

Fair data and open data: differences and consequences

1. To share or not to share: what is fair?
Alex Burdorf, Erasmus MC Rotterdam
2. Data sharing: consequences for informed consent
Marie-José Bonthuis, UMC Groningen

WHO ARE WE ?



FEDERA

Federation of Dutch Medical Scientific Societies

COREON

Commission on regulatory aspects of medical research:

- Code of Conduct for use of data in health research
- Code of Conduct for responsible use of human tissue



Voor een gezond onderzoeksklimaat

Did you know?

ZonMW general terms and conditions governing grants:

20.2. The grant recipient must draw up a data management plan in which it indicates how data will be shared, when the data will be made available to third parties and the manner in which the data will be made accessible. That plan must be submitted to ZonMw for approval.

20.3. Databases, together with the related explanatory notes, will in any event be made available in the manner indicated by ZonMw for use for the benefit of further scientific and/or academic research. Therefore, the data that are collective within a project must be documented and saved in an accessible manner that is in accordance with the applicable standards.

US National Institutes of Health:

Timely release and sharing” is defined as no later than the acceptance for publication of the main findings from the final data set.

Did you know?

Lancet:

Authors may be required to provide the raw data for research papers when they are under review and up to 10 years after publication in The Lancet.

PLOS One:

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a Data Availability Statement describing compliance with PLOS's policy.

Data sharing developments

Guiding principles for scientific data management and stewardship

F = findability; unique and persistent identifier, description of metadata

A = accessibility; retrievable in open formats, license and access conditions

I = interoperability; ready to be combined with other datasets by computer systems

R = Reusability; documentation needed to understand the data and analysis

Open source data repository

* Dataverse: data repository software, formal data citation

* DANS Dataverse: <https://dataverse.nl/dvn> (Dutch universities)

Open Data/ Open Access

- Open Data
- Open Access
 - results (publications)
 - data (full, restricted)
- ‘Open if possible’, restricted if necessary’
 - dataprotection
- *Exception: presence of personal data*



Data sharing in Dutch research

5 research organisations:

Interviews with junior researchers, senior researchers

Interviews with data managers

Evaluation ZonMW checklist data management:

Usability?

Support of open access?

VERMENIGVULDIGEN DOOR DELEN

*Data delen in gezondheidsonderzoek,
van ambitie naar praktijk*

Een rapport van COREON/Federa, in opdracht van ZonMw

Auteurs: Danielle van den Brink
Alex Burdorf
Ernst de Graag
René Melis

Datum: 23 december 2015



Data sharing in Dutch research

Views of researchers:

- * structured data management plan provides guidance, creates awareness, increases quality
- * time needed for appropriate data stewardship
- * lots of work...for whom ? Will data ever be used ?

Important issues in data stewardship:

- * open versus restricted access
- * privacy and informed consent, storage of identifiers for future linkage
- * timing of open access, fixed datasets versus ongoing data collection
- * linked, enriched data, ownership
- * sustainable storage
- * scientific recognition, data citation

Data sharing in Dutch research

Facilities and support in organisations:

- * provision of data protocols and practical support for data stewardship
- * agreement on informed consent
- * retention period of data (15 years?)
- * provision of infrastructure for data sharing (technical, legal)
- * parameters of acknowledgements for data collection (credits)
- * evaluation of scientific and societal benefits of data sharing (data graveyards)

Data sharing recommendations

Funding agencies:

1. Differentiate in requirements for open data access
2. Facilitate open access to research data (infrastructure, protocols, instruments)
3. Facilitate research with reuse of data (grants)
4. Lobby for data protection regulations that support appropriate research

Researchers and their organizations:

1. Provide training and support for researchers and departments
2. Provide adequate infrastructure (IT, software, legal)
3. Design and implement local policies



HANDS

Handbook for Adequate Natural Data Stewardship

[What is data stewardship?](#)

[You are a data steward](#)

[FAIR principle](#)

[Accountability and responsibility](#)

[Preparing your study](#)

[Creating and processing your data](#)

[Analysing your data](#)

[Preserving your data \(archiving\)](#)

[Giving access to your data](#)

[About this Handbook](#)

[Glossary](#)

[Appendices](#)

Handbook for Adequate Natural Data Stewardship



The Federation of Dutch UMCs (NFU) has developed this handbook to promote good data stewardship at the eight Dutch university medical centres (UMCs). HANDS was written by experts with knowledge of the current regulations and best practices for data stewardship. The handbook should help you become a good data steward. Please consult experts within your UMC or within your research field if you have questions about your own study.

Fair data and open data: differences and consequences

1. To share or not to share: what is fair?
Alex Burdorf, Erasmus MC Rotterdam
2. Data sharing: consequences for informed consent
Marie-José Bonthuis, UMC Groningen

Personal data

- (In-)direct identifiable versus anonymous
 - dynamic
 - based on context
- Re-use= real risk?
Focus on collecting or audit trail?



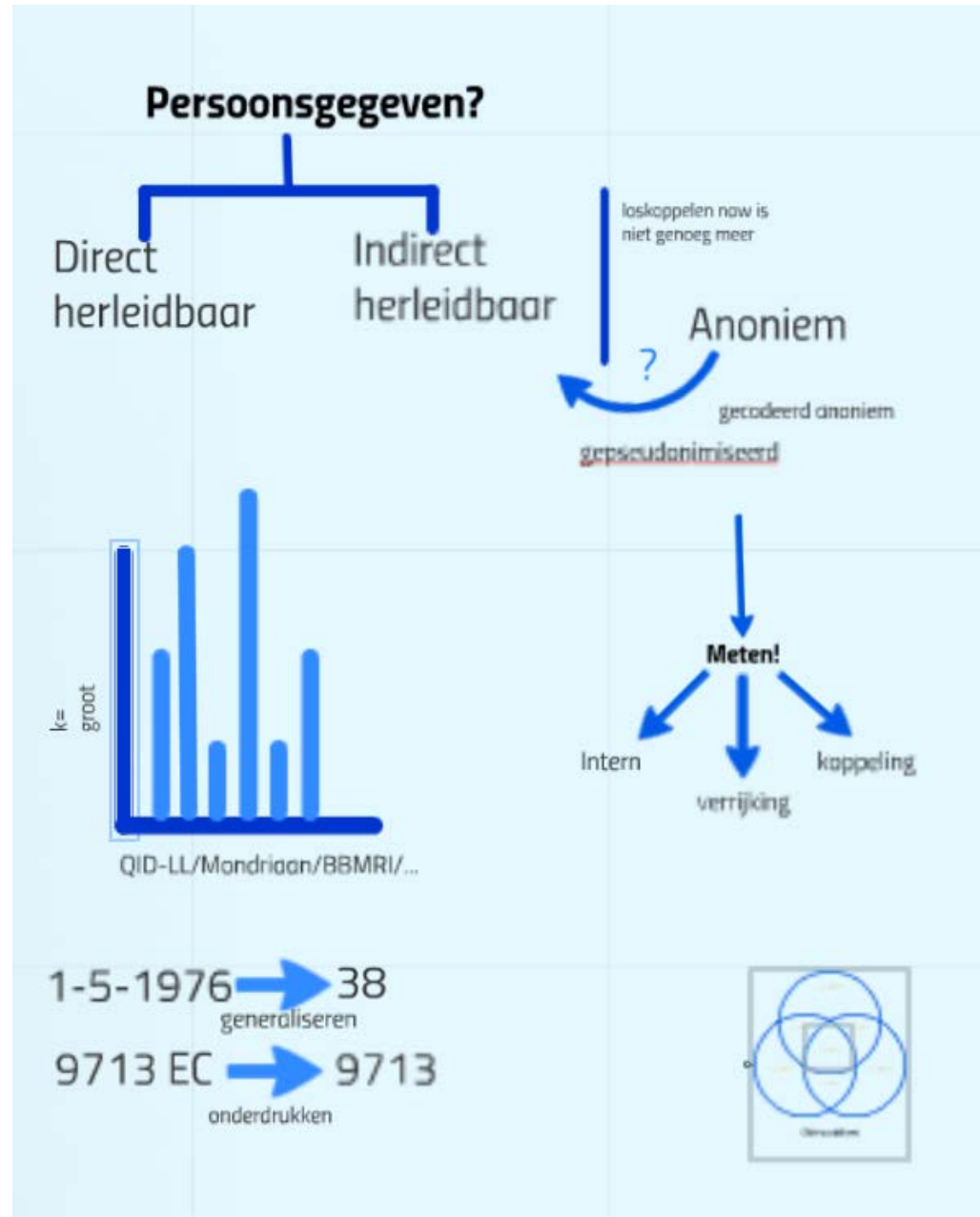
Working Party 29's opinion on anonymization

- Singling out (k-anonymity)
- Linkability
- Re-identification

->*Dynamic approach*



K-anonymity



Measuring and predicting ‘re-identification’

Table 3.2: Anonymity set size k for various (potential) quasi-identifiers

Quasi-identifier:	# of sets	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
PC4	388	2	3,278	7,090	7,188	10,300	22,330
PC6	66,883	1	24	35	41	50	1,322
PC4+DoB	2,267,700	1	1	1	1	1	42
PC6+DoB	2,759,422	1	1	1	1	1	5
PC4+gender	776	1	1,652	3,536	3,594	5,151	11,730
PC6+gender	133,012	1	11	18	21	25	954
gender+YoB	221	1	5,219	14,570	12,550	19,740	25,580
gender+YoB+MoB	2,699	1	397	1,177	1,028	1,594	2,326
gender+YoB+MoB+PC4 ^a	635,679	1	2	3	4	6	40
gender+YoB+MoB+municip. ^b	34,790	1	6	18	80	96	733
gender+DoB	71,318	1	21	40	39	54	571
gender+DoB+PC4	2,488,828	1	1	1	1	1	22
gender+DoB+PC6	2,766,475	1	1	1	1	1	4
town+gender	134	1	222	1116	20,700	3259	347,100
town+YoB	5,642	1	6	29	492	101	14,270
town+YoB+MoB	49,207	1	2	5	56	20	1,262
town+DoB	463,134	1	1	2	6	7	419
town+YoB+gender	10,492	1	4	17	264	60	7,515
town+YoB+MoB+gender	83,172	1	1	3	33	14	695
town+DoB+gender	697,875	1	1	2	4	5	226

^a QID_A , see Section 3.3.2.

^b QID_B , see Section 3.3.3.

Wrap up

- Dynamic approach personal data (k-anonymity, WP29)
- Transparency; audit trail
- Consequences for consent



Agenda setting for COREON activities ?

- to be debated -

1. Procedures for informed consent
2. Facilities and (financial) support for open data
3. Facilities for linkage of datasets (trusted third party for anonymisation procedures)
4. Data repositories & big data research
5. VVE: Responsible epidemiologic research practice
reproducibility, research waste