Experiences with open and reproducible epidemiology using sensitive data

Kaitlyn Tsuruda
Department of Research, Cancer Registry of Norway
Norwegian Institute of Public Health

September 26, 2026

Cancer Registry of Norway

NIPH

Overview

- Cancer Registry of Norway
- Open and reproducible science
- In practice
  - Study registration
  - Open code
  - Data sharing
- Wrap-up

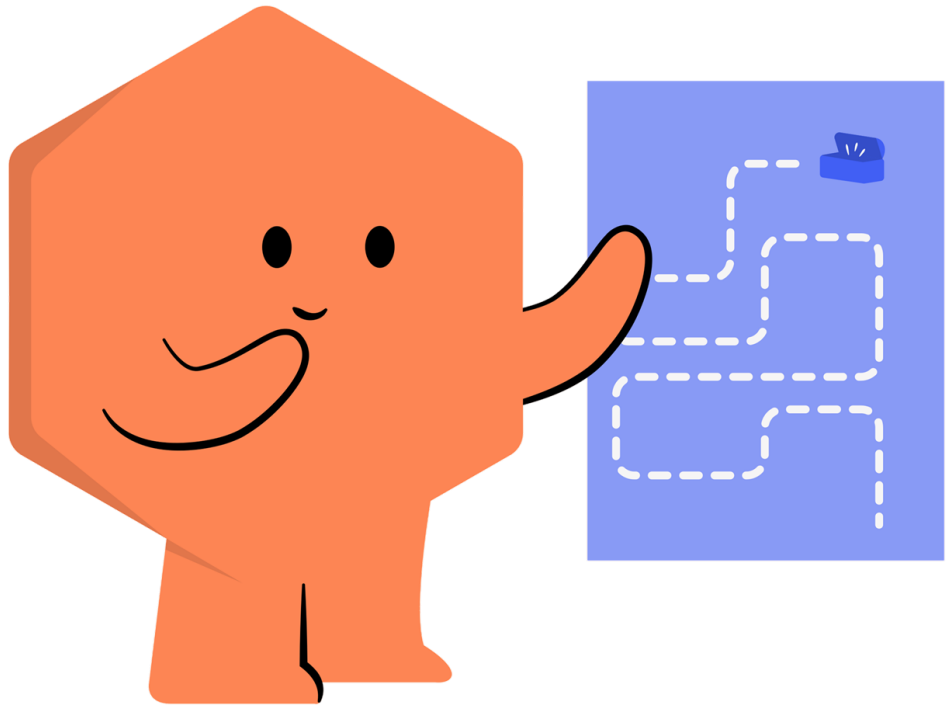


Illustration by Aditya Sahu on Unsplash

Department of Research, Cancer Registry of Norway

- Roughly 35 employees
  - Researchers
  - PhD students and post docs
  - Support staff: administrative + data management/analysis
- Classical, occupational, environmental, molecular, and pharmaco-epidemiology

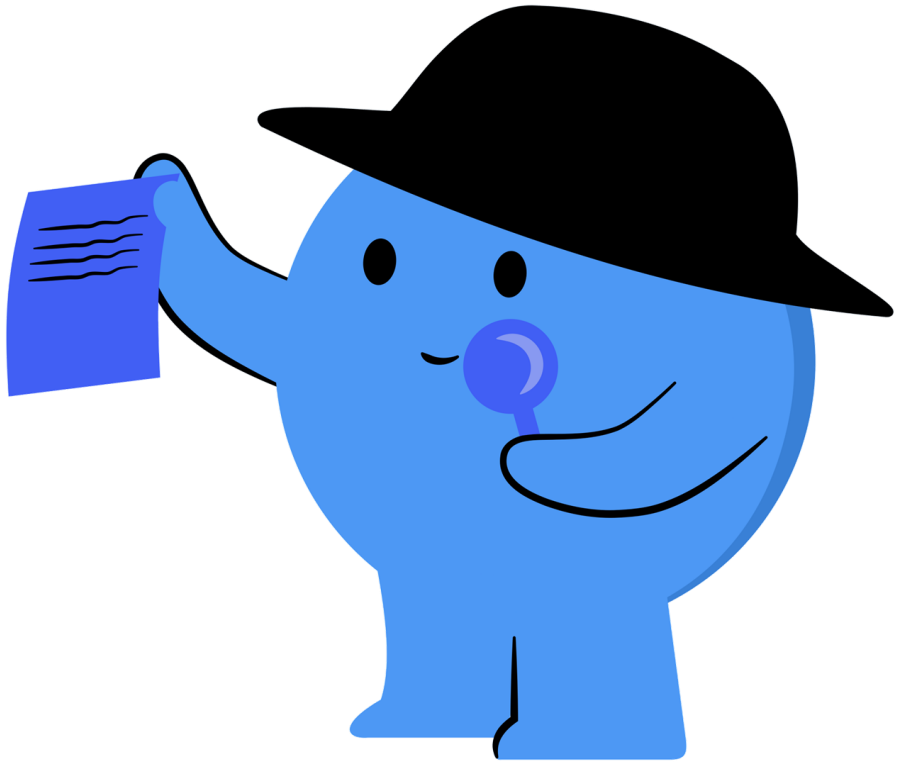


Illustration by Aditya Sahu on Unsplash

Cancer Registry of Norway

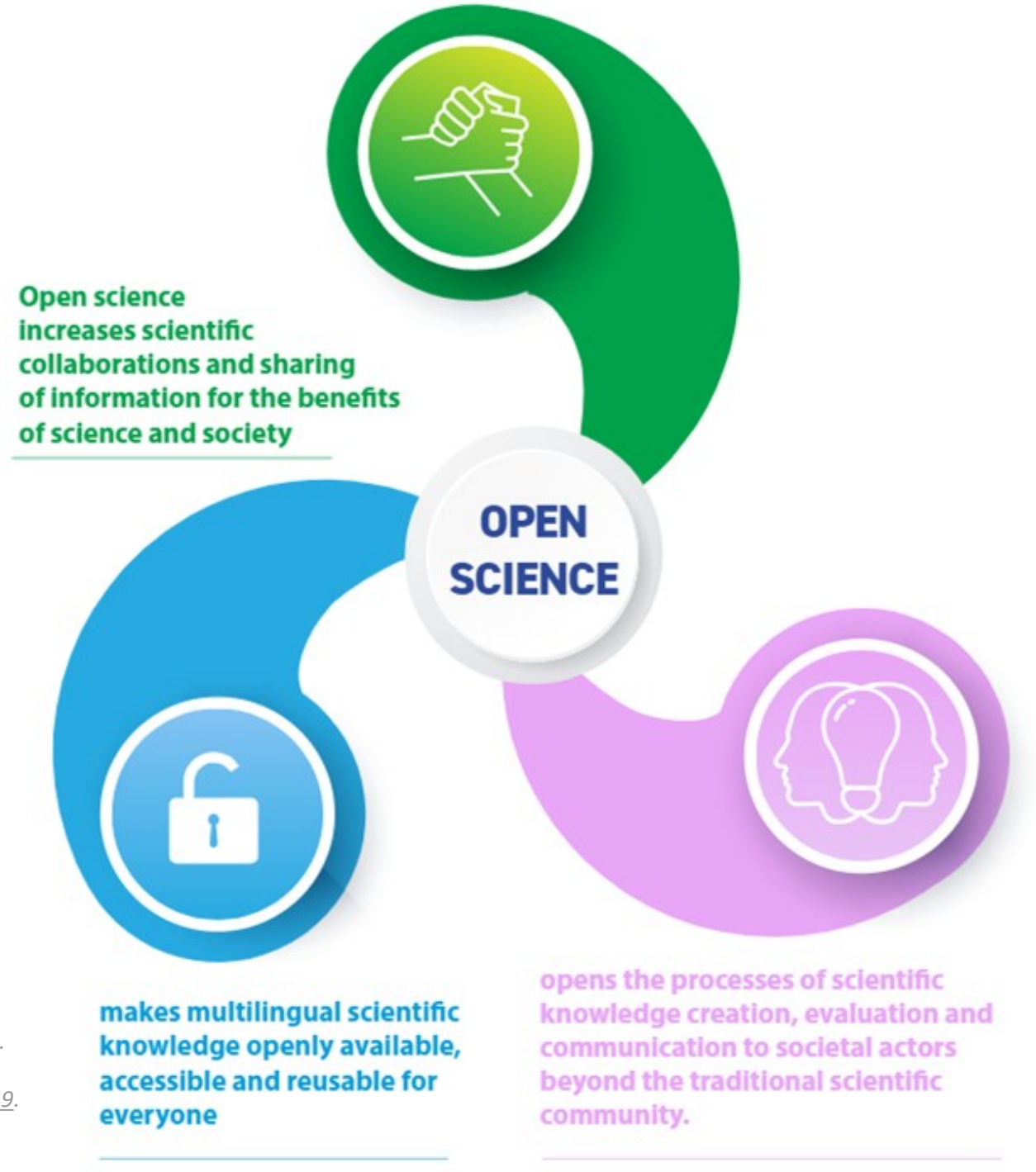
NIPH

Open and reproducible science

Cancer Registry of Norway

NIPH

What is Open science?

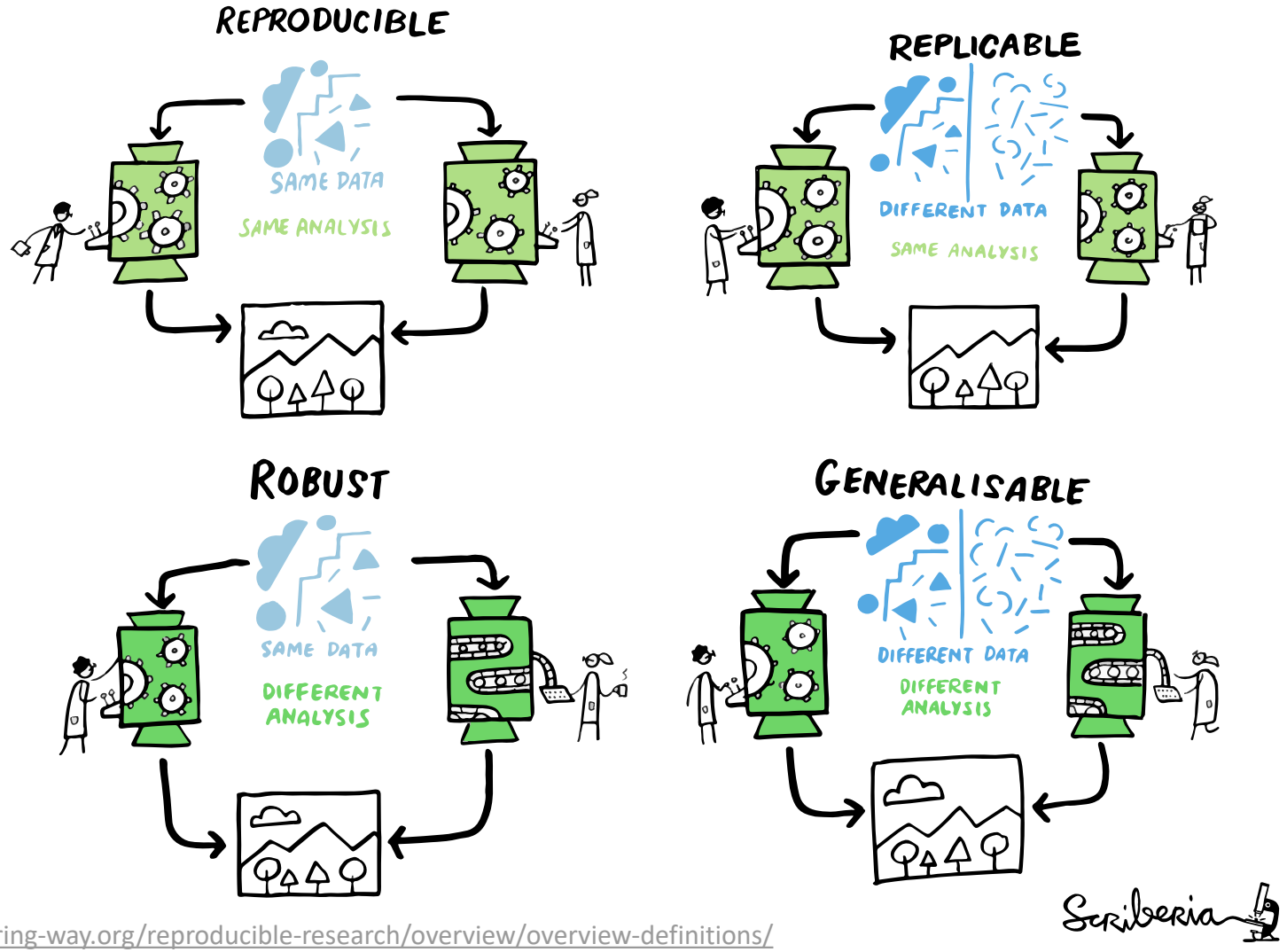


UNESCO. UNESCO Recommendation on Open Science. *2021; https://unesdoc.unesco.org/ark:/48223/pf0000379949. (Licence type* CC BY-SA 3.0 IGO)

Cancer Registry of Norway

NIPH

Research is reproducible if the same analysis on the same data can repeatedly produce the same results



The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: The Turing Way Community & Scriberia (2024).

https://book.the-turing-way.org/reproducible-research/overview/overview-definitions/

Study registration

Cancer Registry of Norway

NIPH

Study registration for RCTs is well established

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

Enter search terms SEARCH

Recommendations | Disclosure of Interest | About ICMJE | News & Editorials

Recommendations

Browse

- I. About the Recommendations
- II. Roles & Responsibilities
- III. Publishing & Editorial Issues
  - A. Corrections and Version Control
  - B. Scientific Misconduct
  - C. Copyright
  - D. Overlapping Publications
  - E. Correspondence
  - F. Fees
  - G. Supplements, Theme Issues, and

Home > Recommendations > Browse > Publishing & Editorial Issues > Clinical Trials

L. Clinical Trials

PAGE CONTENTS

1. Registration
2. Authors' Access to Data
3. Data Sharing

1. Registration

The ICMJE's clinical trial registration policy is detailed in a series of editorials (see News and Editorials and FAQs).

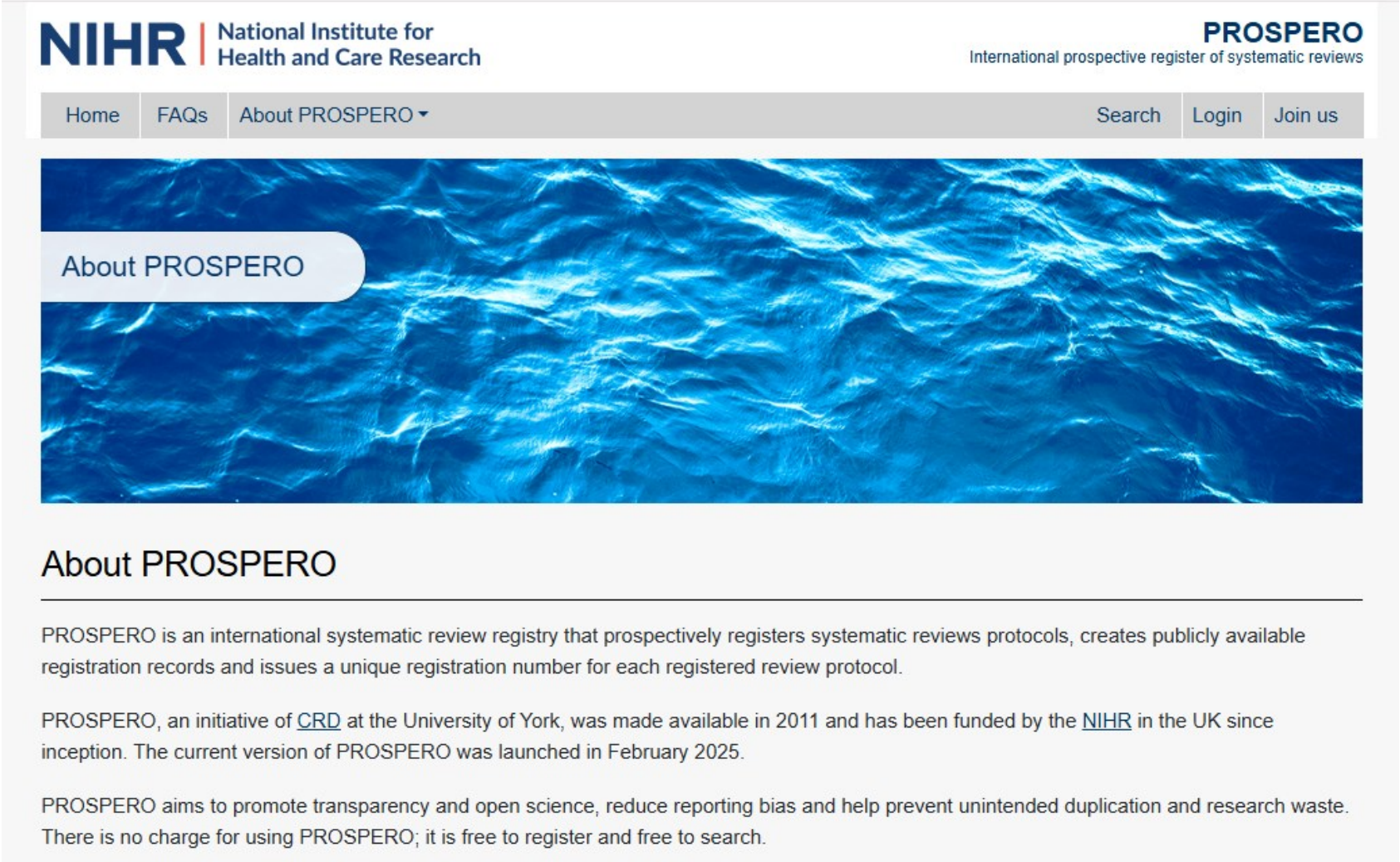
Briefly, the ICMJE recommends that all medical journal editors require registration of clinical trials in a public trials registry at or before the time of first patient consent for enrollment as a condition of consideration for publication.

https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

Cancer Registry of Norway

NIPH

Study registration for systemic reviews is also common



https://www.crd.york.ac.uk/PROSPERO/help/aboutprospero

Registering observational studies is not as common, but also has benefits!

- Improve discoverability
- Define scientific questions before data are accessed/analyzed
- Avoid publication bias
- Improve reproducibility and replicability
  - A “regular” methods section is too short to describe a full statistical analysis plan

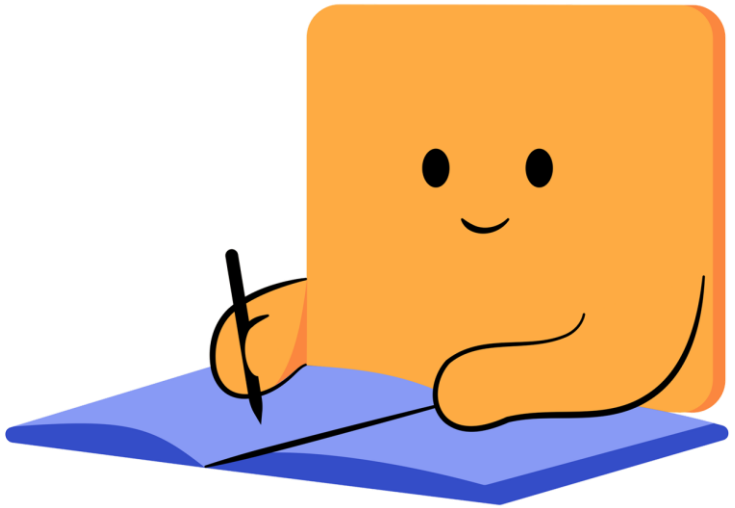
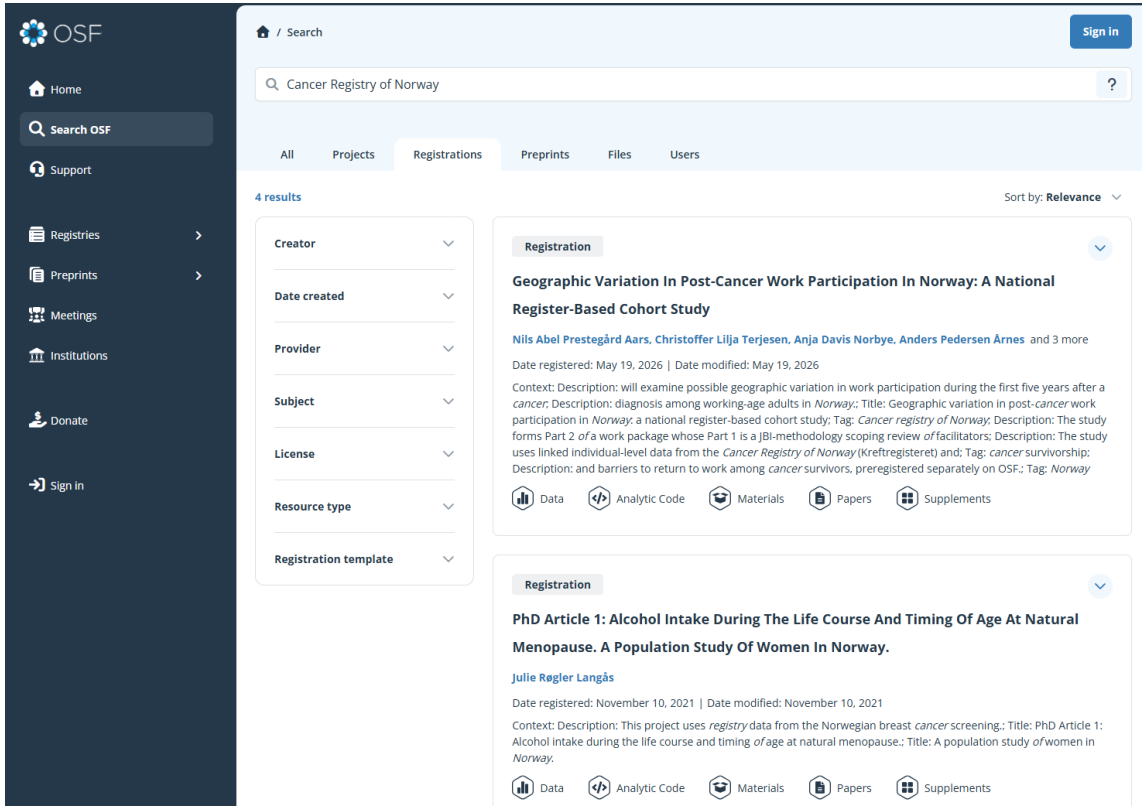
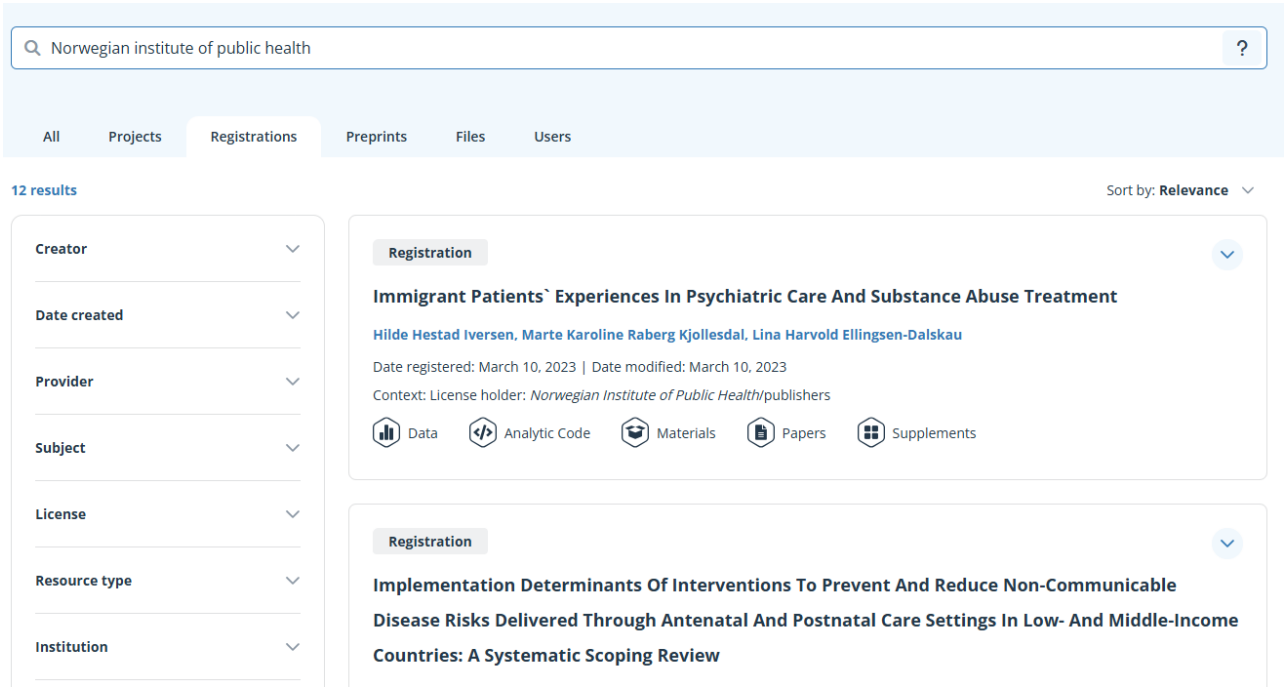


Illustration by Aditya Sahu on Unsplash

Preregistration can be straightforward if you already have a protocol

Fill out a template to upload your protocol to the OSF platform





https://osf.io/search?search=Cancer%20Registry%20of%20Norway&tab=3

https://osf.io/search?tab=3&search=Norwegian%20institute%20of%20public%20health

Cancer Registry of Norway

NIPH

Some ways to register an observational study

	OSF platform	Protocol paper
Version control	Yes	No/very limited
Availability	Yes (online)	Yes (OA), or just to journal subscribers
Cost	Free	Cost of publication
Peer review	No	Yes
Time from submission to publication	Short	Long

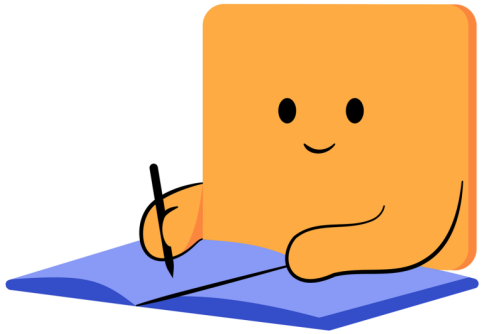


Illustration by Aditya Sahu on Unsplash

Registered reports



ELSEVIER

Journal of Clinical Epidemiology 185 (2025) 111894

Journal of Clinical Epidemiology

REGISTERED REPORT STAGE I

Comparison of AI-assisted and human-generated plain language summaries for Cochrane reviews: protocol for a randomised trial (HIET-1) [Registered Report - stage I]

Declan Devane[a,b,c,*], Johanna Pope[a,b,c], Paula Byrne[a], Evan Forde[d], Steven Woloshin[e,f], Eileen Culloty[g], Darren Dahly[h,i], Ingeborg Hess Elgersma[j], Heather Munthe-Kaas[j], Conor Judge[d], Martin O'Donnell[d], Finn Krewer[d], Sandra Galvin[a,c], Nikita Burke[b,c], Theresa Tierney[k], KM Saif-Ur-Rahman[b,l], Tom Conway[a,b,c], James Thomas[m]

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 191 (2026) 112102

REGISTERED REPORT STAGE II

Comparison of AI-assisted and human-generated plain language summaries for Cochrane reviews: a randomised non-inferiority trial (HIET-1) [Registered Report - stage II]

Declan Devane[a,b,c,*], Johanna Pope[a,b,c], Paula Byrne[a,b,c], Evan Forde[d], Isabel O'Byrne[d], Steven Woloshin[e,f], Eileen Culloty[g], Darren Dahly[h,i], Ingeborg Hess Elgersma[j], Heather Munthe-Kaas[j], Conor Judge[d], Martin O'Donnell[d], Finn Krewer[d], Sandra Galvin[a], Nikita N. Burke[b,c], Theresa Tierney[k], KM Saif-Ur-Rahman[b,l], Tom Conway[a,b,c], James Thomas[m]

Cancer Registry of Norway

https://www.cos.io/initiatives/registered-reports

NIPH

Some ways to register an observational study

	OSF platform	Protocol paper	Registered report
Version control	Yes	No/very limited	Yes
Who has access?	“Everyone” (online)	“Everyone” (OA), or just to journal subscribers	“Everyone” (OA), or just to journal subscribers
Cost	Free	Cost of publication	Cost of 2 publications
Peer review	No	Yes	Yes
Time from submission to publication	Short	Long	Longest

Open code

Sharing code is good practice!

- Gain insight into execution of statistical analysis plan
- Facilitate adoption of new/complex methodologies
- Replicate studies or find errors
- Gain trust from the public and other researchers

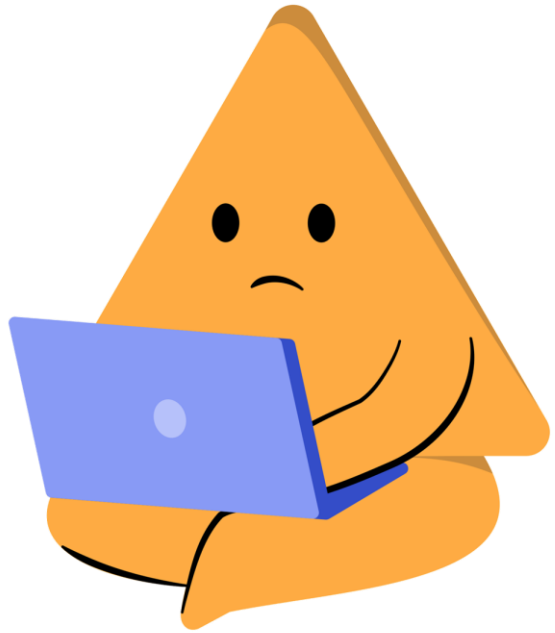


Illustration by Aditya Sahu on Unsplash

Tazare J, et al. Sharing Is Caring? International Society for Pharmacoepidemiology Review and Recommendations for Sharing Programming Code. Pharmacoepidemiol Drug Saf. 2024. https://doi.org/10.1002/pds.5856

But are we motivated to share?

- Can it really be reused?
- Making publication-ready code takes time
- Do we have the skills to share our code?



Modified illustration by Aditya Sahu on Unsplash

Tazare J, et al. Sharing Is Caring? International Society for Pharmacoepidemiology Review and Recommendations for Sharing Programming Code. Pharmacoepidemiol Drug Saf. 2024. https://doi.org/10.1002/pds.5856

The immortal life of poorly managed data: technical debt in registry-based research

Section for Administration and Research Support, Cancer Registry of Norway at NIPH
Kaitlyn M Tsuruda and Ronnie Babigumira

WHAT IS TECHNICAL DEBT?

INEFFICIENCIES CAUSED BY PRIORITIZING QUICK FIXES

Causes

- Organizational: rushed decisions, poor planning
- Technical: messy code

Effects

- Disorganized files
- Buggy code
- Lower reproducibility
- Decreased credibility

A well-managed data set is a technical asset



https://vincentdnl.com/en/drawings/technical-debt

CREATE TECHINCAL ASSETS: PLANNING

BEFORE YOUR DATA ARRIVE

Set roles: Establish team roles and expectations

Get organized: Adopt a file organization system

Make a plan: Write a data management plan and put it to work

AFTER YOUR DATA ARRIVE

Document: Write down key information and decisions

Capture the essentials: Excessive documentation steals time from other tasks, find a balanced based on the complexity of your project

CREATE TECHINCAL ASSETS: CODING PRACTICES

TREAT RAW DATA LIKE GOLD

Never edit your source data

CULTIVATE YOUR DATA

Explore first: Get to know your data with exploratory data analyses

Purposeful formatting: Let your study question dictate if you need long or wide data

CODE RIGHT, FUTURE'S BRIGHT

Be consistent: Name files, scripts, variables, and outputs clearly

Use best practices: Follow established guidelines for writing good code

Stay organized: Use your file organization system to house your data and code

https://zenodo.org/records/15040399

Cancer Registry of Norway

NIPH

README | Code of conduct | Contributing | CC-BY-4.0 license

Resource list for "better" and best-practices for data management in registry-based research

PRs welcome | last commit over 1 year ago

Table of Contents

- Commentary on avoiding technical debt
- Creating a data management plan
- Writing and documenting code
- Tips for data visualization
- Resources for Stata users

https://github.com/tsurudak/technical-debt

Git is a great tool for version control

git keeps detailed track of changes in your code and go back to old versions

GitHub and GitLab help you collaborate and share code

These are very powerful tools.
They can be intimidating to learn, but it’s worth taking the time!



Cancer Registry of Norway

NIPH

Data sharing

Cancer Registry of Norway

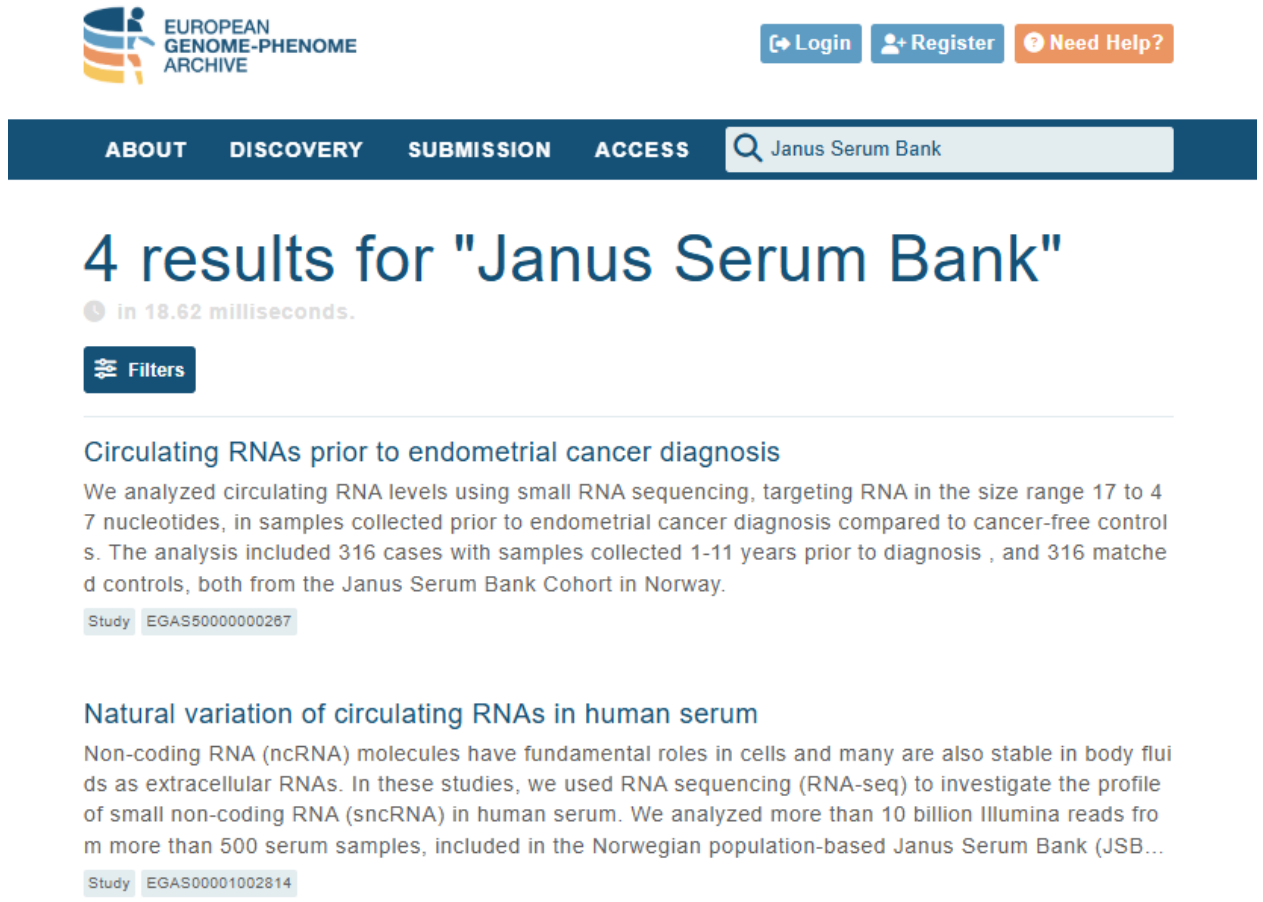
NIPH

Sharing sensitive research data

Federated European Genome-phenome Archive (FEGA) is a network for permanently archiving and sharing personally identifiable data

Strict access control procedures

- data access agreements with terms and conditions for access
- data access committees grant access to data



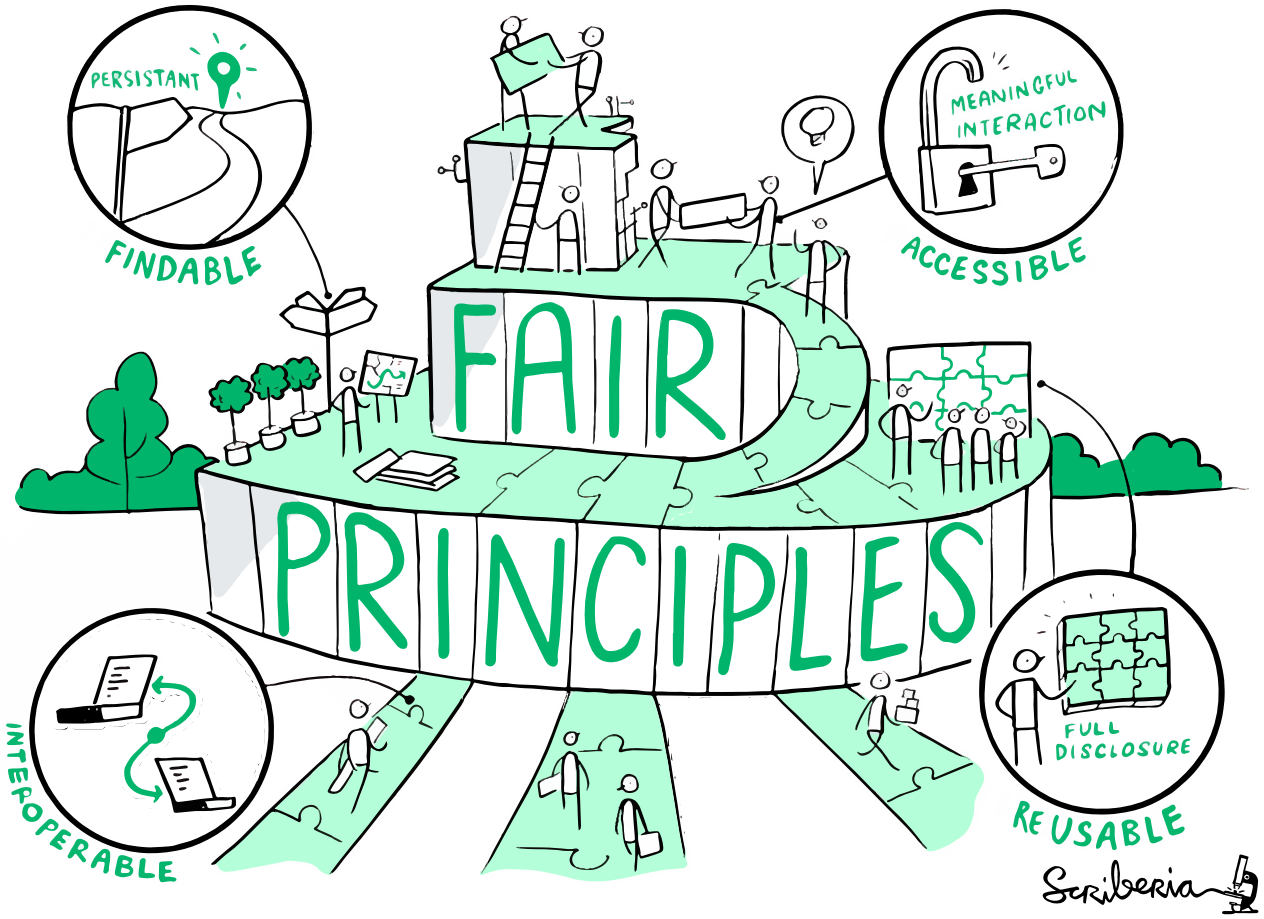
https://ega-archive.org/search/Janus%20Serum%20Bank

Cancer Registry of Norway

NIPH

Open, but what about FAIRness?

- Are data accessible?
- Reusable?
- Study participants can revoke consent at any time and be removed from any subsequent research analyses



The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: The Turing Way Community & Scriberia (2024).

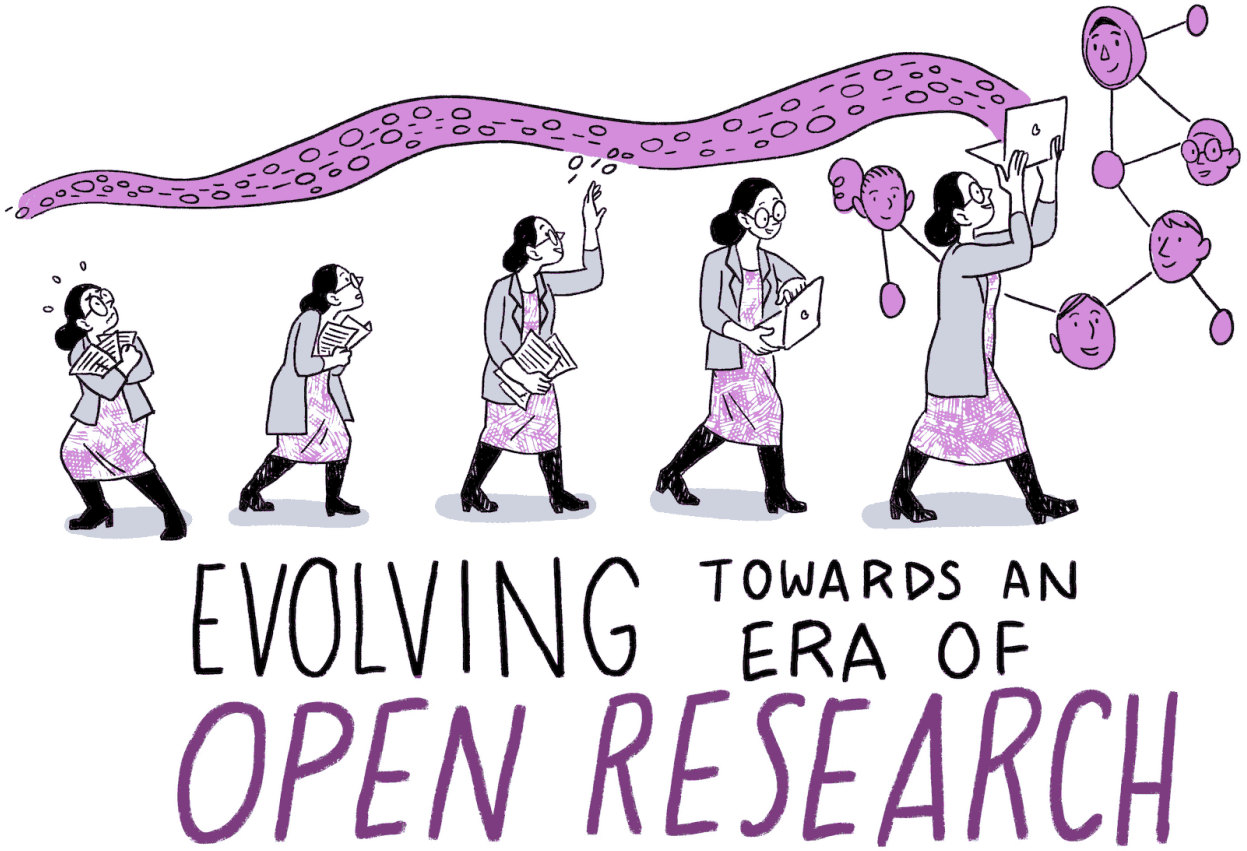
Cancer Registry of Norway

NIPH

Summary

Open and reproducible epidemiology

- Study protocol registration
  - Pre-registration on OSF
  - Protocol papers
  - Registered reports
- Open code
  - Git and GitHub/GitLab are powerful tools for code development, collaboration, and sharing
- Open (sensitive) data
  - Challenging, but not impossible



Cancer Registry of Norway

The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: The Turing Way Community & Scriberia (2024).

NIPH

Thank you

Ronnie Babigumira

Sarah Hjorth

Hilde Langseth

Marie Udnesseter Lie

Jo Steinson Stenehjem

Cancer Registry of Norway

NIPH

And thank *you*!

Questions?

Kaitlyn Tsuruda
kats@fhi.no

Cancer Registry of Norway

NIPH