WELCOME TO THE PI-MEETING

Marjukka Myllärniemi, vice-dean for research
Alise Hyrskyluoto, Senior Advisor
Medical Faculty
PI MEETING, FEBRUARY 25, 2021
- DATA PROTECTION & DATA STORAGE FOR RESEARCHERS -

14:30 – 16:00

**Opening words**
Vice-Dean in Charge of Research,
Professor Marjukka Myllärniemi

**Examples of research data use**
Professor Mika Kivimäki

**Data protection and research**
Lotta Ylä-Sulkava, Legal counsel, Research services

**Research data services in Meilahti – storing sensitive data**
Pekka Hintsanen, IT specialist, Centre for Information Technology

After the event, materials will be added to:
LEGISLATION IS CHANGING - THIS PROMOTES A CHANGE IN RESEARCH PARADIGM

• GDPR - 2018 (data protection EU)
• https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=CELEX:32016R0679&from=FI
• Law on data protection (Finlex)
• Law on the secondary use of health care data 1.4.2020

-> HY and HUS have given statements related to the fees and the use of the law
WHAT DOES THIS MEAN IN PRACTICE?

- All use of sensitive data has requirements – safe environment, strong identification of data and log data on all use, export and views of data
- Pseudonymized is not *anonymized* only the latter is not sensitive
GDPR gives us requirements – how do we answer to these requirements? Unorthodoxically saved sensitive data presents a threat to the safety of patients/healthy volunteers.

- HUS has a registry of research permissions.
- HY researchers have no permission process, only ethical statements, therefore no registry is formed.
• Medical faculty reserarch council (tutkimusneuvosto) had formed a working group for prioritizing IT-infrastructures and acquisitions – **IT for life science**

• Medical faculty has obtained infrastructures for NAS-hardware IT for life science prioritizes new acquisitions -> to be continued

• We have made a suggestion to TINE to form a faculty-specific registry out of the privacy notices (tietosuojailmoitus)

NEXT PI MEETING 19.5.2021 at 2.30pm

CYBER SECURITY

• Speakers including Catharina Candolin, Cyber Security expert from OP Financial Group

• Detailed program will be announced closer to the event
PLEASE GIVE US FEEDBACK OF THE MEETING IN E-FORM:

https://elomake.helsinki.fi/lomakkeet/109824/lomake.html

THANK YOU!
Examples of research data use

The challenge of combining two important principles in science:

(1) Data protection (“Everyone has the right to the protection of personal data concerning him or her”, GDPR).

(2) Data sharing and Open Science (FAIR-principles for scientific data management and stewardship)
Funders want your study

• To be Findable
• To be Accessible
• To be Interoperable
• To be Reusable

The Whitehall II study follows the FAIR principles
Overview of Data Flow in the Whitehall II Study

Data sharing applicants
- Doubly pseudonymised research data*

ONS, NHS Digital, SAIL, NHS Scotland
- Cancer registration
- Death certificates
- Hospital Episodes Statistics

SLMS Data Safe Haven
- Personal and admin data processing, de-identification and analysis
- Analysis of pseudo-anonymised HES records

Clinical examination

Study participants
- Questionnaires
- Results letters

Key
* selection bespoke per each application
SLMS = School of Medical and Life Sciences
** limited to broad Yes/No groupings of disease categories (≥100) together with year of diagnosis/death

WH2 Research data (read only)
- Folder Access restricted to WH2 researchers
- N drive on UCL network

WH2 Researcher selected project-specific research data for analyses
- S drive on UCL network

Pseudonymised questionnaire & clinical exam data

Limited data derived from cancer, death and HES **
Whitehall II
(also known as the Stress and Health Study)

Study update: Temporary Suspension of Stress and Health Study (Whitehall II) Clinics.

Due to the increasing risk of infection with Covid-19 (Corona virus) we are temporarily halting the Stress and Health Study clinics, from Monday 16th March until further notice.

The decision to close our clinic is in line with measures introduced by the NHS for hospital outpatient clinics. We will contact you to rearrange your appointment once the government health authorities confirm that the infection risk is reduced and the public health guidance indicates when the clinic can re-open.

We apologise for any inconvenience this situation may cause you. If you have any questions, please contact us on our Freephone number 0800 068 1562 or by email shstudy@ucl.ac.uk or

General enquiries:
Dr Tingde Kais
Project Manager
Email: j.k.kaio@ucl.ac.uk

Study participants:
Ms Stephanie Smith
Study & Data Co-ordinator
Tel: 0800 068 1562 or
020 7679 5621
Email: shstudy@ucl.ac.uk or whitehall2@ucl.ac.uk

Data sharing enquiries:
Email: whitehall2@ucl.ac.uk

Location:
Please note that data sharing applications sent directly to the Whitehall team may be subject to delays due to the measures being put in place to control the spread of Covid-19. We thank you for your patience.

We are committed to maximizing the use of Whitehall II data to advance scientific knowledge.

We welcome proposals for collaborative and external projects for bona fide research. We make data as widely available as possible while strongly protecting confidentiality, and making sure that we maintain the reputation of the study, its funders and its participants. All proposals are reviewed by the data sharing committee.

For aggregate data, please contact the researcher you wish to collaborate with.

For individual-level data, we follow a controlled access model for sharing, which is described in our Whitehall II data sharing policy (PDF). This policy conforms to the MRC Policy on Research Data Sharing.

For general data sharing enquiries, please contact whitehall2@ucl.ac.uk
For projects involving genetics, please contact Dr Meena Kumari.

Please note that we charge a modest fee to cover the administrative costs associated with the management and generation of the bespoke datasets.
Applying for data

For individual level data, please submit the following two documents:

1. the WII Data Sharing Application Form in word format (also in PDF).

2. the WII Excel data dictionary with the list of variables needed for the project. Please highlight only the variable names needed for your project.

The WII research team meets monthly to evaluate all applications received. Upon approval of an application, applicants will be asked to sign the WII Data Sharing Agreement (last page of application form) and proceed with the payment.

Within two weeks of the receipt of this document and of the payment, the data manager will release an anonymised dataset, tailor-made for each project.

For aggregate data, please contact the researcher you wish to collaborate with.
Data documentation

You can obtain a detailed description of the WII variables in the following sources:

- Excel data dictionary
- Questionnaires
- Clinical Measures

On request, we can supply documentation such as copies of other WII questionnaires (FFQ, MMSE, GKQ, cortisol diaries, etc), clinical protocols and syntax of derived variables.
7. DATA REQUESTED

7.1 DATA FORMAT NEEDED

Please indicate what format you will be using the data (SAS, SPSS, Stata, csv):

7.2 DATA REQUIRED FOR ANALYSIS

1) Download the WHII data dictionary from the Whitehall II website (www.ucl.ac.uk/whitehallll/data-sharing). It is an Excel file containing the exact variables names for every collection phase.

2) **Highlight** all of the variables needed. The list of highlighted items and data collection phases must be consistent with the project proposal.

3) Send the highlighted WHII data dictionary file at the same time as this application form.
Whitehall II Data Application Form

Please note that a separate application form is required to access WHII Genetics data.

Preferred member of the WHII team to be the WHII Contact Researcher (if any):

1. PRINCIPAL INVESTIGATOR (PI)

For PHD projects, the PI on the application must be the supervisor, not the student. Please do not enter the name of the student as the PI.

<table>
<thead>
<tr>
<th>Title, forename, surname:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Address:</td>
</tr>
<tr>
<td>Telephone:</td>
</tr>
<tr>
<td>Email:</td>
</tr>
</tbody>
</table>

Please attach the PI's Curriculum Vitae.

2. RESEARCH TEAM / CO-APPLICANTS

Details of each Research Team member involved in the proposed project who will have access to the data.

<table>
<thead>
<tr>
<th>Name of co-applicant</th>
<th>Employing organisation</th>
<th>Position in organisation</th>
<th>Email address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.1 PROJECT TITLE

6.2 SUMMARY
A brief summary of up to 200 words describing the aims of the study/research project.

6.3 CONTEXT
Where research is part of a larger programme, please give details.

6.4 PROJECT DESCRIPTION
Full description of the purpose/s for which the data are requested (maximum 4 A4 sides).
- Background

6.5 PLANNED SCIENTIFIC OUTPUTS
Intended outputs/publications arising from the use of these data, including abstracts, posters and
Whitehall II Policy on Data Sharing

Information for researchers interested in analysing existing Whitehall II data

Whitehall II Study
Department of Epidemiology & Public Health
University College London
1-19 Torrington Place
London WC1E 6BT
email: whitehall2@ucl.ac.uk

March 2018
Whitehall II Data User’s Agreement

To be signed once the data sharing application has been approved.

Main applicant

Title, forename, surname: ...............................................................
Work Address: .................................................................
...............................................................................................
...............................................................................................
Telephone: .................................................................
Email: ...........................................................................
Project title: ........................................................................

I agree that my project will use the requested Whitehall II data and is to be conducted according to the contents of the document “Whitehall II Policy on Data Sharing”.

I agree that both myself and my collaborators will abide by the terms and conditions outlined there within, also summarised in the next page.

I agree that failure to comply with these terms and conditions will result on any future data sharing applications from me and/or my collaborators being refused by the WHII team.

Signature of main applicant: ............................................................
Name in block capitals: ............................................................
Date: ...........................................................................

Please sign, date and return to the WHII Contact Researcher in electronic or paper form.
The Data Sharing Committee (DSC) comprises the Whitehall II study Director and PI, co-Investigators and associated senior study researchers, statisticians and the core operational team.

The committee meet once a month to review the scientific merit and the suitability of the Whitehall II Study, for the data sharing requests as well as the ability of the applicants(s) to deliver on the project. The proposals can be approved fully, or approved pending further information or rejected, with an explanation to the applicant.

Data Sharing Committee Members

Prof Mika Kivimaki, Director and PI, The Whitehall II study
Professor Eric Brunner, (Co-PI)
Professor Archana Singh Mantoux, (Co-PI)
Professor Jenny Head, (Statistician)
Professor Annie Britton, (Epidemiologist)
Professor Meena Kumari, (Genetics studies)
Dr Martin Shipley, (Statistician)
Dr Adam Tabak, (Clinician and Senior Research Associate)
Dr Severine Sabia (Senior Research Associate)
Dr Jatinderpal Kalsi, Mrs Beverley Milne, Mrs Stephanie Smith (Core Study Team)
Data sharing FAQ

Q: Who can apply for Whitehall II data?

Q: What are individual level data and aggregate data?

Q: What does bona fide research mean?

Q: Why do I really need to fill in an application form, even if I am a close collaborator?

Q: What do I need to do to apply for Whitehall II data?

Q: I am a PhD student, can I apply for data?

Q: When will the latest wave of data collection be available?

Q: How long does it take to receive the data?

Q: Who should I contact if I have any queries?

Q: Could my project be turned down?

Q: Can I request extra variables?

Q: How do you ensure that the released data are anonymised?
• Approval rate:
• 227 at first submission
• 14 re-submission requested (10 received)
• 1 refused
Evidence from Whitehall is used in clinical guidelines and policy documents

<table>
<thead>
<tr>
<th>No.</th>
<th>Disease area</th>
<th>Guideline/Publication Title</th>
<th>Whitehall II studies used as evidence in the guideline/policy document</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Topic</td>
<td>References/Notes</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>


| 5.  | Dementia/Cognitive decline   | NICE guideline [NG16] Dementia, disability and frailty in later life – mid-life approaches to delay or prevent onset. Published date: 20 October 2015.  


Big data to reduce type 1 and type 2 errors.
Systolic blood pressure in relation to CVD mortality rates:
the Prospective Studies Collaboration

Rory Collins
Systolic blood pressure in relation to CVD mortality rates: the Prospective Studies Collaboration

Rory Collins
Systolic blood pressure in relation to CVD mortality rates: the Prospective Studies Collaboration

Rory Collins
Providing access to research data rather than sharing data
The Dementias Platform UK (DPUK) Data Portal

Sarah Bauermeister1 · Christopher Orton2 · Simon Thompson2 · Roger A. Barker3 · Joshua R. Bauermeister1 · Yoav Ben-Shlomo4 · Carol Brayne5 · David Burn6 · Archie Campbell7 · Catherine Calvin1 · Siddharthan Chandran8 · Nishi Chaturvedi9 · Geneviève Chêne10 · Iain P. Chessell11 · Anne Corbett12 · Daniel H. J. Davis9 · Mike Denis13 · Carole Dufouil10 · Paul Elliott14,39,40 · Nick Fox15 · Derek Hill16 · Scott M. Hofer17 · Michele T. Hu18 · Christoph Jindra1 · Frank Kee19 · Chi-Hun Kim1 · Changsoo Kim20 · Mika Kivimäki21 · Ivan Koychev1 · Rachael A. Lawson22 · Gerry J. Linden19 · Ronan A. Lyons2 · Clare Mackay1 · Paul M. Matthews23 · Bernadette McGuinness19 · Lefkos Middleton14 · Catherine Moody24 · Katrina Moore15 · Duk L. Na25 · John T. O’Brien26 · Sebastien Ourselin27 · Shantini Paranjothy28 · Ki-Soo Park29 · David J. Porteous30 · Marcus Richards9 · Craig W. Ritchie8 · Jonathan D. Rohrer15 · Martin N. Rossor15 · James B. Rowe3 · Rachael Scähill15 · Christian Schnier31 · Jonathan M. Schott15 · Sang W. Seo25 · Matthew South1 · Matthew Steptoe12 · Sarah J. Tabrizi15 · Andrea Tales33 · Therese Tillin34 · Nicholas J. Timpson4 · Arthur W. Toga35 · Pieter-Jelle Visser36 · Richard Wade-Martins37 · Tim Wilkinson31 · Julie Williams38 · Andrew Wong39 · John E. J. Gallacher1

Received: 14 February 2020 / Accepted: 10 April 2020
© The Author(s) 2020

Abstract
The Dementias Platform UK Data Portal is a data repository facilitating access to data for 3 370 929 individuals in 42 cohorts. The Data Portal is an end-to-end data management solution providing a secure, fully auditable, remote access environment for the analysis of cohort data. All projects utilising the data are by default collaborations with the cohort research teams generating the data. The Data Portal uses UK Secure eResearch Platform infrastructure to provide three core utilities: data...
Welcome to the Data Portal

The DPUK Data Portal brings together records of over 3 million people in a free-to-access resource.

Researchers can identify which cohorts are relevant to them, apply for access to the data and then analyse it in a secure, remote environment complete with data linkage and analysis packages.

• To be Reusable
Researchers access and analyse data in DPUK safe-haven server (data sharing is avoided; standardized data protection)
Log files from data analyses are documented (evidenced appropriate use of the data)
There is a possibility to close access for non-compliant researchers
<table>
<thead>
<tr>
<th>ACRONYM</th>
<th>FULL NAME</th>
<th>INSTITUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIMROD</td>
<td>Neuroinflammation and Inflammation in Memory and Other Disorders (the NIMROD Study)</td>
<td>University of Cambridge</td>
</tr>
<tr>
<td>OPDC Discovery</td>
<td>Oxford Parkinson's Disease Centre Discovery Cohort</td>
<td>University of Oxford</td>
</tr>
<tr>
<td>PICNICS</td>
<td>Parkinsonism: Incidence and Cognitive heterogeneity in Cambridgeshire</td>
<td>University of Cambridge</td>
</tr>
<tr>
<td>PREVENT</td>
<td>The PREVENT Research Programme</td>
<td>West London Mental Health Trust, Imperial College London and University of Edinburgh</td>
</tr>
<tr>
<td>PRIME</td>
<td>PRIME (étude PRospective sur l'Infarctus du MyocardE)</td>
<td>Queen's University Belfast, Institute of Clinical Sciences</td>
</tr>
<tr>
<td>PROTECT</td>
<td>Platform for Research Online To investigate gEnetics and CogniTion and ageing</td>
<td>King's College London</td>
</tr>
<tr>
<td>SABRE</td>
<td>Southall And Brent REvisited</td>
<td>University College London</td>
</tr>
<tr>
<td>SMC Amyloid PET</td>
<td>Samsung Medical Center Amyloid PET Cohort</td>
<td>SAMSUNG MEDICAL CENTER</td>
</tr>
<tr>
<td>TRACK HD</td>
<td>TRACK HD</td>
<td>University College London</td>
</tr>
<tr>
<td>UK Biobank</td>
<td>UK Biobank</td>
<td>UK Biobank</td>
</tr>
<tr>
<td>Whitehall II</td>
<td>Whitehall II</td>
<td>University College London</td>
</tr>
</tbody>
</table>
DATA PROTECTION FOR RESEARCHERS

PI MEETING 25.2.2021 MEILAHTI
WHAT IS BEHIND THIS GDPR HAZE?
WHAT IS PERSONAL DATA?

- Name
- Address
- Opinion
- IP Address
- Picture
- Location
- Feature
- Behaviour
- And so on...

Any information that can be connected to an individual.
SPECIAL CATEGORIES OF PERSONAL DATA

- Also called as "sensitive data"
- As defined in the regulation:
  - Racial or ethnic origin
  - Political opinions
  - Religious or philosophical beliefs
  - Trade union membership
  - Genetic data
  - Biometric data when it is processed for the purpose of uniquely identify a person
  - Data concerning individuals health
  - Data concerning the persons sexual life or sexual orientation
  - Data relating to criminal convictions and offences

It is allowed to process special category data in research but the data must be secured more carefully.
PSEUDONYMISATION, ANONYMISATION AND UNDIRECT IDENTIFICATION

Undirect identification
In the light of the regulation, data can be personal data even though you are not processing direct identifiers such as name or social security number if individuals can be identified from the data by other means.

- F.ex. In cases where you are processing information about persons rare disease or feature.
- Or if you are collecting lot of data from an individual so that the identity can be concluded from it.

Pseudonymised data
- Pseudonymised information is still personal data!
- Direct identifiers (such as names) are removed or replaced with other identifiers (like number).
- It is still possible to restore the data to be identifiable but it requires additional information.

Anonymised data
- It is not possible by any means to connect the information to an individual.
- There remains quite often the risk that the data is still identifiable.

Presentation Name / Firstname Lastname 25/02/2021 5
WHO IS RESPONSIBLE FOR DATA PROTECTION?

**Controller** is the one who is responsible for the legal compliance and possible damages to data subjects.

**Processor** Analyses, stores, destroys or by other means processes personal data on behalf of the controller.

**Joint controller:** Two organisations and/or persons determine the purposes and means for the processing together.
WHO IS A CONTROLLER – UNIVERSITY OR THE RESEARCHER / STUDENT?

University of Helsinki

- When the research project is approved by the University and the funding for the research is allocated to the University ("UH's research project")
- When the research is performed under an employment relationship with UH
- UH is the controller for research materials or other research result that includes personal data when it is disclosed or collected to UH to be used within university's research project etc.

Reseacher / Student

- Researcher/student, who is not in employment relationship with UH or is not performing the research as part of UH's research project but is performing the research independently
  - Defines the purposes and means for the research and collected personal data independently
  - He/she still has designated mentor/couch at the University but is working independently
“The principal lead of a research project is responsible for ensuring that the project complies with data protection legislation and these data protection principles. In addition, it is their duty to ensure that researchers are trained in the practices of processing personal data before the processing commences. The principal lead specifies the responsibilities and obligations for processing of personal data in research materials based on the roles of the staff (the person responsible/contact person/processor) and, where applicable, in accordance with the division of responsibilities described above.”
GENERAL PRINCIPLES

Lawfulness, fairness and transparency

- Process personal data only when you have a legal mandate to do so (f.ex. and usually: Research carried out in public interest)
- Be fair towards the research participants and other examinees
- Be open about the processing of personal data

Purpose limitation

- Use the data only for purposes it was collected and in a manner that was informed to data subjects
- Data can be used for future research projects if the purpose for processing is "compatible" with the primary purpose and if the data subjects can assume that the data can be used for other purposes also (informing!) and/or if the possible data permits allows

Data minimisation

- Collect only the data you need for the research
- Erase all data you don't need anymore (also pseudonymisation and anonymisation)
GENERAL PRINCIPLES

Accuracy

- Processed personal data should be accurate and kept up to date where necessary

Storage limitation

- Store personal data for as long it is needed
- Personal data can be stored for future research purposes also (be transparent towards the data subjects and comply with these general principles)

Integrity and confidentiality

- Protect the data adequately, towards unauthorized access, destruction or damage
Accountability means that you should be able to demonstrate that you comply with the regulation.

You can do it by documenting / doing:
- Data management plan (incl. information on data collection, informing of data subjects, processing and protection up until the destruction/archiving of the data)
- Data protection preliminary evaluation to see whether you need to perform a Data protection impact assessment (DPIA) + DPIA when needed
- Data protection statement and possible consent forms
- Agreements with partners (incl. possible agreements on the international transfer of data)
DIFFERENT KIND OF CONSENTS

- Informed consent *for participation* to medical research (regulated by Medical Research Act, and regarding clinical drug research by EU-regulation and national law)
- Consent to collect and store an organ, tissue or cells (regulated by Kudoslaki)
- Biopankconsent (regulated by Biopank Act)

- Consent *for processing personal data* (regulated in the GDPR)
  - you usually do not need to ask this because the processing of personal data can be based of "research carried out in public interests" (and it is highly recommended)

- As we speak the regulation is evolving around medical research…
When you are planning your research, think of the risks to data subjects and plan how to mitigate them.

Define also how you will implement the general principles.

Risks for the data are f.ex. Accidental or unauthorized

- destruction,
- losing,
- alteration,
- disclosure/leakage

1) Analyse what kind of risks there are or may be to data subjects
2) Define the measures for minimising these risks

It is advisable also to check and update the measures later on if needed - are the safeguards enough and up to date?

Factors affecting the risk level are f.ex.:
- Processed data
- Amount of data
- Focus group
- Partners
- Resources and tools used
POSSIBLE RISKS & IMPACTS TO DATA SUBJECTS

For example:
• Bodily harm
• Loss of freedom of movement
• Loss of control over the purposes of processing of his/her personal data
• Fraud
• Negative impact to employment
• Sosial damage/damage to reputation, embarrassment

• Nuisance
• Verbal abuse
• Feeling of insecurity
• Inability to exercise rights
• Loss of freedom of political/religious opinions or beliefs
• Discrimination
POSSIBLE MITIGATION ACTIONS

- Safeguards can be, f. ex:
  - Pseudonymisation, anonymisation
  - Encryption
  - Aggregation
  - Using encrypted email/connection when sending sensitive data
  - Contracts
  - Instructions for persons processing the data
  - Defining access rights
  - Using systems that logs the processing of data
  - And so on…
The idea is to identify and analyse the risks that the processing of the personal data causes to data subjects.

- Helps you to manage the risks to data subjects.

In practice:
- The processing of personal data is described.
- The necessity and proportionality of personal data processing is analyzed.
- The risks to data subjects' rights and freedoms are identified and analyzed.
- Needed measures to mitigate the risks are defined.

The main point is to analyze and argument (plus document) whether the processing is acceptable in the first place.

(You can find the template from FLAMMA).
INTERNATIONAL TRANSFERS

- Transfer of personal outside of EU is restricted
- EU-Commission have declared some countries to have adequate level of data protection: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en
- If you are transferring or disclosing personal data (NB! Even pseudonymous data) outside of EU and if the Commission have not declared that the country have adequate level of data protection you need extra safeguards to protect the data and to transfer the data legally
- For example and usually: Commissions Standard Contractual Clauses
  - New standard Contractual Clauses coming.
  - "New" EU ruling have turned international transfers upside down and it led to stricter requirements for controllers to know where the data is going (how does the country respect the right to privacy and data protection)
    - SCC may not be enough -> you may still need extra safeguards (f.ex. pseudonymization, encryption)
- You may also ask a explicit consent for the transfer
  - The data subject need to be informed about the possible risks when the data is disclosed to a country that does not provide adequate level of data protection
- Research lawyers (and DPO) can help you with international transfers!
INFORMING

- Medical research:
  - Requirements coming from: Tutkimuslaki, tutkimusasetus (+kudoslaki)
  - Regulates how you need to inform about a research that includes intervention
  - Concentrates more to the possible impacts of the research measures than data protection
  - *Regulation is under renewing...*
- TENK also provides some criteria for informing (from ethical point of view)
- GDPR: regulates what you need to inform about the processing of personal data
  - UH have template for privacy notice that you can use: FLAMMA
    - Included guinace and all the questions you need to answer.
• Personal data breach = "Breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed"

• Contact tietoturva@helsinki.fi immediately if you suspect a data breach has occurred.
  • F.ex. Your computer or username and password gets stolen, you have lost your USB drive etc.

• In urgent matters, please contact a member of the data security team directly

• Must be notified to the Finnish data protection authority within 72 hours and in many cases also to the data subjects
ACT ON SECONDARY USE OF HEALTH AND SOCIAL DATA

- Regulates the use of register data stored in a social and health care providers registers for research, innovation and education purposes
  - Only aggregated data can be given to innovation and development activities
- You need a data permit to process "sote-data" for research
  - From Findata if you collect the data from more than one controller
- 1.5.2021 requirements for the secure processing environment will apply when you want’t to process individual level data
  - register data may be processed only in environments that comply with Findata’s requirements.
  - Researcher may also use Findata's own remote access environment.
- More information: https://www.findata.fi/
WHERE TO FIND INFORMATION?


• Researchers guidance for processing of personal data and tools for implementing the requirements: https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/tutkimuksen-tietosuoja-asiat

• General information about the regulation and processing of personal data: https://flamma.helsinki.fi/en/group/turvallisuus/tietosuoja

• Information Security: https://flamma.helsinki.fi/group/turvallisuus/tietoturva

• UH datasupport: –https://datasupport.helsinki.fi/

• National Authority: http://www.tietosuoja.fi
CONTACT DETAILS

Research related questions: researchlawyers@helsinki.fi

UH Data Protection Officer: tietosuoja@helsinki.fi

Helpdesk: helpdesk@helsinki.fi

Information Security Team: tietoturva@helsinki.fi

Datasupport team: datasupport@helsinki.fi
RESEARCH DATA SERVICES IN MEILAHTI

STORING SENSITIVE DATA

PI meeting 25.2.2021
Pekka Hintsanen
IT Center / IT For Science / ONCOSYS
STORING SENSITIVE DATA

- Helsinki University Offering
- CSC (IT CENTER FOR SCIENCE; owned by the Finnish state and higher education institutions) Offering
- others
STORING SENSITIVE DATA: UNIVERSITY OF HELSINKI OFFERING

- **Umpio**
  - Designed for sensitive data
  - **Access only from secure VDI machine**
  - For research groups / personal use

- **Medical faculty owned storage**
  - Designed for sensitive data, fullfills GDPR requirements
  - Access from university workstations, CSC ePouta and university servers
  - "Cheap", subvented by medical faculty, ~50€ / TB / storage lifecycle (~5 years)
  - No backups, resilient, snapshots
  - For research groups

- **Storage attached to server**
  - physical server
    - Max 200TB, price > 8000 k€
  - Virtual server
    - Max 10TB, price for standard disc 0,11€ / GB / year ( > 100 € / TB / year)
  - For research groups

- **group folder (P drive)**
  - Max 10TB, price 0,019€ / GB / month (~ 200€ / TB / year )
  - e.g. moderate risk personal info, pseudonymized
  - For research groups

- **home folder (Z drive)**
  - Max 50GB, free, personal
STORING SENSITIVE DATA: CSC OFFERING

• Computing storage
  - Epouta is **computing** environment for sensitive data, currently the only option available
  - storages:
    1) Work storage for computing; from local ePouta infrastructure
    2) Long term storage (NetApp storage)
      - In addition to use in ePouta, can be connected to university servers also
      - Can also be used for computing

• Long term storage
  - NetApp storage (see above)
  - Fairdata-PAS (PAS = P**itkääkaisSäilytys** = Long term storage)
    - ministry of education and culture owns, CSC produces
    - for digital preservation of research data for several decades, or even centuries.
  - CSC about to launch new long term storage for sensitive data
    - In co-operation with EGA, national federated EGA node
    - Despite of name, not only for genome data
  - Note! IDA is not suitable for sensitive data anymore
• Pilot project started: UH’s own computing cluster for sensitive data
  – Currently CSC ePouta only option when computing resources are needed
  – Epouta resources are in Kajaani; big latency ("delay") and slow I/O to faculty storage
  – Epouta computing resources are virtualized
  – Infrastructure: we reuse the retired Taito supercomputer HW
• Computing environments / storages:
  - https://wiki.helsinki.fi/display/it4sci/IT+for+Science+group
  - it4science@helsinki.fi

• Long term storage when the project is completed:
  - datasupport@helsinki.fi

• CSC: servicedesk@csc.fi (exception: in ePouta cases epouta-support@helsinki.fi)