Introduction of the team

The session presenters today are:
» Tiemen Folkers
» Jitka Vavra

We are part of the Research Data Office (RDO)
Researchdata@rug.nl
Our road-map for today

› Duration: approx. 90 min.

› **Part 1 (40 min)**
  › Research data management: what is it and why it is useful
  › One common ground: the FAIR data principles

› **Lunch + Are you FAIR ready? (20 min)**

› **Part 2 (30 min)**
  › Close look: the NWO data management plan
  › Questions
Student offered thief to keep laptop, and take $1,000, but return Ph.D thesis

FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you $1000 and use whatever way you offer to send you the money. The price is

Two years of research, gone

A burglary in the SSH house on Plutolaan has cast a sinister question mark over the future of PhD student Xiaoniu Huang. The thief made off with her laptop – and two years of PhD research.
“The commission says to be hindered by the ‘unstructured research practice’ which makes it difficult or impossible to reproduce or check the conclusions based on the research data”
 RESPONSIBLE RESEARCH DATA MANAGEMENT AND THE PREVENTION OF SCIENTIFIC MISCONDUCT

Introduction

Studying the role of the immune system in the interaction between mental and physical health is challenging. To study individuals with an intensive, longitudinal study design that requires repetitive sampling in their daily life, non-invasive sampling techniques are a necessity. Urine can be collected in a non-invasive way, but this may be demanding for participants and little is known about fluctuation of inflammatory markers in urine over time. The aim of this study was to investigate the feasibility of non-invasive sampling, and to explore intra-individual differences in inflammatory markers in urine.

Materials & methods

Ten healthy individuals collected 24-hour urine for 63 consecutive days. In a pilot analysis, 39 inflammatory markers were examined for detectability in urine, stability over time and under storage conditions, and daily fluctuations. Multiplex analyses were used to quantify levels of eight selected markers: C-reactive protein (CRP), Fractalkine, Interleukin-1 receptor-antagonist (IL-1Ra), Interferon-α (IFNα), Interferon-γ (IFNγ), Interferon gamma-induced protein 10 (IP10), Macrophage inflammatory protein-1α (MIP-1α), and Vascular Endothelial Growth Factor (VEGF). Cross-correlations were calculated between the overnight and 24-hour samples were calculated, to examine whether 24-hour urine could be replaced by the overnight portion for better feasibility. We examined intra- and interindividual differences in the levels of inflammatory markers in urine and the fluctuations thereof.
What is research data?
Please give some examples of research data

Join by Web

1. Go to PollEv.com
2. Enter RDO
3. Respond to activity

Join by Text

1. Text RDO to 22333
2. Text in your message

Total Results: 0
What is research data?

- Spreadsheets, field notebooks, diaries, photographs, spectra, questionnaires, artifacts, models, etc..
What is research data?

“Research data is data that is collected, observed, or created, for purposes of analysis to produce original research results.”

(Source link)

"the material underpinning a research assertion"

(Laura Molloy, University of Glasgow)
Why manage your research data?
Why manage your research data?

Join by Web

1. Go to PollEv.com
2. Enter RDO
3. Respond to activity
Reasons for data management

› To do sound and solid research
› To make sense of data
› To make a budget
› Take measures to prevent, e.g.:
  • data loss
  • unauthorized access & data breaches
› To get credit
› To work uniformly in collaborations
› Requirements of institutions, funding agencies and journals
› **Bottom line: It’s part of the job!**
Privacy and research data

Are you aware...
that there are people behind data?

Become a leader in the field of responsible research
RDO can support you!
Find out more at www.rug.nl/researchdata

*GDPR: General Data Protection Regulation
GDPR

› Applicable when processing **personal data**
› Requires **data minimization**
› Anonymize/pseudonymize personal data (describe)
› **Inform participants**: processing personal data & their rights
  • Get Consent!
GDPR

When you are processing personal data...

- You have to protect the **rights** and the **freedom** of the participants.
- Evaluate the possible **risks** and provide **safeguards** to avoid or limit these risks.
- Be **transparent** with the participants about the use of their data.
GDPR

This isn’t just a legal matter...

› Careful handling of data ensures citizens’ confidence in scientific research and reduces the risk of reputational damage.
› Compliance with GDPR is explicitly requested by funders (such as H2020, ERC, NWO).
› Journals are also increasingly asking for compliance with the GDPR.

Become a leader in the field of responsible research!
The aim of good data management is...

- Work consistently, efficiently & effectively
- Have a clear plan for now and in the future
  - Well-saved and well-documented
  - Reverification & re-use
  - Be accountable
Use a template

> DMP’s or checklists:
  - RDMP Webtool for faculty templates (UG-only)
  - DMP online
  - DMO – University of Helsinki

> Funder will provide you with a template
  - ZonMW, NWO, ERC, H2020, ....

> Collaborating? Maybe a plan already exists

Even though the templates are different, they are based upon the same 4 principles.....
Findable

Accessible

Interoperable

Reusable
“ZonMw promotes FAIR data management and stewardship to make your research data verifiable and reusable, and enhance the quality of your research”

“research data that emerges from NWO-funded research should be as accessible and reusable as possible”
**F**indable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

**A**ccessible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

**I**nteroperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

**R**eusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

https://www.cessda.eu
The FAIR guiding principles:

<table>
<thead>
<tr>
<th>To be Findable</th>
<th>To be Accessible</th>
<th>To be Interoperable:</th>
<th>To be reusable</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1. (meta)data are assigned a globally unique and persistent identifier</td>
<td>A1. (meta)data are retrievable by their identifier using a standardized communications protocol</td>
<td>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation</td>
<td>R1. (meta)data are richly described with a plurality of accurate and relevant attributes</td>
</tr>
<tr>
<td>F2. data are described with rich metadata (defined by R1)</td>
<td>A1.1. the protocol is free, open and universally implementable</td>
<td>I2. (meta)data uses vocabularies that follow FAIR principles</td>
<td>R1.1. (meta)data are released with a clear and accessible data usage license</td>
</tr>
<tr>
<td>F3. metadata clearly and explicitly include the identifier of the data it describes</td>
<td>A1.2. the protocol allows for an authentication and authorization procedure, where necessary</td>
<td>I3. (meta)data include qualified references to other (meta)data</td>
<td>R1.2. (meta)data are associated with data provenance</td>
</tr>
<tr>
<td>F4. (meta)data are registered or indexed in a searchable resource</td>
<td>A2. metadata are accessible, even when the data are no longer available</td>
<td></td>
<td>R1.3. (meta)data meet domain relevant community standards</td>
</tr>
</tbody>
</table>

https://doi.org/10.1038/sdata.2016.18
What is Metadata?

A photo

[Image of metadata information]

https://dataedo.com/kb/data-glossary/what-is-metadata
NO METADATA
NO FUTURE
Lunch + Are your FAIR ready?

http://rug-nl.libwizard.com/RDM
Questions about the quiz?
How to make a plan
What is in the plan?

› Storage of data during and after
  ▪ what kind of formats
  ▪ the size and the shape (digital and physical)
  ▪ Where: repository?
  ▪ Costs
› Personal data?
› Collaborators?
  ▪ internal, national, international
  ▪ commercial
› How is the data acquired
  ▪ created yourself
  ▪ reuse
  ▪ external source: third party
  ▪ consent?
What is in the plan?

› Will the data be reusable?
  ▪ How?
  ▪ Conditions?
  ▪ “Ownership”

› How will the data be made findable
  ▪ metadata
NWO template model

› We are going to use the NWO dmp template as a model
› All dmp’s more or less want you to demonstrate how you uphold the FAIR model

(General Information)
1. Description of the data
2. Data storage: during research
3. Data storage: after the research
4. Standards and metadata
5. Making data available
1. Description of the data

› Describe the data that will be collected/generated within the project.
   ▪ data flow chart

› Specify the type and format of the data.
   ▪ data table

› It’s not only about digital data!
   ▪ written questionnaires, soil samples, etc.
2. Data storage: during research

› What is the volume of the data and where will the data be stored?
› Is there sufficient storage capacity during the project? Y/N
› Will the data be backed-up regularly during the project?
  ▪ Who is responsible for this?
› What are the expected costs? Please specify these and state an amount that is as realistic as possible. How will these costs be covered?
› Who has or has to have access?
› Special storage for raw data?
3. Data storage: after the research

› Specify in which trusted repository the data will be stored after the project.
› If the data will not be stored in a trusted repository how will the data be made findable, accessible and reusable?
› Will a persistent identifier be used to make the data findable? Y/N
› For how long will the data be archived?
› What are the expected costs? Please specify these and state an amount that is as realistic as possible.
   How will the costs be covered?
› Who is “owner” of the data?
4. Standards and metadata

› How will the data be documented? What metadata standard will be used to make the data accessible and reusable?

› Metadata is "data that provides information about other data".
  ▪ Dublin Core
  ▪ Machine readable

› Used for
  ▪ dataset - determined by the repository
  ▪ explaining the data itself - discipline specific
5. Making data available

› Are the data available for reuse after the project?
  ▪ If not, please explain why the data are not suitable and/or available for reuse.

› If data are only made available after a certain period then please state the reason for this.

› If part of the data cannot be made (directly) available then please state the part concerned.

› Are there any conditions for the reuse of the data?
  ▪ If so, are these conditions defined in a (consortium) agreement?
Further information

› FAIR
› Research Data Management Plans
› Institute or Faculty RDM policy
› ICT SOLUTIONS (storage, infrastructure, access, collaboration)
  ▪ Virtual Research Workspace (VRW)
  ▪ DataverseNL
  ▪ Portfolio Tools & Services
› METADATA
  ▪ What is Metadata
  ▪ Disciplinary Metadata
› GDPR
  ▪ Consent
  ▪ Data Protection Impact Assessment (DPIA)
Questions?

researchdata@rug.nl