

Tessa Pronk, researcher
KWR Watercycle Research Institute
Utrecht University Library

Data sharing games

.....

To deposit or not to deposit, that is the question

Structure:

- Research data sharing: (dis)advantages
- Why a model?
- Results
- Model take home messages

'The Data Gap'

67%

Said lack of access to others' data is a major impediment to scientific progress.

84%

Scientists who said that they would use other researchers' data sets if they were easily available.

36%

Scientists who said that others can access their data easily.

Visualisation from:

Nature 495, 430–432 (28 March 2013) | doi:10.1038/495430a

Data from:

C. Tenopir et al. PLoS ONE 6, e21101; 2011

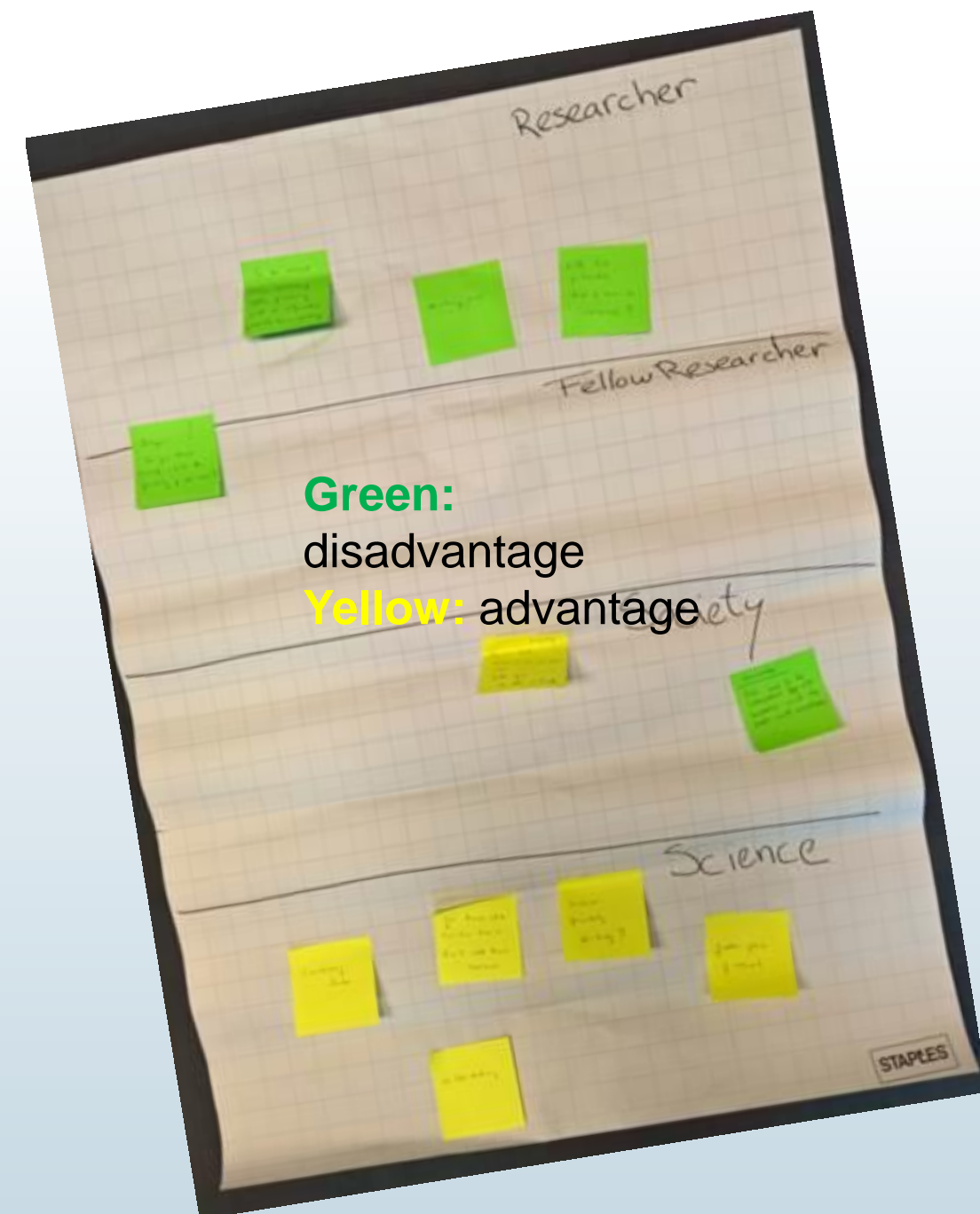
More disadvantages
than advantages?

Advantages for 'others',
disadvantages for sharing researcher?

Why a model?

Structure the problem. Define players, options, factors.
Make explicit your assumptions. Gain insight.

**“all models are wrong, but
some are usefull”**



<https://peerj.com/articles/1242/>

A game theoretic analysis of research data sharing

Tessa E. Pronk¹, Paulien H. Wiersma¹, Anne van Weerden¹ and Feike Schieving²

¹ Utrecht University Library, Utrecht University, Utrecht, The Netherlands

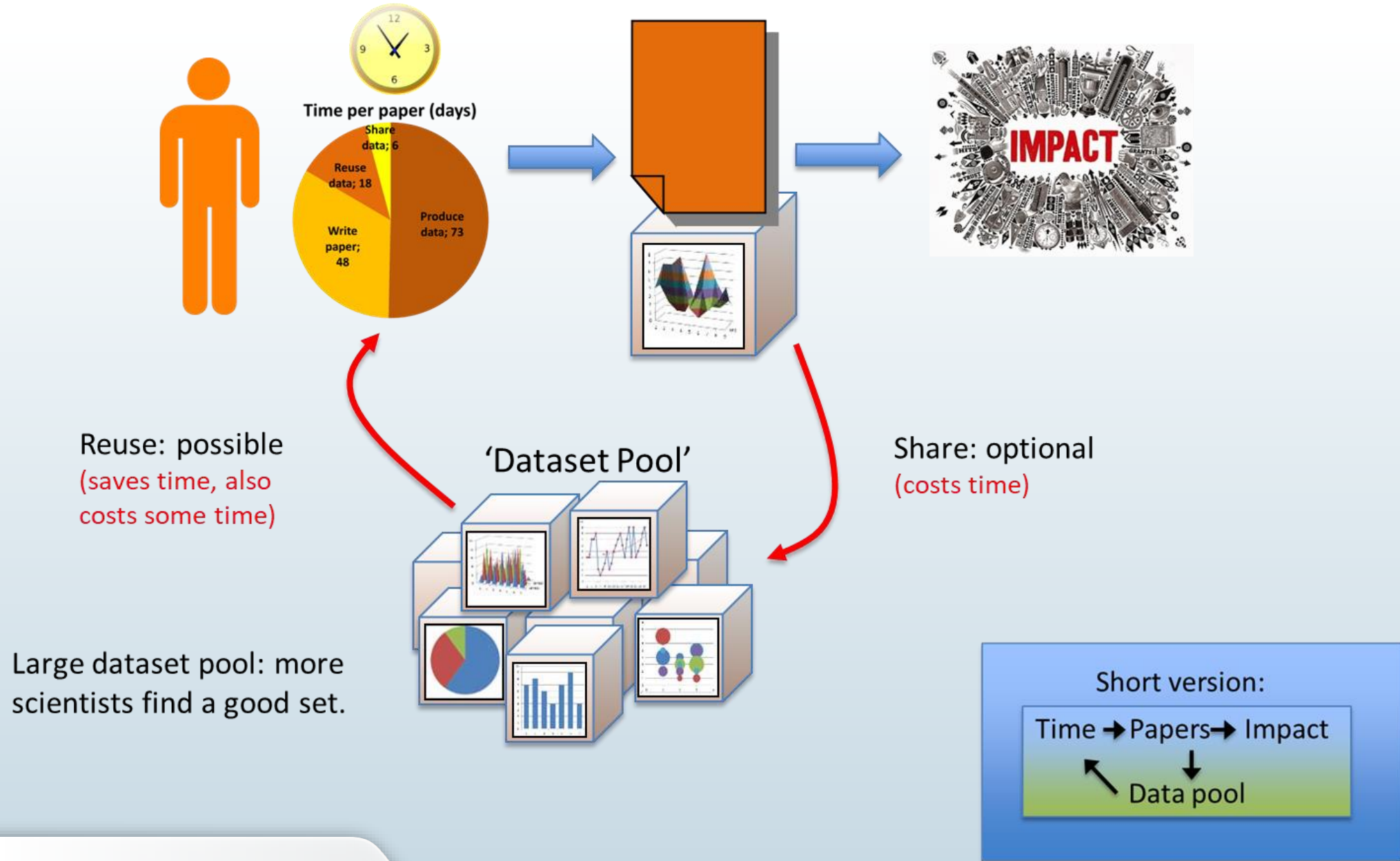
² Ecology and Biodiversity, Utrecht University, Utrecht, The Netherlands

Game theory is: - a way of structuring

- competition or conflict
- decision making
- optimal strategy (cost and benefit)
- interaction with other players

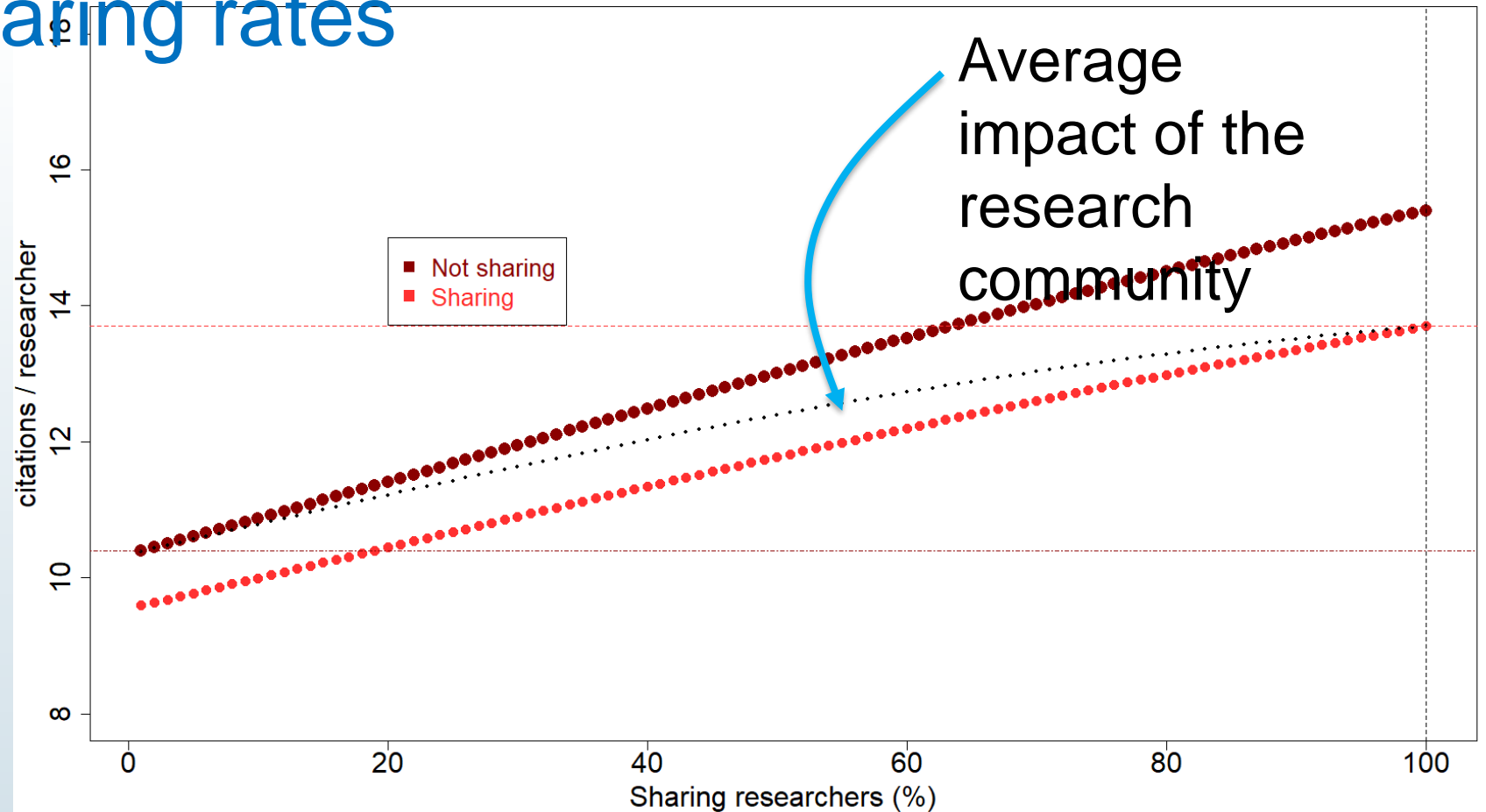
The model

NO sharing / reusing: about 3 papers per year



Results:

Research community yearly citations sharing / not sharing researchers at different sharing rates

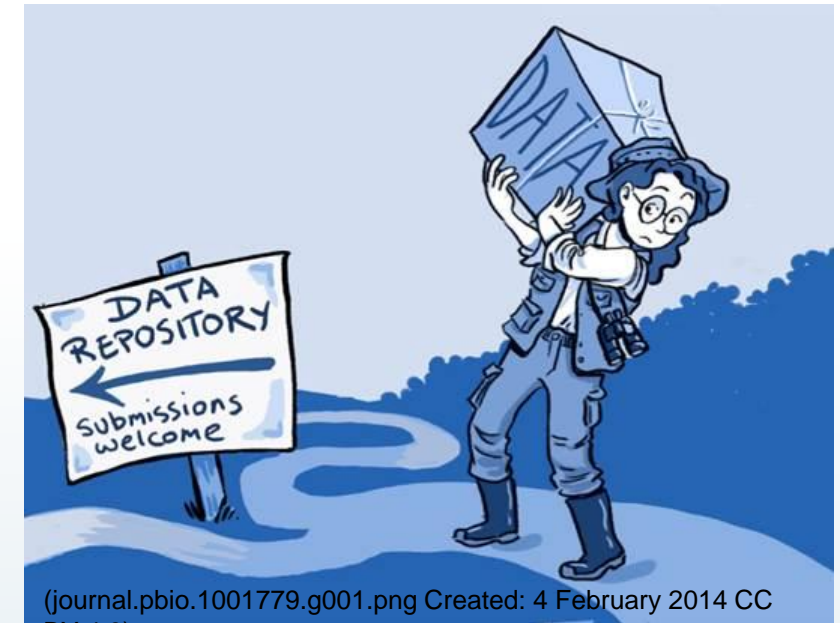


Sharing research is, in the end, profitable for all.

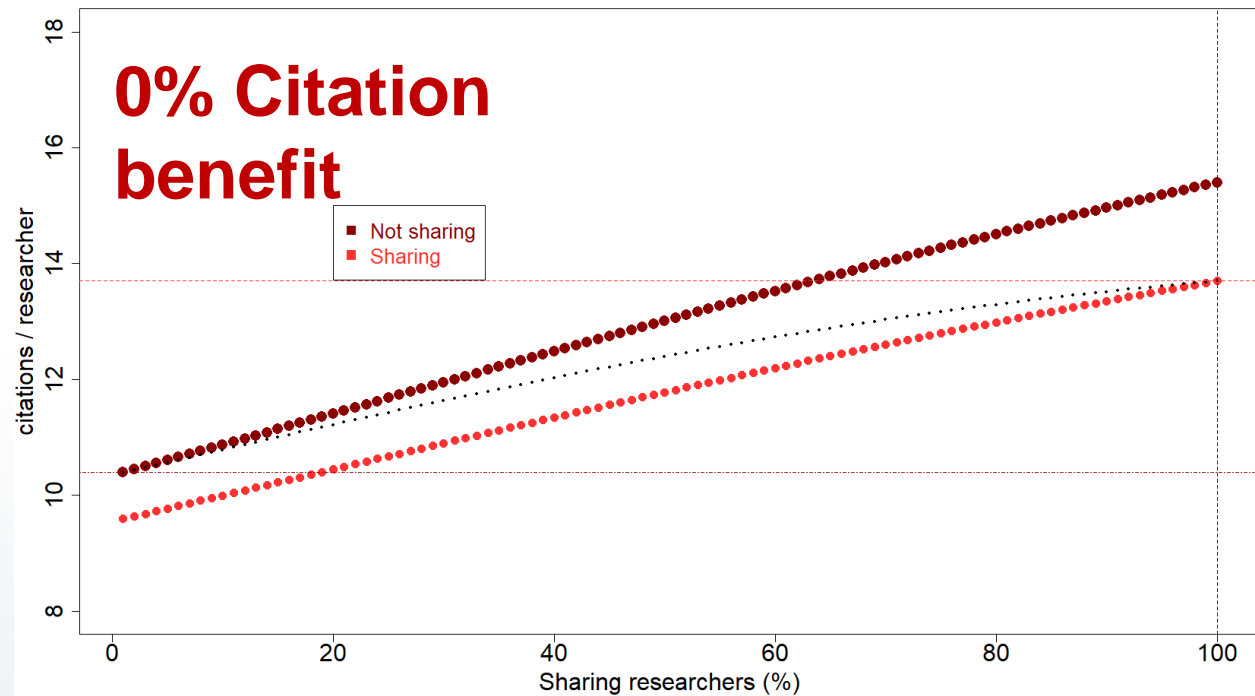
However,

The optimal strategy for the community,
- sharing -
is not the optimal strategy for the individual.

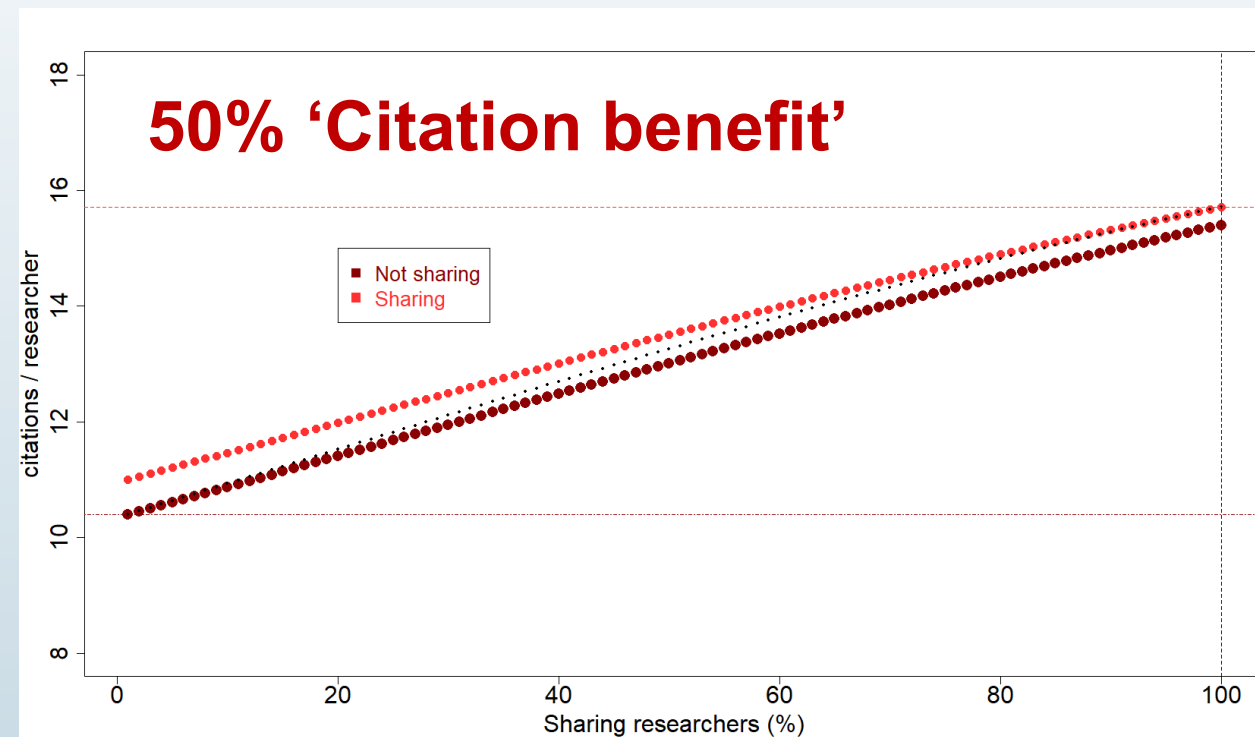
The strategy is, under our standard assumptions, *unstable*. There is a conflict of interest.



(journal.pbio.1001779.g001.png Created: 4 February 2014 CC BY 4.0)



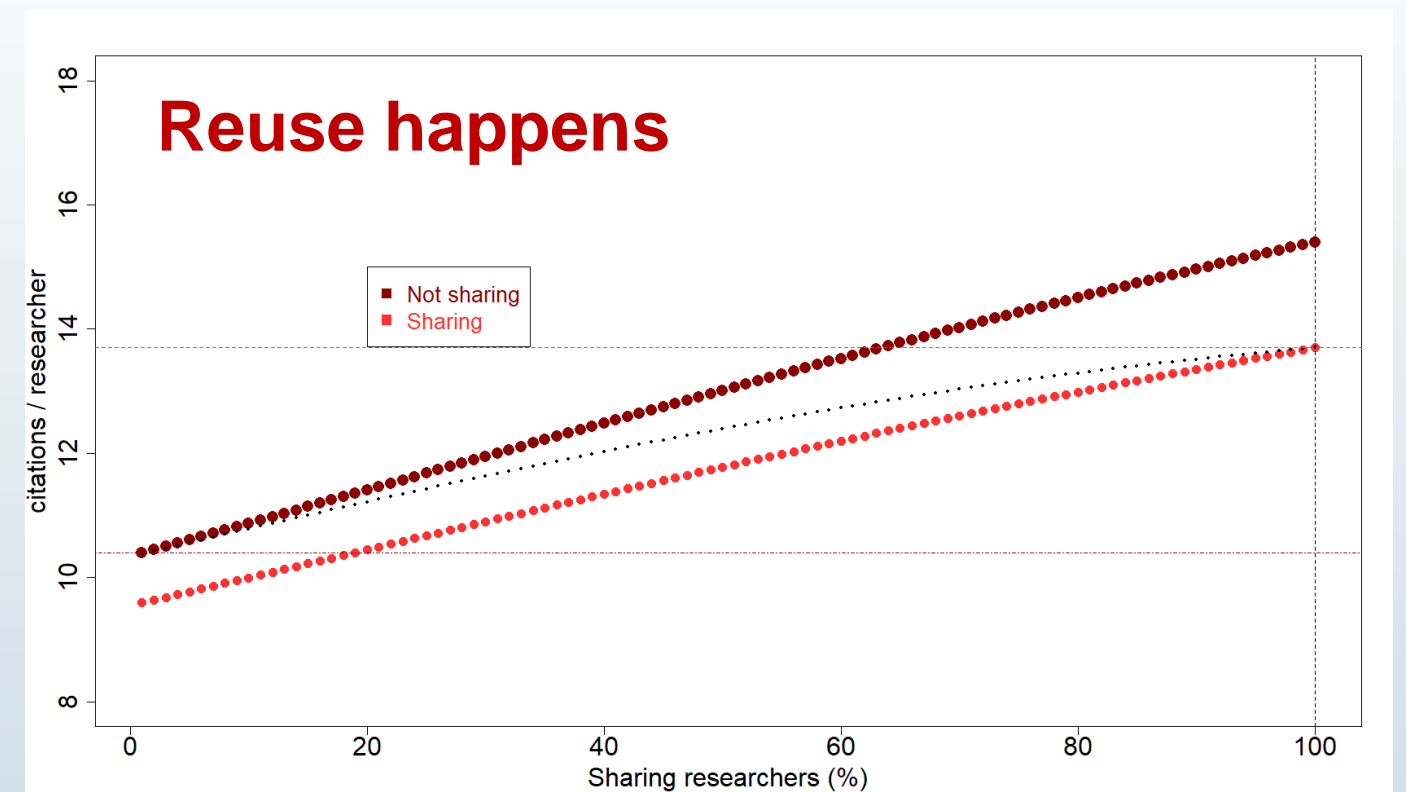
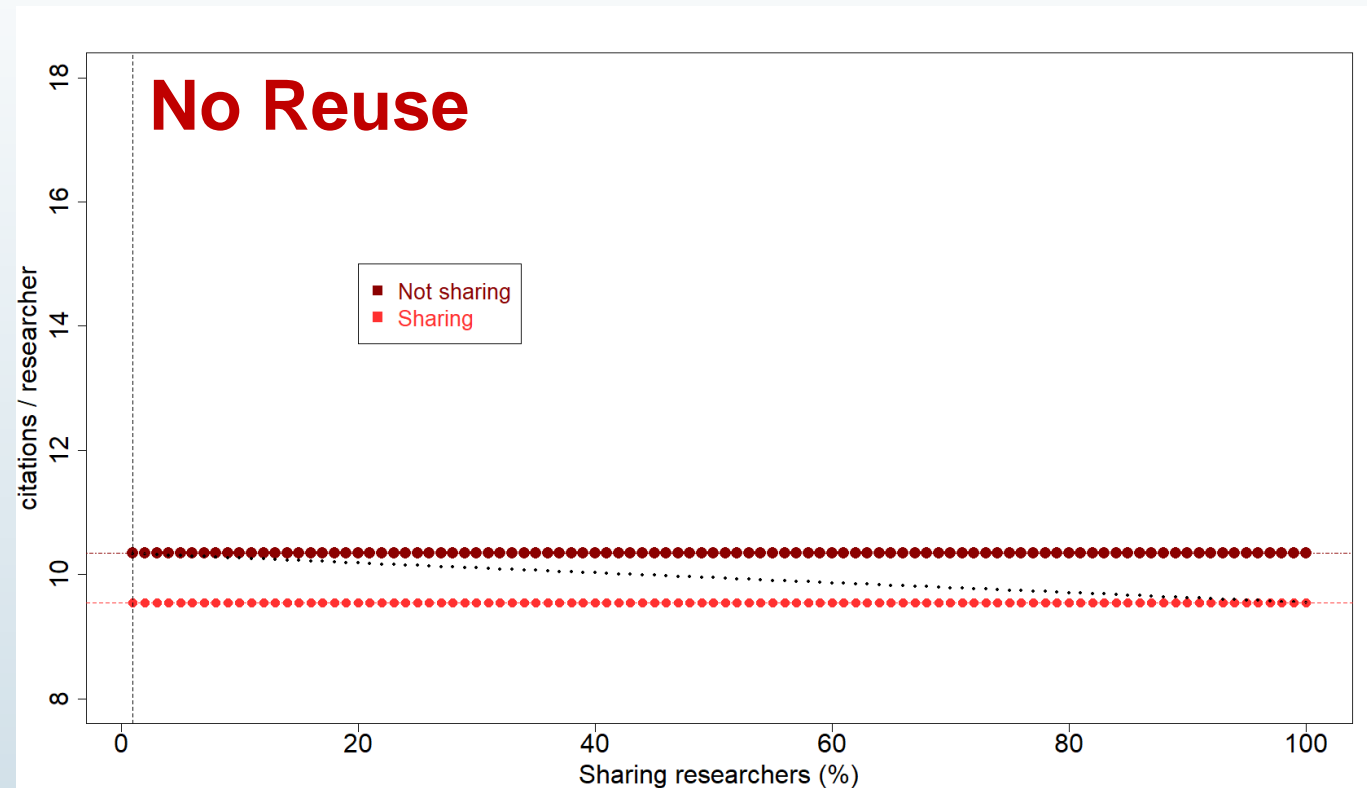
Remedy:
Get the average for sharing researchers
- ABOVE -
That of *not* sharing researchers.



Making data available would in that case become the most profitable, and therefore stable, strategy. This means researchers would willingly make their datasets available, and arguably in the best possible way to enable reuse.

Complication:

Reuse is a defining factor for efficiency with sharing

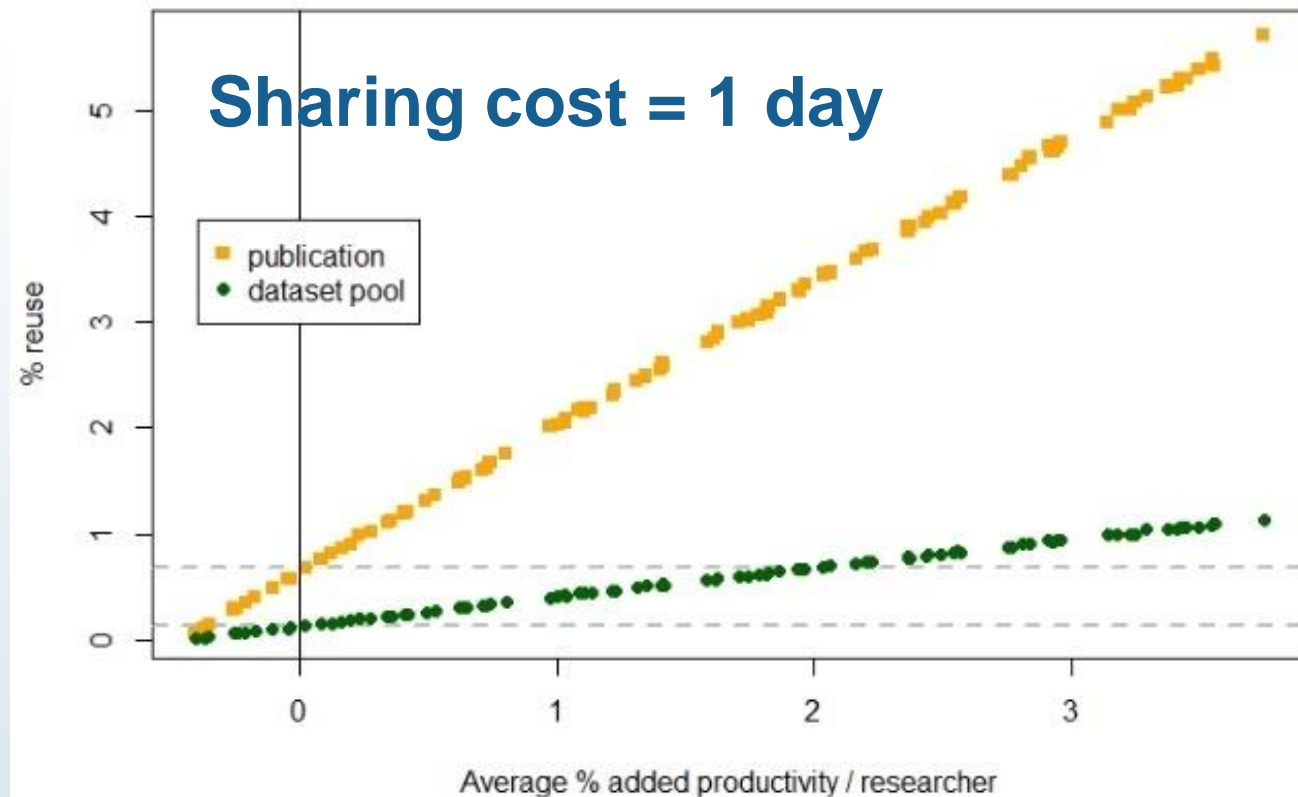




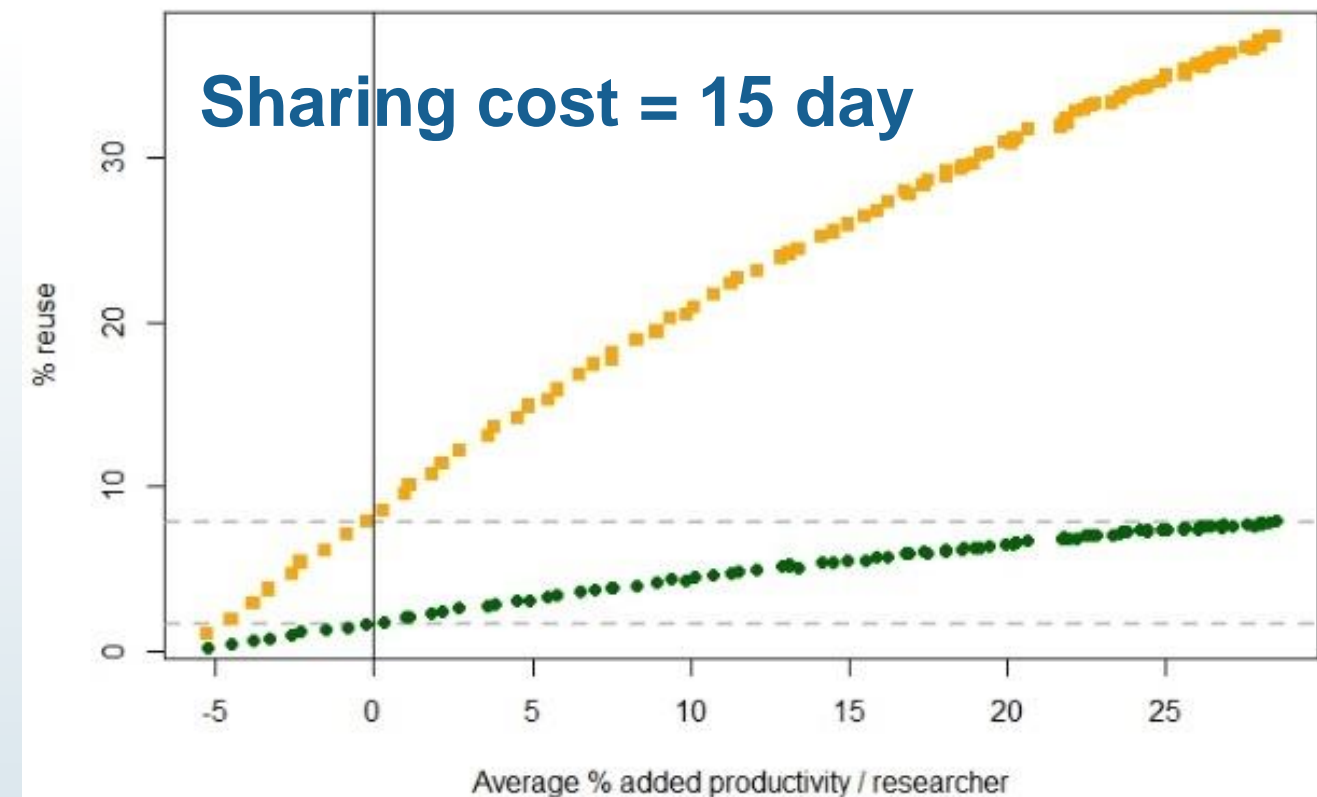
Considering time costs for sharing

What **reuse level** needs to be achieved to break even?

Break-even point: Time costs for sharing = Time gain by



Minimum reuse = 0.7 % of papers
or 0.2 % of available datasets



Minimum reuse = 8 % of papers
or 2% of available datasets

<https://tessapronk.shinyapps.io/ReuseResearchDataMinimum/>



Messages from the model:

- Incentives for sharing researchers
- Stimulate reuse
- Be critical on what to share
 - ✓ reuse foreseen?
 - ✓ investment worthwhile?

Thanks to:

Anne van Weerden
Paulien Wiersma
Feike Schieving

University Library Utrecht
University Library Utrecht
Ecology and biodiversity, Utrecht University