can be accessed:
  - Available for everyone or access restricted to academic research
- **Without informed consent forms** personal data may only be stored/accessed for research under very strict conditions – dependent on national legislation.
- In all cases:
  - principle of **data minimisation**: collect & store as little personal data as possible;
  - processing personal data will be subject to **appropriate technical and organisational safeguards for the rights and freedoms of the data subject** (respondent, test subject...).

<http://www.eugdpr.org/> EU general GDPR portal

Based on information provided by Heiko Tjalsma, DANS legal officer

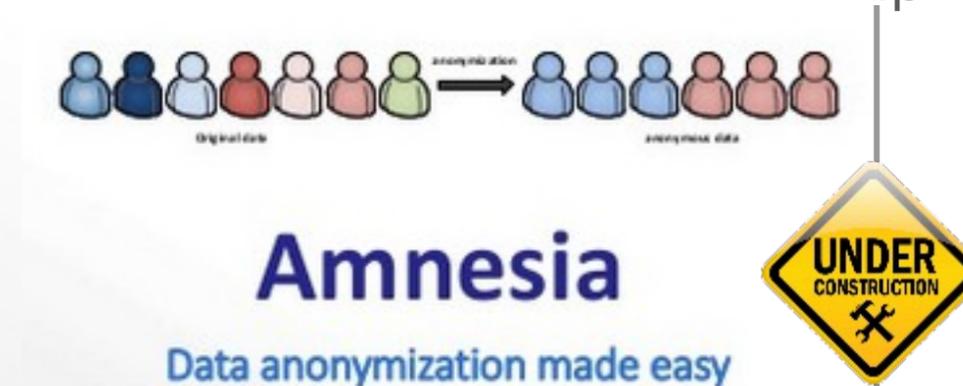
The OpenAIRE webinar tomorrow morning is about **Legal aspects of Open Science**

# Legal and ethical – anonymisation

“How is anonymisation implemented in text analysis, and in large surveys in social sciences?”

In general:

- Don't collect identifiable data – data minimisation principle!
- Anonymise data: remove, generalise, aggregate or distort data identifiers. Plan anonymisation at the time of transcription or initial write-up.



- Use pseudonyms or replacements that are consistent within the research team and throughout the project.
- Generalise the meaning of a detailed text variable by replacing potentially disclosive free-text responses with more general text.
- Get informed consent for sharing personal data (tricky but might be worth trying).

In short: preventing identification should be part of the study design.

# 5. RDM training

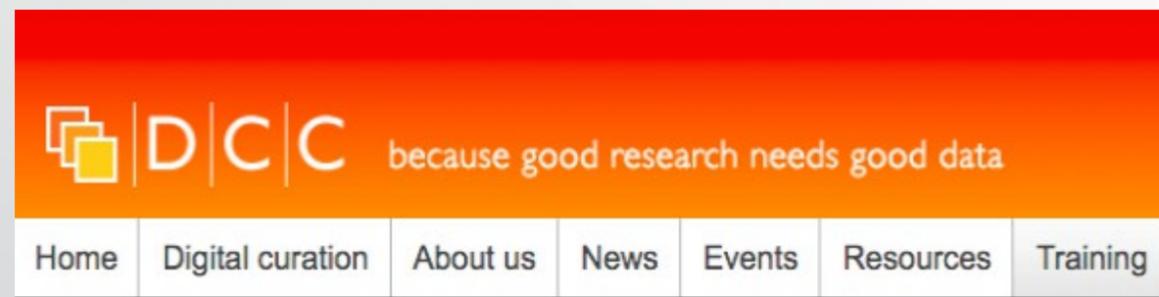
“Which online course will teach my team to manage the project’s data coherently, reproducible. open access?”



RDNL’s Essentials 4 Data Support  
<http://datasupport.researchdata.nl/en/>  
all content available for blended learning under CC-BY-SA



Many resources on Open Science and RDM  
<https://www.fosteropenscience.eu/resources>



Summary of training materials from various organisations  
<http://www.dcc.ac.uk/training/data-management-t-courses-and-training>



MOOC by University of North Carolina at Chapel Hill & University of Edinburgh  
<https://www.coursera.org/learn/data-management>

# Bonus: updated GA

Version 5.0, 18 October 2017, shows changes in

- Article 29.3 “Open access to research data”: to provide for third party access to research data **in health actions** in cases of **public health emergencies**. Access may also be required of health actions *not* participating in the Open Data pilot.

- Article 34 “Ethics and research integrity”: to align the provisions on ethical and **research**

**integrity principles** to the new (2017) European

[http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf)

<http://www.allfea.org/wp-content/uploads/2017/03/All-FA-European-Code-of-Conduct-for-Research-Integrity-2017-1.pdf>



# GA Article 34.1

Grant Agreement article 34.1 “Obligation to comply with ethical and research integrity principles”

In addition, the beneficiaries must respect the fundamental principle of research integrity — as set out in the European Code of Conduct for Research Integrity<sup>51</sup>.

This implies compliance with the following fundamental principles:

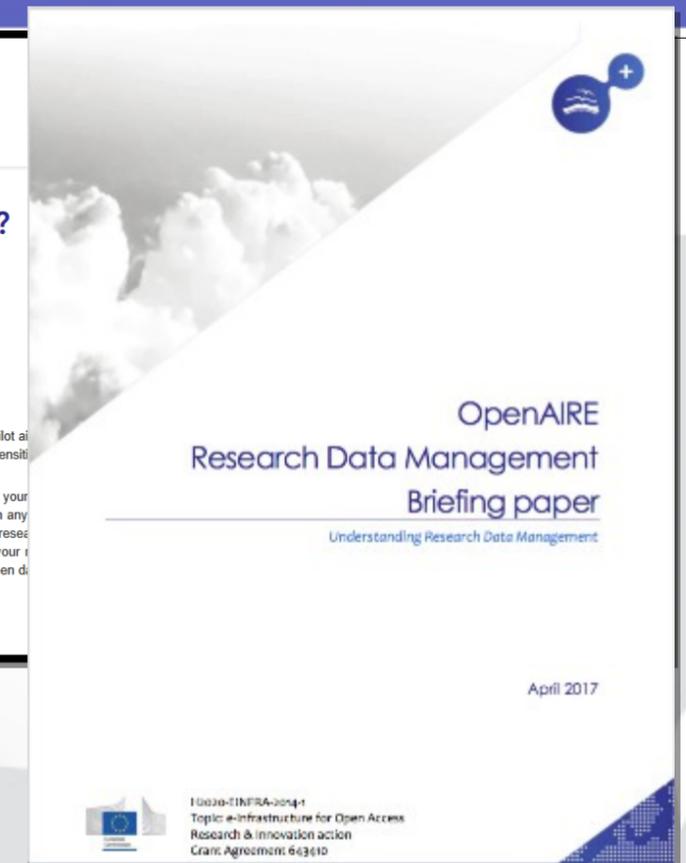
- **reliability** in ensuring the quality of research reflected in the design, the methodology, the analysis and the use of resources;
- **honesty** in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair and unbiased way;
- **respect** for colleagues, research participants, society, ecosystems, cultural heritage and the environment;
- **accountability** for the research from idea to publication, for its management and organisation, for training, supervision and mentoring, and for its wider impacts

and means that beneficiaries must ensure that persons carrying out research tasks follow the good research practices and refrain from the research integrity violations described in this Code.

# OpenAIRE support materials

<https://www.openaire.eu/opendatapilot>  
<https://www.openaire.eu/support>  
[https://www.slideshare.net/openaire\\_eu](https://www.slideshare.net/openaire_eu)

- Briefing papers, factsheets, webinars, workshops, FAQs
- Information on:
  - Open Research Data Policy
  - Creating a data management plan
  - Selecting a data repository



# Related webinars

[Introductory RDM webinar](#), Tony Ross-Hellauer & Marjan Grootveld, 30 May 2017:

- Reasons to manage data
- How to manage and share data (+ how to respond to concerns about sharing)
- OpenAIRE services
- Slides: [https://www.slideshare.net/OpenAIRE\\_eu/20170530open-research-data-in-horizon-2020](https://www.slideshare.net/OpenAIRE_eu/20170530open-research-data-in-horizon-2020)



[Open Research Data in H2020 and Zenodo](#), Marjan Grootveld & Krzysztof Nowak, 26 October 2016:

- Sustainable file formats differ across domains and repositories
- Funders embrace the FAIR data principles - implications for Data Management Planning?
- Slides: [www.slideshare.net/OpenAIRE\\_eu/openaire-webinar-on-open-research-data-in-h2020-oaw2016](http://www.slideshare.net/OpenAIRE_eu/openaire-webinar-on-open-research-data-in-h2020-oaw2016)
- Q&A document: <https://www.openaire.eu/public-documents?id=843&task=document.viewdoc>

[How to write a DMP](#), Sarah Jones & Marjan Grootveld, 7/14 July 2016:

- What is a Data Management Plan and why to write it?
- Example DMPs in different domains, with lots of links!
- Guidance, e.g. storing  $\neq$  archiving; how to find a repository; file-naming conventions

[FAIR data in Trustworthy Data Repositories](#), Peter Doorn & Ingrid Dillo, 12/13 December 2016:

- Proposal for scoring datasets on Findability, Accessibility and Interoperability = Reusability levels
- Inspired by the Data Seal of Approval criteria for Trustworthy Data Repositories
- Slides:  
<http://www.slideshare.net/EUDAT/fair-data-in-trustworthy-data-repositories-webinar-1213-december-2016>

# Any more questions?

Thanks to all who submitted questions when registering for the webinar.

And enjoy tomorrow's webinars!

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

 @openaire\_eu

 facebook.com/groups/openaire

 linkedin.com/groups/OpenAIRE-3893548

 [marjan.grootveld@dans.knaw.nl](mailto:marjan.grootveld@dans.knaw.nl)