# The case for data

Grace Baynes

Data & New Product Development Director, Open Research, Springer Nature

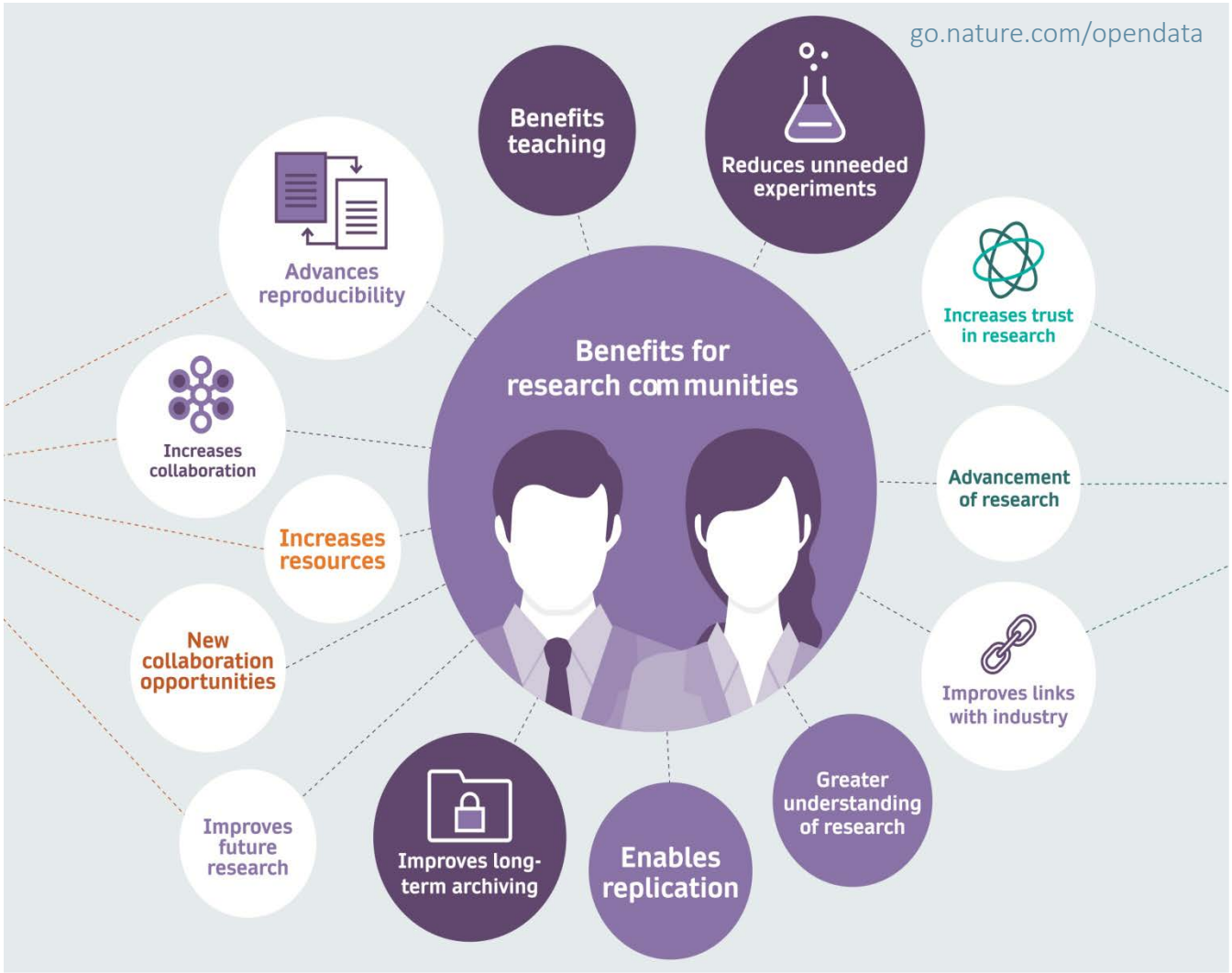
2017



SPRINGER NATURE

# THE CASE FOR DATA

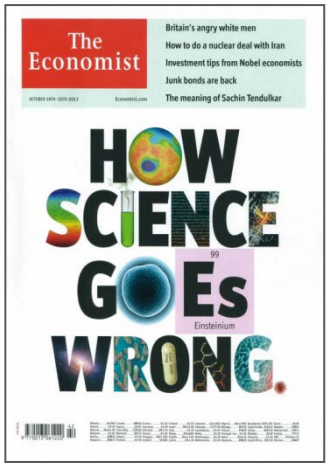
*Reproducibility crisis creates a compelling case for more openness*



# 2.1. THE CASE FOR DATA: REPRODUCIBILITY IS A CRITICAL ISSUE IN RESEARCH

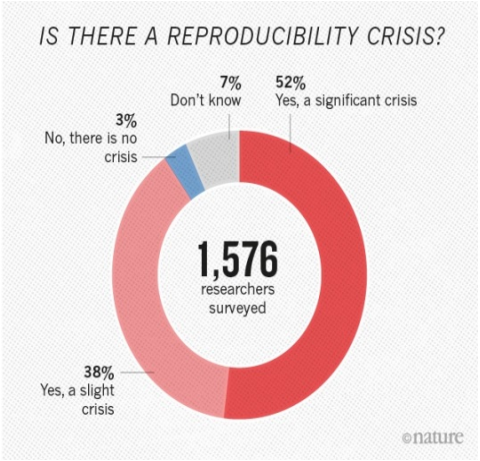
*Data availability has been shown to enable reproducibility*

## Evidence is mounting on costs & scale of the issue




- Irreproducible biology research costs US $28 billion per year[1]
- Pharma companies report 75%+ failure rates replicating conclusions of peer-reviewed papers [2,3]

## A recent *Nature* survey[4] highlights concern in the research community




>50% of researchers couldn't reproduce their own experiments

>70% couldn't reproduce the work of others

## There is evidence that data availability increases reproducibility

A study[5] of eighteen *Nature Genetics* papers found :

- Two could be reproduced fully
- Six were reproduced partially
- Ten could not be reproduced

*"The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis."*

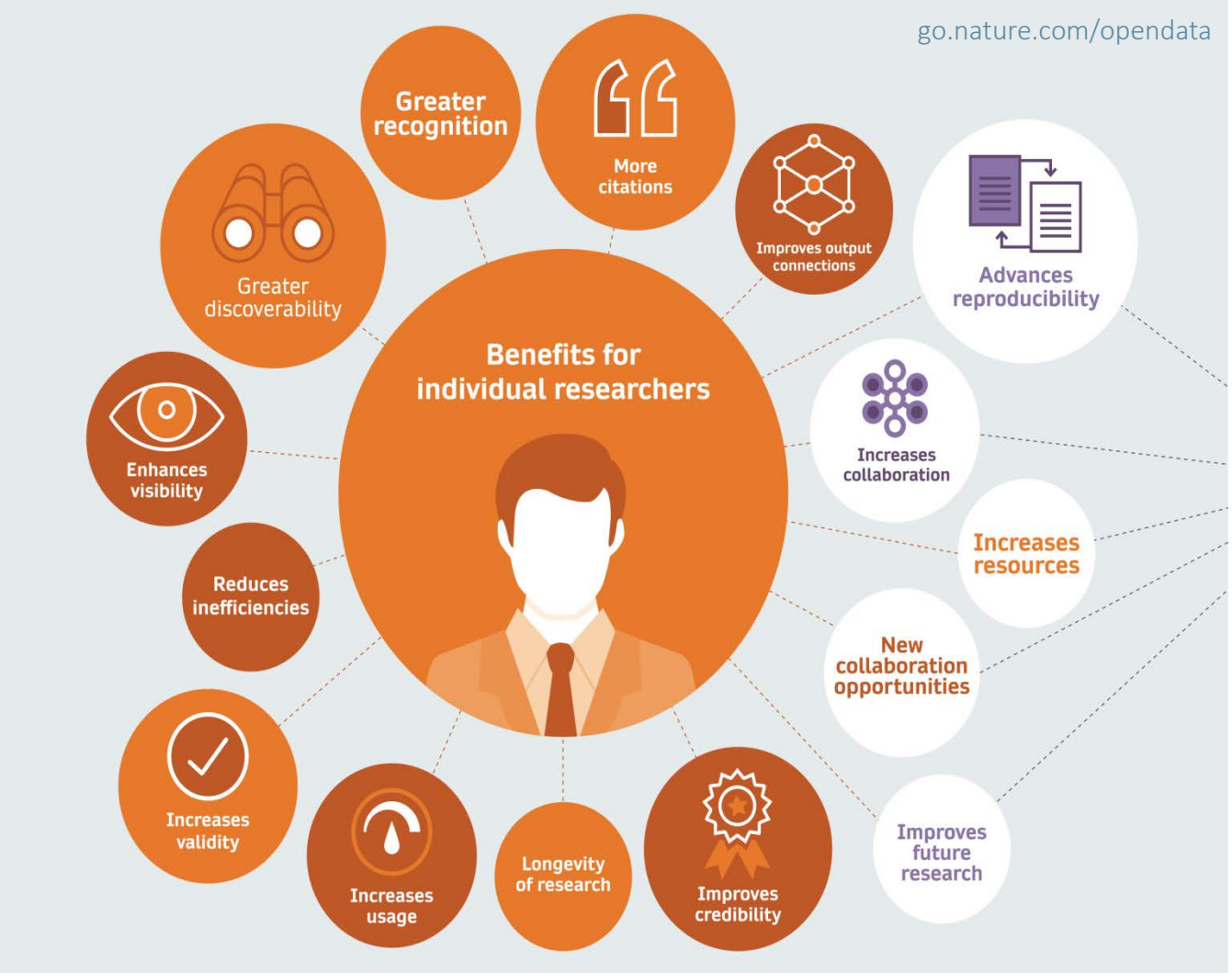
— *Nature Genetics* **41,** 149–155 (2009)

1. Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. *PLoS Biol.* **13**, e1002165 (2015) http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165
2. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012), 3. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10**, 712 (2011)
4. Baker (2015) http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970
5. Ioannidis et al (2009) https://www.nature.com/ng/journal/v41/n2/full/ng.295.html

SPRINGER NATURE

# THE CASE FOR DATA

*Sharing data has benefits for researchers, and is not just a public good*



# THE CASE FOR DATA: EVIDENCE OF BENEFITS TO RESEARCHERS & SCIENCE

*Citation advantage & productivity increase in multiple fields*

## Data archiving can double the publication output of studies

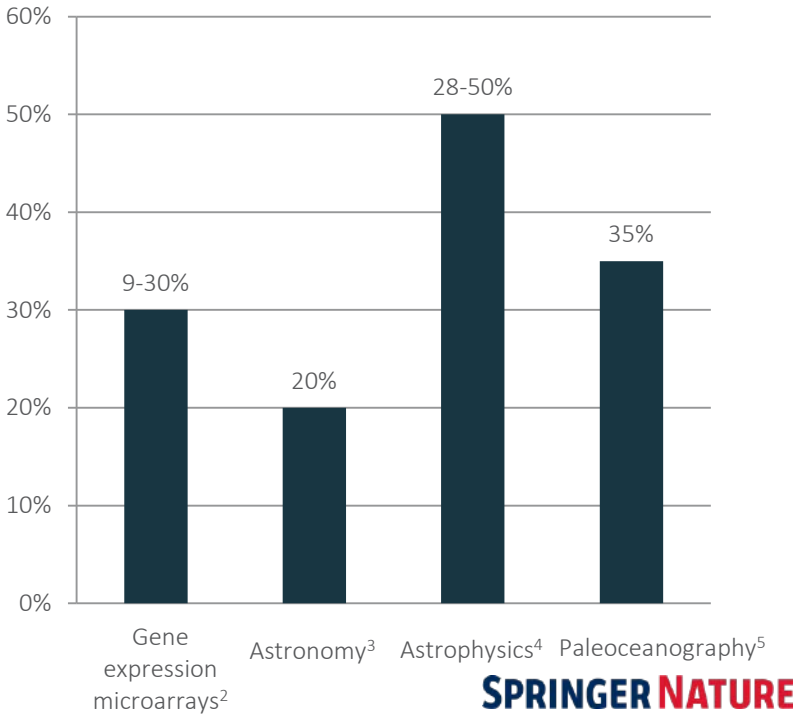
A study of 7,000 NSF and NIH research projects in social sciences found that:

- Those with archived data resulted in 10 (median) publications;
- Those without archived data resulted in 5 publications[1]

Principal investigators who archived their data were more likely to publish more articles per project, and to see others build on their work

## Research articles with open data are cited up to **50% more**

Analysis shows that articles with data available are cited 9-50% more, depending on the field

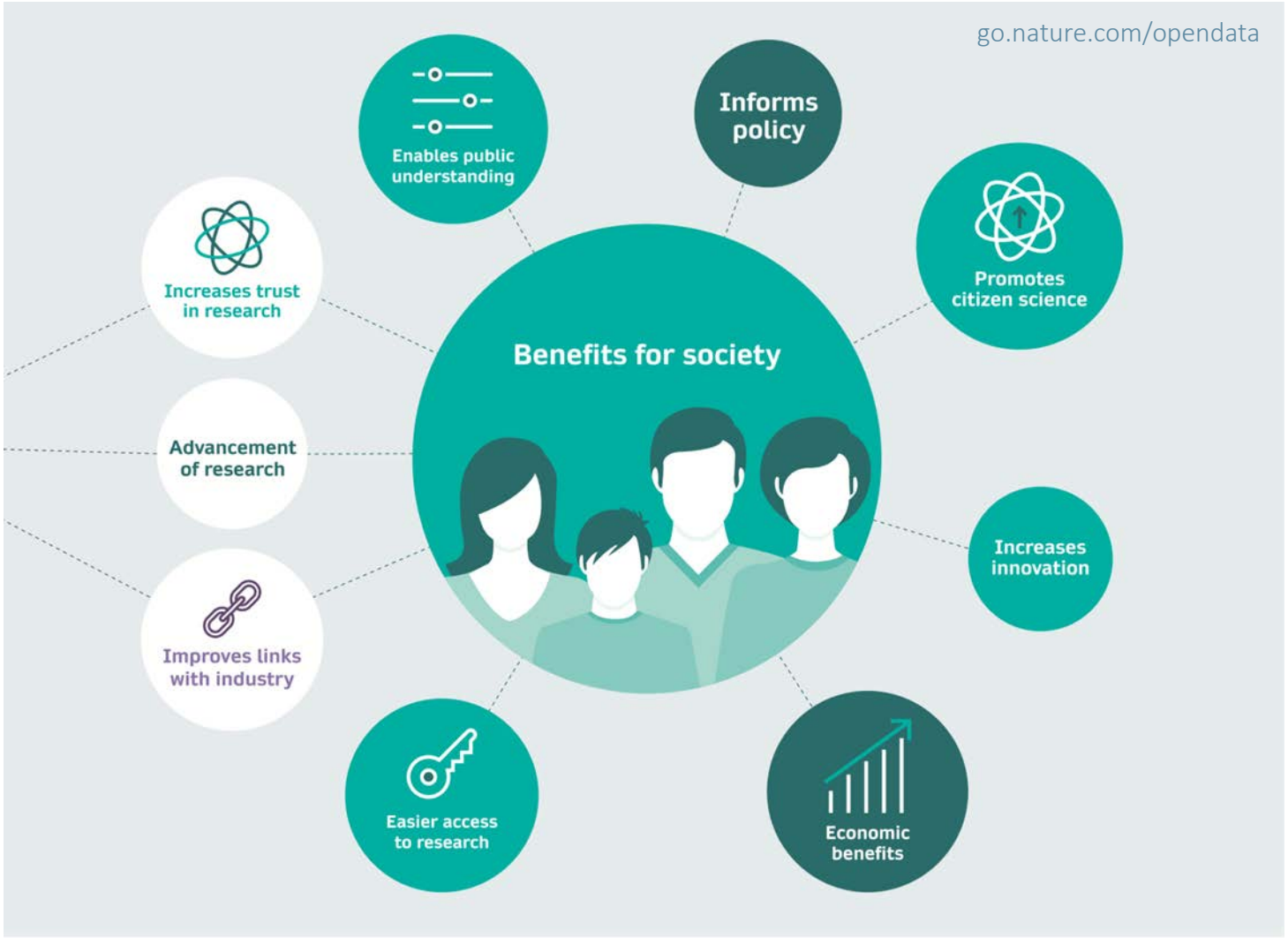




1. Pienta et al (2010) https://deepblue.lib.umich.edu/handle/2027.42/78307
2. Piwowar & Vision (2013) https://doi.org/10.7717/peerj.175
3. Henneken & Accomazzi (2011) https://arxiv.org/abs/1111.3618
4. Dorch et al (2015) https://arxiv.org/abs/1511.02512
5. Sears et al (2011) https://figshare.com/articles/Data_Sharing_Effect_on_Article_Citation_Rate_in_Paleoceanography/1222998/1

# THE CASE FOR DATA

*Some compelling evidence of benefits to society, more is needed*



# **CASE FOR DATA:** SOME EVIDENCE OF SOCIETAL BENEFITS

*Two case studies show significant benefit to the economy*

## CASE STUDY: Human Genome Project


**$14.5 billion:** Total US government investment in sequencing human genome

**$1 trillion:** Estimated contribution to the US economy in the decade since, according to a report by the Battelle Memorial Institute[1]

**178-to-1:** The return on investment for every US$ spent

The whole human genome sequence data is open to anyone

## CASE STUDY: European Bioinformatics Institute


**£47 million:** Annual operating cost

**£1 billion**: Annual efficiency savings to researchers worldwide, according to an independent report[2]

**£920 million:** Estimated annual estimate of future research impacts

The European Bioinformatics Institute (EBML-EBI) is a UK-based organization that hosts and manages large open datasets in various disciplines of the life sciences

More evidence is needed

1. http://www.unitedformedicalresearch.com/advocacy_reports/the-impact-of-genomics-on-the-u-s-economy
2. http://www.ebi.ac.uk/about/news/press-releases/value-and-impact-of-the-european-bioinformatics-institute

researchdata.springernature.com
@grace_baynes

SPRINGER NATURE