Advanced Methodology
and Data Science

Good Data Science Practice:
Towards a code of practice for drug development

Mark Baillie
September 21st, 2021
Data Ethics Session

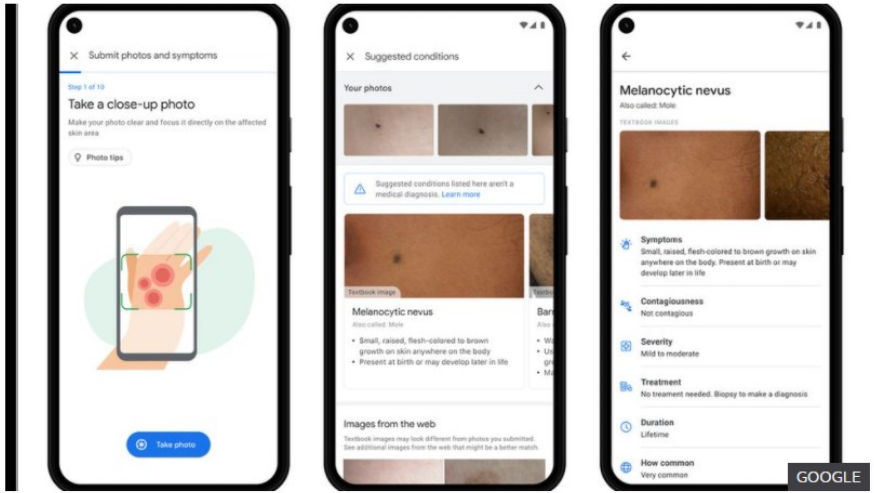
NOVARTIS | Reimagining Medicine

Advances in data science

Google AI tool can help patients identify skin conditions

By Zoe Kleinman
Technology reporter

20 hours ago



Google has unveiled a tool that uses artificial intelligence to help spot skin, hair and nail conditions, based on images uploaded by patients.

Source: https://www.bbc.com/news/technology-57157566

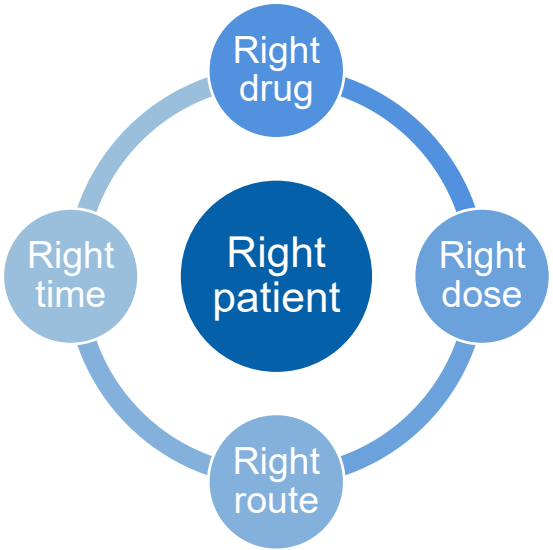
NLTK
Spark
theano
Numpy
Keras
Pytorch
TensorFlow
mxnet
Scikit-learn
Pandas
git
Bitbucket
GitLab
ggplot2
forcats
readr
stringr
purrr

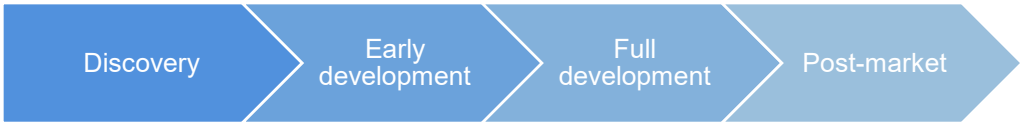
NOVARTIS | Reimagining Medicine

Can “data and digital” accelerate drug development?

By **learning from existing and future data** using advances in science, statistics, machine learning, computation, AI, etc. to:

- increase our understanding of drug, disease and patients,
- accelerate and improve our development projects, and
- inform our decision making.





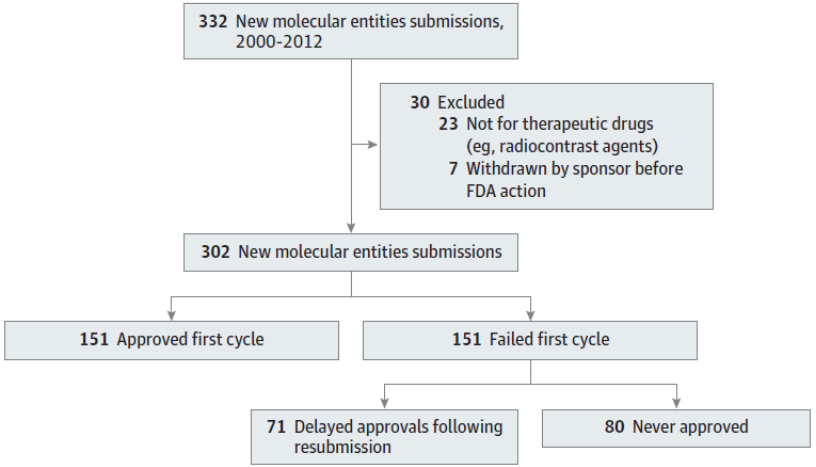
Destination without a map?
We don't always get the five rights right

Original Investigation

Scientific and Regulatory Reasons for Delay and Denial of FDA Approval of Initial Applications for New Drugs, 2000-2012

Leonard V. Sacks, MBBCh; Hala H. Shamsuddin, MD; Yuliya I. Yasinskaya, MD; Khaled Bouri, PhD, MPH; Michael L. Lanthier, BA; Rachel E. Sherman, MD, MPH

Figure. Flow Diagram of Outcomes for New Molecular Entities Submissions to the Center for Drug Evaluation and Research of the US Food and Drug Administration Between 2000 and 2012

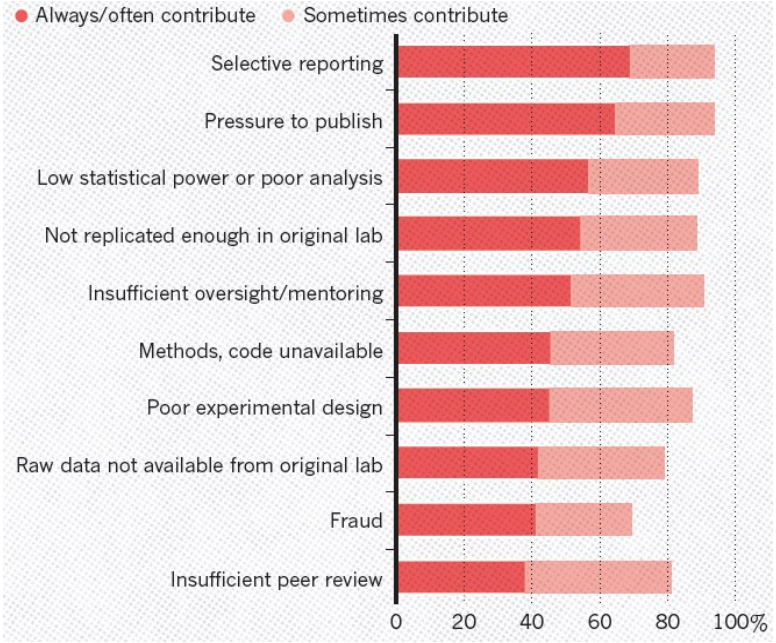


A crisis facing science?

1,500 scientists lift the lid on reproducibility

Survey sheds light on the ‘crisis’ rocking research.

What factors contribute to irreproducible research?
(Nature’s survey of 1,576 researchers)

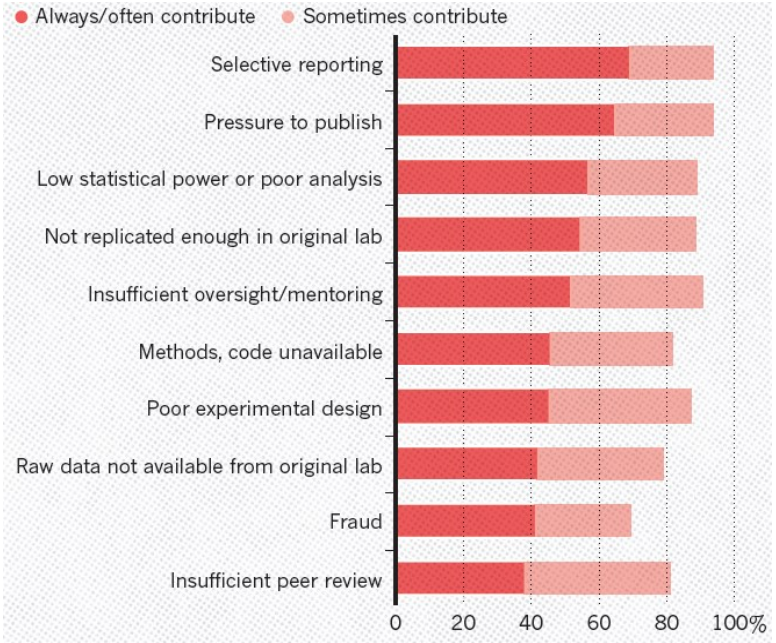


Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533,** 452–454 (2016). https://doi.org/10.1038/533452a

Different perspectives on the challenges facing science

- Computational
- Statistical
- Scientific
- Ethical and legal
- Human

What factors contribute to irreproducible research?
(Nature's survey of 1,576 researchers)



Good data science practice in drug development?

By learning from existing and future data using advances in science, statistics, machine learning, computation, AI, etc. to:

- increase our understanding of drug, disease and patients,
- accelerate and improve our development projects, and
- inform our decision making.

What do we mean by data science?

What are the practices of data science?

What do we mean by “good” practice in the context of drug development?

13 Core principles of ICH-GCP

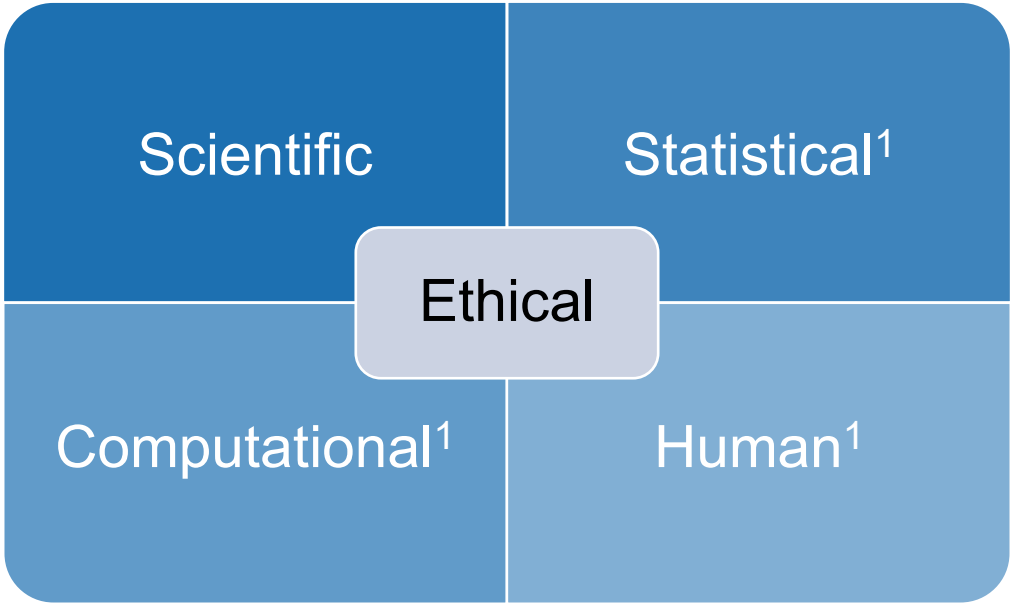
1. Ethical Principles of Declaration of Helsinki	2. Benefit justifies risk	3. Rights, safety, wellbeing	4. Adequate information to support trial	5. Clear, scientifically sound protocol
6. IRB/EC approval prior to initiation	7. Medical care / decisions by qualified physician	8. Researcher training, education and experience	9. Freely given informed consent	10. Accurate data handling and storage
	11. Data Protection and confidentiality	12. Good Manufacturing Practice	13. Quality assurance systems	

Principles of Good Clinical Practice | SpringerLink

A view of data science for drug development

A set of integrated thinking skills and practices refocused for answering questions with data

Greater statistics tend to be inclusive, eclectic with respect to methodology, closely associated with other disciplines, and practiced by many outside of academia and often outside professional statistics.
(John Chambers, 1993)

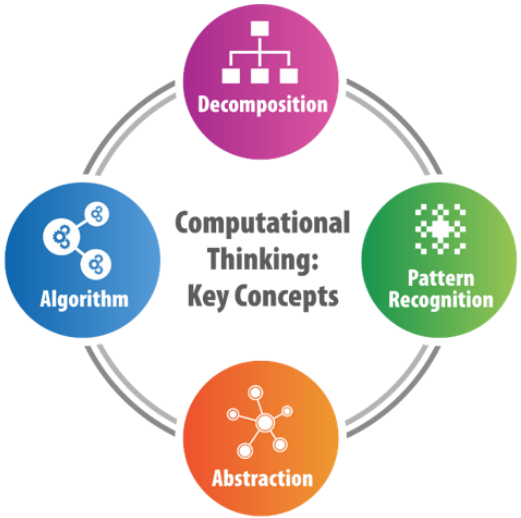


[1] *Blei & Smyth (2017) Science and data science. PNAS 114 (33):8689-8692*

Computational perspective

A big computer, a complex algorithm and a long time does not equal science.
(Robert Gentleman)

- Computational thinking is the design, implementation and execution under constraints.
- Focuses on the algorithmic implementation of methods, and a way to understand and compare their computational footprints.
- Design and application of computation software, packages, libraries, languages to solve a specific problem
- Practices for technical reproducibility and accuracy
- Computational thinking interfaces with scientific and statistical thinking
  - Advances in Bayesian inference and deep learning driven by improvements in numerical integration (automatic differentiation)

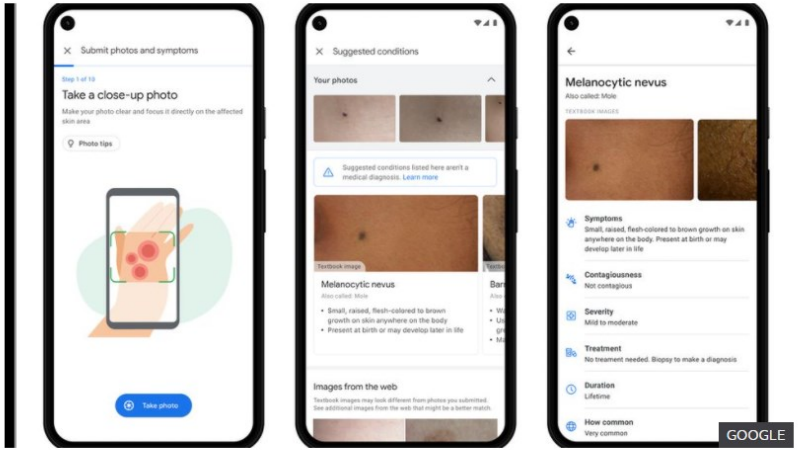


Advances in data and digital

Google AI tool can help patients identify skin conditions

By Zoe Kleinman
Technology reporter

20 hours ago



Google has unveiled a tool that uses artificial intelligence to help spot skin, hair and nail conditions, based on images uploaded by patients.

https://www.bbc.com/news/technology-57157566

NOVARTIS | Reimagining Medicine

JAMA Dermatology | Original Investigation

Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition

Julia K. Winkler, MD; Christine Fink, MD; Ferdinand Toberer, MD; Alexander Enk, MD; Teresa Deinlein, MD; Rainer Hofmann-Wellenhof, MD; Luc Thomas, MD; Aimilios Lallas, MD; Andreas Blum, MD; Wilhelm Stolz, MD; Holger A. Haenssle, MD

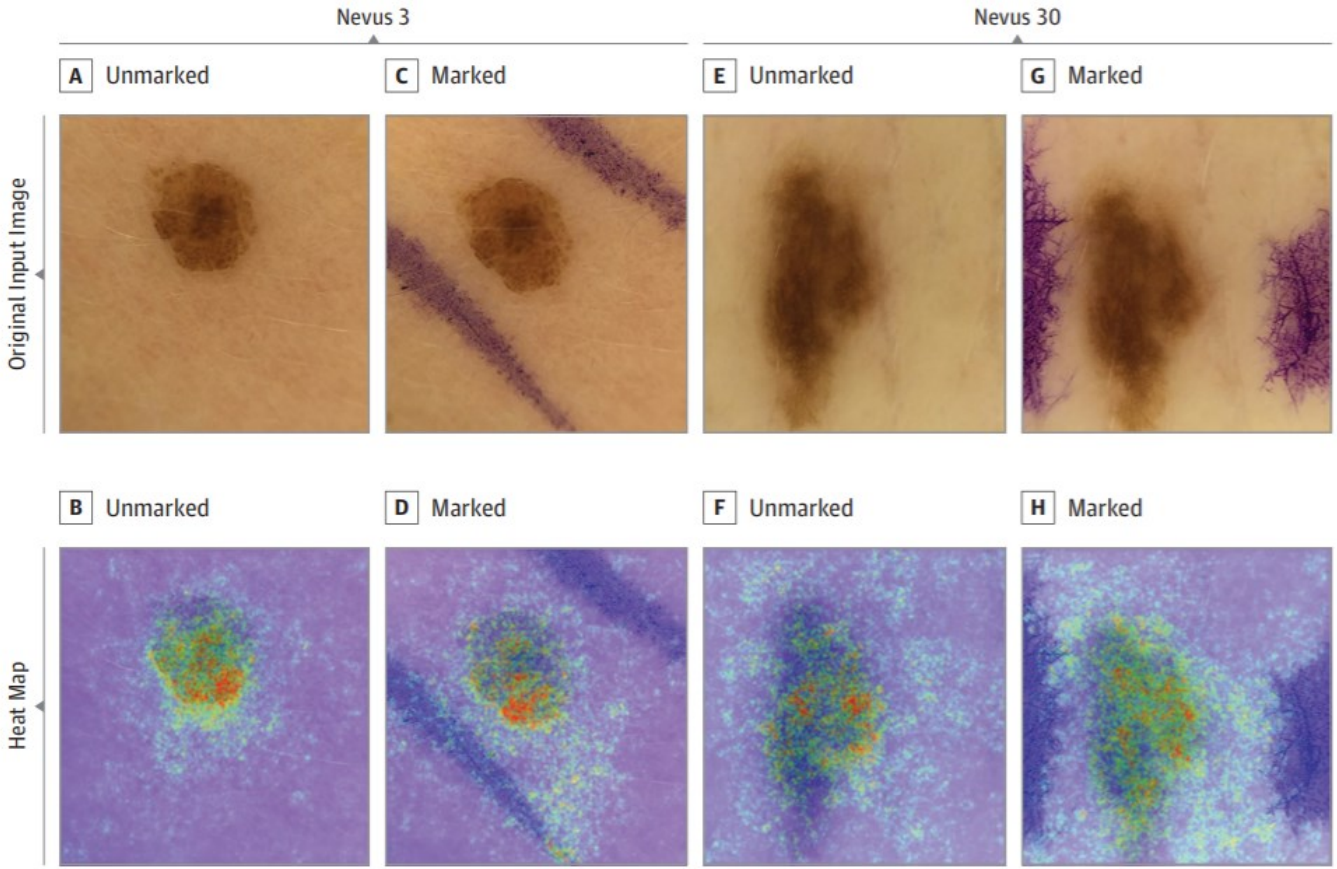
IMPORTANCE Deep learning convolutional neural networks (CNNs) have shown a performance at the level of dermatologists in the diagnosis of melanoma. Accordingly, further exploring the potential limitations of CNN technology before broadly applying it is of special interest.

Editorial page 1105

Supplemental content

https://jamanetwork.com/journals/jamadermatology/fullarticle/2740808

Figure 3. Heat Maps of 2 Benign Nevi With Unchanged Melanoma Probability Scores After Addition of In Vivo Skin Markings



ining Medicine

Statistical perspective

Statistical thinking provides strategies and methods to answer scientific questions.

- Problem translation and formulation
  - Question types (i.e. description, prediction, explanation, intervention)
- Designing appropriate analytical strategies and methodology
  - Design (mapping questions to analytical strategies)
  - Designing experiments vs strategies for found data
- Understanding measurement, variation, bias and uncertainty

There are no routine statistical questions, only questionable statistical routines.
(David R. Cox)

Estimands: a framework for thinking

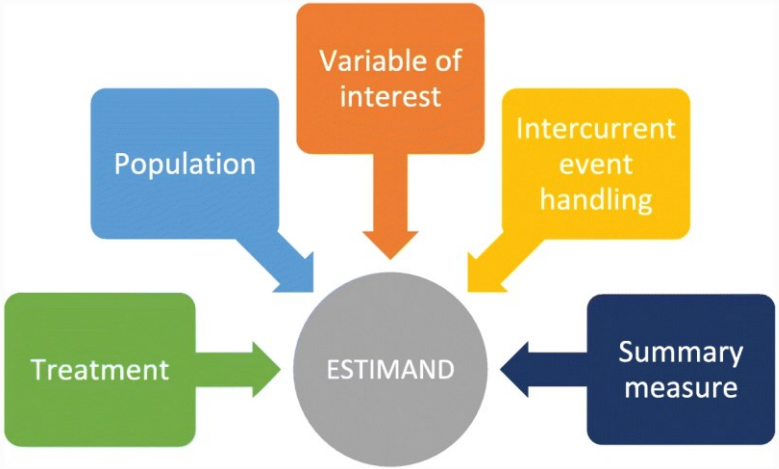
EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

1 30 August 2017
2 EMA/CHMP/ICH/436221/2017
3 Committee for Human Medicinal Products

4 ICH E9 (R1) addendum on estimands and sensitivity
5 analysis in clinical trials to the guideline on statistical
6 principles for clinical trials

ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials - Step 2b (europa.eu)

From: What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials?



The five attributes of an estimand according to the ICH E9 (R1) addendum

Lawrance, R., Degtyarev, E., Griffiths, P. et al.

What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials?. J Patient Rep Outcomes 4, 68 (2020). https://doi.org/10.1186/s41687-020-00218-5

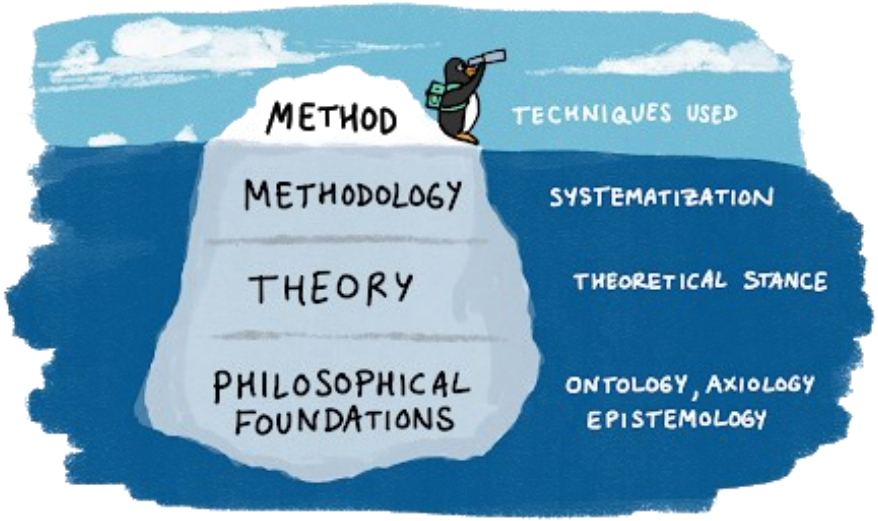
NOVARTIS | Reimagining Medicine

Scientific perspective

World view (systems thinking)

- Provides the context – the why
- Ensures clarity on the purpose, outcome, value and impact
- Ensures that prior knowledge can be navigated and leveraged.
- (Scientific) theories determine what to measure and how

exploring the context—obtaining sufficient background information to formulate the problem carefully
(Chris Chatfield)



Ethical perspective

We need less research, better research, and research done for the right reasons
(Doug Altman)

Professional code of conduct – guiding principles:

- Selflessness: Place the needs and concerns of those who depend on us above our own, and prevent harm
- Skill: Continuously aim for excellence in our knowledge and skill
- Trustworthiness: Take responsibility for personal behavior and conduct
- Discipline: Follow prudent procedure and functioning with others

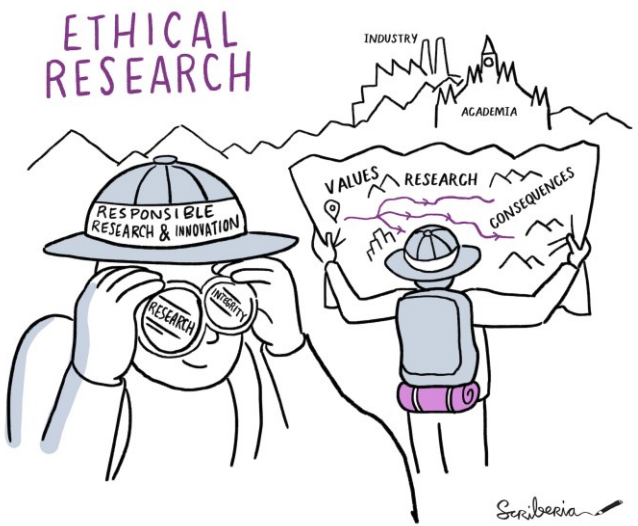


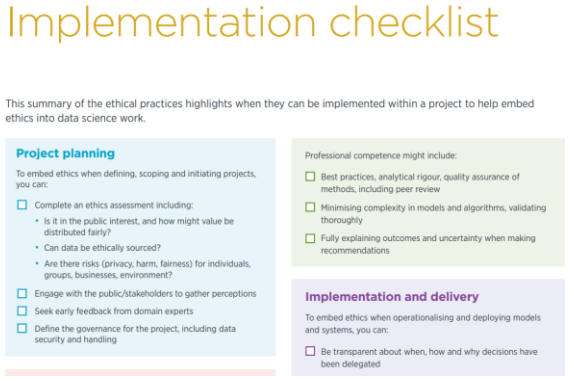
Figure: Turing Way Community (2019) A Handbook for Reproducible Data Science. Zenodo.

Institute and Faculty of Actuaries

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

A Guide for Ethical Data Science

A collaboration between the Royal Statistical Society (RSS) and the Institute and Faculty of Actuaries (IFoA)



Implementation checklist

This summary of the ethical practices highlights when they can be implemented within a project to help embed ethics into data science work.

Project planning

To embed ethics when defining, scoping and initiating projects, you can:

- ☐ Complete an ethics assessment including:
  - Is it in the public interest, and how might value be distributed fairly?
  - Can data be ethically sourced?
  - Are there risks (privacy, harm, fairness) for individuals, groups, businesses, environment?
- ☐ Engage with the public/stakeholders to gather perceptions
- ☐ Seek early feedback from domain experts
- ☐ Define the governance for the project, including data security and handling

Professional competence might include:

- ☐ Best practices, analytical rigour, quality assurance of methods, including peer review
- ☐ Minimising complexity in models and algorithms, validating thoroughly
- ☐ Fully explaining outcomes and uncertainty when making recommendations

Implementation and delivery

To embed ethics when operationalising and deploying models and systems, you can:

- ☐ Be transparent about when, how and why decisions have been delegated

https://rss.org.uk/RSS/media/News-and-publications/Publications/Reports%20and%20guides/A-Guide-for-Ethical-Data-Science-Final-Oct-2019.pdf

NOVARTIS | Reimagining Medicine

https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf

Human perspective

“What’s the collective noun for a group of statisticians?”
"A quarrel.“ (John Tukey)

Strive for balance: prevent one person or perspective from dominating a team and leading to weaker solutions

Communication across fields: understanding & openness to learn from each other

Ensure reproducibility, replicability, ...

Translating, assimilating and operationalizing knowledge

Figure: Turing Way Community (2019) A Handbook for Reproducible Data Science.

Data science requires communication and collobration

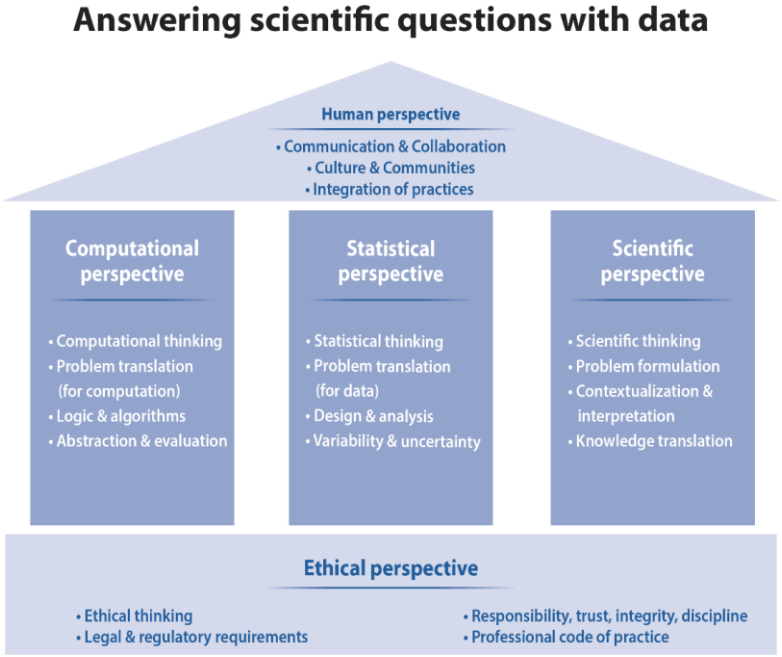
Good data science practice takes a holistic approach, and includes people that **collectively & collaboratively** span a range of perspectives

Shared principles provide a starting point and compass, but good practice is a journey!

An open mind and **balancing diverse skillsets are key** – with a mindset of continuously improving how we work, learn, and communicate as an interdisciplinary team of quantitative scientists.

Practicing good data science is a team sport: the team greatly benefits if we can communicate our deep, diverse skillsets in an understandable manner!

"What would data science look like if its key critics were engaged to help improve it, and how might critiques of data science improve with an approach that considers the day-to-day practices of data science?" (Neff et al. 2017)



Acknowledgments

- Conor Moloney
- Carsten Philipp Mueller
- Lukas Widmer
- Jelena Cuklina
- Frank Bretz
- Janice Branson
- David Ohlssen