

SELF-REFLECTION GUIDE Human-Related Big Data Research (MBDO)

This Self-Reflection Guide is a tool for researchers and research teams within the Faculty of Humanities (HUM), designed to invite them to self-reflect on their research plans, so that they can conduct non-contact Human-Related Big Data Research (MBDO) while abiding to the law and frameworks of scientific integrity. The goals are to ensure this type of studies are conducted in an ethical and lawful manner within the Faculty of Humanities, on the basis of a 'yes-but' model, taking into account ethics, privacy and data management at all stages of the research. In principle, MBDO research does not have to be assessed by the FETC-H. Plans for Big Data studies where researchers interact with data subjects should apply for ethical approval from the FETC-H.

Upon assignment by the Vice Dean for Research and Impact of the Faculty of Humanities, in 2023 the Self-reflection Guide was developed by a steering committee consisting of researchers, chairs of both FETC-H chambers, privacy officers, and data managers. This Self-Reflection Guide was established based on a comparative literature review of:

- A. The [Netherlands Code of Conduct for Research Integrity](#) en the advice [Big data in wetenschappelijk onderzoek met gegevens over personen](#) (KNAW, 2018, which includes [a summary in English](#));
- B. Existing procedures, guidelines and protocols used by research platforms (e.g. the Data School and the [Centre for Digital Humanities](#)), privacy officers and ethical review committees within the various faculties at UU (see e.g. [Tactful non-contact research. Manual for researchers](#), available in Dutch only) Dutch sister faculties and universities abroad (e.g. the [LSE Research Ethics Committee, 2022](#));
- C. Ethical guidelines established by international professional organisations (e.g. [Association of Internet Researchers](#)) and scholarly literature (e.g. D'Ignazio & Klein, 2020; Fleming & Bruce, 2021; Townsend & Wallace, 2016; Zimmer, 2018).

Definitions

Human-related: "any research in which data of participants are collected and which is carried out by researchers under the auspices of the Faculty of Humanities." ([Regulations FETC-H](#), section 4.1, p. 6).

Big data: datasets, originating from natural persons, that have already been published and/or collected elsewhere and where those persons (cannot) be contacted (directly) by the researcher (derived from [Regulations FETC-H](#), p. 1). This refers to data obtained from sources such as mobile phones, smart devices, social media, transactional data or sensors. Big data research does not focus on data from individual people but aggregates data points to establish patterns, trends or correlations through, for example, data mining, machine learning and statistical modelling.

Data subject: someone whose (personal) data are used for research. Within big data research, individuals whose data are used for research are not active 'participants', which is why we use the term 'data subject' here to refer to data that is traceable to an involved natural person ("data subject", General Data Protection Regulation (GDPR), article 4.1).

Contactless: research in which data from data subjects are collected in an unsolicited manner. Researchers are not in (direct) contact with data subjects about providing data and data subjects do not give direct consent to researchers (Read more about non-contact research in Gerritsen, [Tactful non-contact research](#), 2021, p. 1).

Basis: The GDPR article 6 outlines a total of 6 legal bases to collect and process personal data. In research the legal bases of 'General Interest' and 'Consent' are most commonly used. Within

this document, we assume that 'General interest' is chosen as the legal basis for doing human-related big data research, as direct contact with participants is often not possible or difficult to achieve. More on the legal bases can be found in the [Data Privacy Handbook](#).

For whom is this Self-Reflection Guide intended?

This guide is intended for researchers working in the Faculty of the Humanities, conducting human-related research using big data. This refers to research with large amounts of personal data collected without contact, without correspondence between researcher and data subjects.

Note that **extra care** is requested under the following circumstances:

1. You work with data on sensitive issues.
For example, polarisation, extremism, radicalisation, lgbtqia+.
2. You work with data on '[special categories of personal data](#)'.
Note that based on the GDPR you need to invoke 'an exception ground' to be allowed to work with these special categories of personal data. In research, this usually requires the explicit provision of consent, unless that is demonstrably too much trouble, or a data subject has already disclosed the data themselves.
3. You work with data from/about vulnerable groups.
For example, persons incapable of giving informed consent, minors, refugees, newcomers, prisoners, homeless people, people with dementia, or persons from socially disadvantaged groups (e.g. from ethnic, cultural or sexual minorities, see [Regulations FEtC-H](#), p.2).
4. You work with data from external parties.
For example, UMCU, publishers, social media platforms. Extra care is required here because there may be contractual (intellectual property) issues involved.

How does it work?

Walk through the different sections of this Self Reflection Guide and reflect on the conditions for responsibly carrying out data processing actions. Document (in writing) your considerations in the cells of the tables below, under 'My reflection.' Save the completed document under a unique filename, and store it as part of your research project. By documenting the pros and cons for each data act, you will have laid a solid foundation for conducting a responsible study, and after completion you may begin your research.

Are you unable to elaborate your reflection on one or more acts? Or you don't know how? Then ask for assistance from, for example, the FEtC-H, the privacy officer or the data manager, as indicated in the "assistance" column of the Self-reflection Guide.

Keep the completed Self-reflection Guide (the present document) for the duration of your research and update it when changes in the design of the research may pose risks to the data subjects (see also [Regulations FEtC-H](#), 'Acceptable burden and risks for the participants'). Upon completion of the study, the Self-Reflection guide is subject to a retention period of at least 10 years, for more information see the [Utrecht University policy framework for research data](#).

Self-reflection guide MBDO

If you have any questions about contents of this Self Reflection Guide, please do not hesitate to contact any of the undermentioned parties.

Contact details

Party	Website	Contact
Faculty Ethics Assessment Committee Humanities (FETC-H)	https://fetc-gw.wp.hum.uu.nl/en/	fetc-gw@uu.nl
Data manager	https://uu.nl/rdm https://cdh.uu.nl/research/data-management-privacy/	info.rdm@uu.nl datamanagement.gw@uu.nl
Privacy officer	https://intranet.uu.nl/en/knowledge-base/privacy-at-uu https://uu.nl/privacyhandbook	privacy.gw@uu.nl
Research engineer	https://cdh.uu.nl/ https://utrechtuniversity.github.io/research-engineering/	cdh@uu.nl research.engineering@uu.nl
Information security	https://intranet.uu.nl/en/information-security-at-the-uu	informatiebeveiliging@uu.nl
Research Coordination	https://intranet.uu.nl/en/knowledgebase/research-coordination	researchcoordinators.hum@uu.nl
Information and Technology Services (ITS)	https://intranet.uu.nl/en/knowledgebase/advice-and-it-solutions	https://uu.topdesk.net/
Advisor on knowledge security	https://intranet.uu.nl/en/knowledgebase/knowledge-security	kennisveiligheid@uu.nl

Responsible collection (pre-analysis phase)

Processing data

Act	Additional information	Assistance
Data processing		
I have (transparently) documented how the researcher and the data are protected against threats and dangers (offline and online), all during data processing acts such as storage, transportation, processing, sharing, publishing, etc.	<ul style="list-style-type: none"> ▪ Storing and preserving data ▪ Working safely with research data from home ▪ Information security at UU ▪ Encryption 	data manager, research engineer, knowledge security adviser, information security
My reflection:		
Data management		
<p>In my data management plan, I have discussed the points below:</p> <p>A. In light of the research question, an inventory was made beforehand listing the data to be collected and it is detailed what the dataset to be collected will look like (structure, format, etc.).</p> <p>B. Data used for publications may be processed and archived (considering terms of service, personal data protection, intellectual and legal property).</p> <p>C. The data, a processed version thereof, or the metadata of the dataset are published in a data archive, and are made available for reuse (FAIR) where possible.</p>	<ul style="list-style-type: none"> ▪ Data management planning ▪ DMPonline 	data manager, privacy officer (to discuss conditions for archiving)
My scholarly reflection:		
Provenance		
<p>I have reflected on the origins of the data ('data provenance'). Data provenance is not problematic, which means:</p> <p>A. use falls within the 'Terms of Service' (ToS), e.g.: users of a platform are made aware that the data they share with that platform may be used by third parties, AND</p> <p>B. data have been obtained lawfully, i.e. data are not obtained from hacking or theft, or obtained under duress or threat AND</p> <p>C. ownership of data is not disputed, AND</p> <p>D. use of data for scientific research is allowed.</p>	<ul style="list-style-type: none"> ▪ Data Privacy Handbook. 	FETC-GW, data manager, privacy officer (in case of questions about the Terms of Service)

My scholarly reflection:		
User- and processing agreements		
I have made all necessary agreements with any platform and/or database owners and processors of personal data.		data manager, privacy officer
My scholarly reflection:		
Storage		
A solution offered by the UU will be used to store the data during the research project OR I am not making a copy of the data, but it is clear how the data can be analysed/used for research at the original location.	<ul style="list-style-type: none"> ▪ Yoda - A research data management service ▪ Storage finder 	data manager
My scholarly reflection:		
Collaboration		
In the case I cooperate with other institutions: I have laid down in an agreement or arrangement, among other things, where the data are stored, how access to the data is regulated, who informs the data subjects and to whom the data subjects can go to exercise their rights.	<ul style="list-style-type: none"> ▪ Agreements ▪ Sharing personal data with collaborators 	RSO, privacy officer
My scholarly reflection:		
Data integrity		
I have reflected on the reliability of the platform and on how data quality and integrity can be ensured.		FETC-GW, data manager
My scholarly reflection:		
Additional finding		
'Additional findings' includes data that are not needed to answer the research question, but that were inevitably collected during the pre-analysis phase anyway. I have chosen a line of conduct from my field for dealing with additional findings, or formulated one myself.		FETC-GW, data manager, privacy officer

My scholarly reflection:

Working responsibly with persons and personal data (pre-analysis phase)

Act	Additional information	Assistance
Personal data		
<p>One of the following scenarios applies:</p> <ul style="list-style-type: none"> • <u>no</u> personal data is processed OR • I work with personal data in an archive under the control of a public authority and for which rules have been drawn up on the re-use of personal data OR • I work with data that, through anonymization techniques (describe the techniques you use), are no longer traceable to individuals, either directly or indirectly, and are therefore no longer considered personal data OR • I work with data that are extremely difficult to trace back to individuals through the use of pseudonymization techniques (describe the techniques you use), for example, because the data and key file are stored separately from each other OR • I use personal data and a protocol is in place for handling personal data, e.g. based on a Privacy Scan. 	<ul style="list-style-type: none"> ▪ What are personal data? ▪ Privacy scan ▪ Pseudonymisation & anonymisation 	privacy officer
My scholarly reflection:		
Special personal data		
<p>I use special personal data and I have contacted the privacy officer for advice on determining the need for the use of special personal data and on mitigating the risks by taking appropriate measures.</p>	<ul style="list-style-type: none"> ▪ Special categories of personal data 	privacy officer
My scholarly reflection:		
Provision of information to data subjects		
<p>I tried to inform data subjects about the processing of their personal data as part of the</p>	<ul style="list-style-type: none"> ▪ Data Privacy Handbook 	privacy officer

research (Art. 14 AVG), e.g. by publishing a privacy statement on a publicly accessible website and posting a message with a link to that statement on a relevant (social) media platform.		
My scholarly reflection:		
Dignity of the data subject		
I ensure the dignity of the data subject is respected: <ul style="list-style-type: none"> • Representativeness, or possible bias, is considered when including and selecting (for admission) the data subjects; AND • There is consideration of the cultural context, which includes power relations around the data subject and context-specificity of the data collected EN. • Reasonable expectations of data subjects regarding the public/private nature of information are respected, taking into account that the data subject is not explicitly asked to cooperate. 	For additional information for example read the book ' Data Feminism ', by Catherine D'Ignazio & Lauren Klein	privacy officer, FETC-GW
My scholarly reflection:		

Responsibly analysing/studying/profiling data (analysis phase)

Act	Additional information	Assistance
Technical solutions		
I will use UU's existing (technical) solutions for storing/analysing/processing the data. If not, I will contact the privacy officer and data manager.	<ul style="list-style-type: none"> ▪ https://tools.uu.nl/ ▪ Collecting data from the web ▪ High performance and cloud computing 	data manager, research engineer, privacy officer
My scholarly reflection:		
Coincidental findings		
I have adopted or formulated a course of action on how to deal with unforeseen coincidental findings or conclusions (such as: finding child	<ul style="list-style-type: none"> ▪ See FETC-GW regulations, 5.2.2. Acceptable burden and risks of participants ▪ DPIA/Privacy Scan 	FETC-GW, knowledge security adviser, privacy officer

<p>abuse, terrorist threat, crime, undiagnosed condition).</p> <p>If necessary, the privacy officer can supervise a Data Protection Impact Assessment (DPIA).</p>	<ul style="list-style-type: none"> ▪ Knowledge security UU ▪ Knowledge Safety Desk 	
<p>My scholarly reflection:</p>		

Responsible sharing/publishing (post-analysis phase)

Act	Additional information	Assistance
<p>Impact findings on data subject</p>		
<p>I have conducted a risk analysis of the potential impact of the outcomes of the research on data subjects or on third parties (e.g. discrimination or exclusion, financial hardship, risks of violation of other fundamental rights, etc.).</p> <p>The starting point in human-related research is the <i>do-no-harm</i>-principle. The risks to the data subjects must be acceptable and in reasonable proportion to the expected benefits of the research.</p>	<ul style="list-style-type: none"> ▪ Seek advice from the FETC-GW (see FETC-GW regulations, 5.2.2. Acceptable burden and risks of participants) ▪ Privacy risk assessment 	<p>FETC-GW, privacy officer</p>
<p>My scholarly reflection:</p>		
<p>Impact findings on researcher or university</p>		
<p>I have made a risk analysis of the possible consequences of the outcomes of the research in the context of knowledge security, in terms of the position of the researcher themselves or the university (e.g. damage to the university's reputation, inadvertent disclosure of contact or personal data of individual researchers or adverse consequences of a data breach).</p>	<ul style="list-style-type: none"> ▪ Knowledge Safety Desk ▪ Knowledge security UU ▪ Data classification 	<p>FETC-GW, Knowledge security adviser, information security, privacy officer, Director of Research</p>
<p>My scholarly reflection:</p>		

Making data available		
The data, a processed version thereof, or (in the case of sensitive data or subjects) at least the metadata of the dataset are published in a data archive after the study is completed, and where possible made available for reuse (FAIR).	<ul style="list-style-type: none">▪ How to make your data FAIR▪ Metadata and documentation	data manager
My scholarly reflection:		

References

D'Ignazio, C. and Klein, L. F. (2020). *Data feminism*. MIT press.

European Commission (2021) *Ethics in Social Science and Humanities*, Section 3.3 Internet research and social media data in research. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-in-social-science-and-humanities_he_en.pdf

European Commission (2021). *Ethics and Data Protection*. Section VII. Use of previously collected data ('secondary use'). Available at https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

FETC-GW. (October 10, 2022) *Regulations Faculty Ethics Review Committee Humanities*. Utrecht University.

Fleming, G., & Bruce, P. C. (2021). *Responsible Data Science*. John Wiley & Sons.

General Data Protection Regulation (AVG) (June 12, 2023) <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1528874672298&uri=CELEX:02016R0679-20160504>

Gerritsen, 2021. Tactful non-contact research - *Manual for AVG-compliant researchers. Assigned to the Data School*. Available at: <https://dspace.library.uu.nl/handle/1874/406645>

[KNAW \(2018\). Big data in wetenschappelijk onderzoek met gegevens over personen. Amsterdam: KNAW. Beschikbaar op: https://www.knaw.nl/nl/publicaties/big-data-wetenschappelijk-onderzoek-met-gegevens-over-personen](https://www.knaw.nl/nl/publicaties/big-data-wetenschappelijk-onderzoek-met-gegevens-over-personen)

LSE Research Ethics Committee (2022). *Using Data From The Internet and Social Media in Research: Ethics & Consent*. London School of Economics and Political Science. Beschikbaar op: <https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-Using-internet-and-Social-media-data-v8.pdf>

NESH (2019). *A Guide to Internet Research Ethics*. National Committee for Research Ethics in the Social Sciences and the Humanities. Beschikbaar op: <https://www.forskningsetikk.no/en/guidelines/social-sciences-and-humanities/a-guide-to-internet-research-ethics/>

Townsend, L. & Wallace, C. (2016). *Social media research: A guide to ethics*. University of Aberdeen. Beschikbaar op: https://www.gla.ac.uk/media/media_487729_en.pdf.

Utrecht University RDM (2023) *Data Privacy Handbook*. Available via: <https://uu.nl/privacyhandbook>.

Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media + Society*, 4(2). Beschikbaar op: <https://doi.org/10.1177/2056305118768300>