Computational Challenges in Modeling & Simulation of Complex Systems

OPEN SCIENCE

Simon J.E. Taylor

Modelling & Simulation Group
Department of Computer Science, Brunel University London (UK)

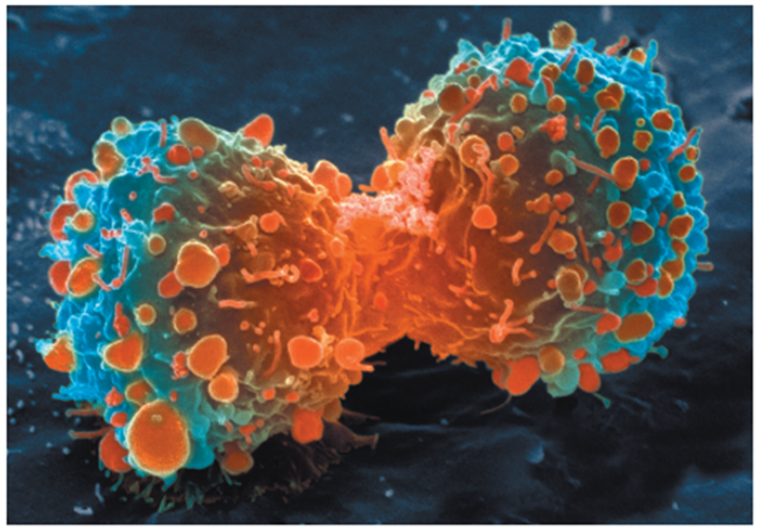
Brunel University London

Modelling & Simulation Group

A Reproducibility Crisis in Science!

COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534 | EARTH SYSTEMS Past climates give valuable clues to future warming p.537 | HISTORY OF SCIENCE Descartes' lost letter tracked using Google p.540 | OBITUARY Wylie Vale and an elusive stress hormone p.542



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will

investigators must reassess their approach translating discovery research into gre clinical success and impact.

Many factors are responsible for the h failure rate, notwithstanding the inh ently difficult nature of this disease. C tainly, the limitations of preclinical to

47/53 "landmark" publications could not be replicated

Begley and Ellis, *Nature* **483**, 531–533 (29 March 2012) doi:10.1038/483531a

Must try harder

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorou at the data — and at themselves.

Error prone

Biologists must realize the pitfalls of work on massive amounts of data.

If a job is worth doing, it is worth doing twice

Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues Jonathan F. Russell.

The case for open computer programs

Six red flags for suspect work

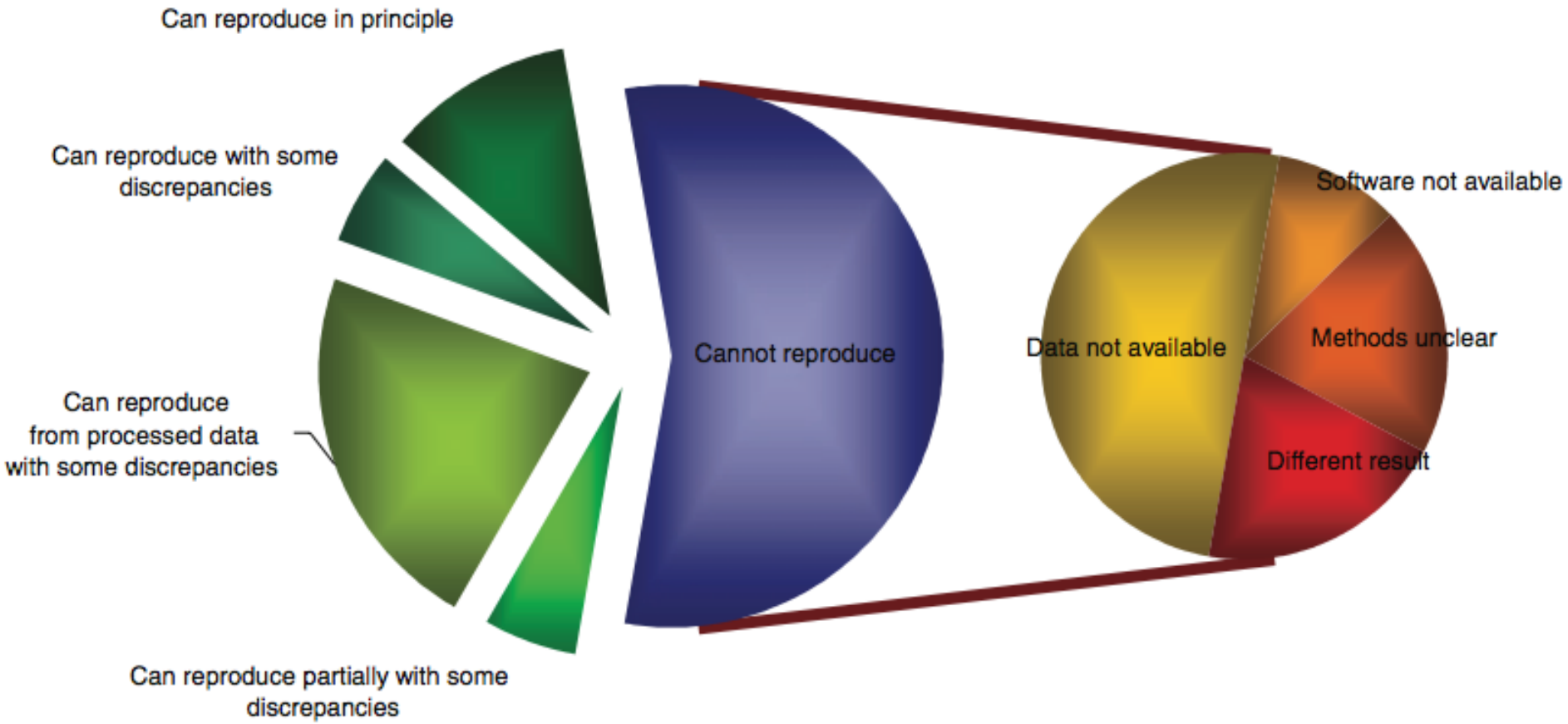
C. Glenn Begley explains how to recognize the preclinical papers in which the data won't stand up.

Know when your numbers are significant

Sci-GaIA
www.sci-gaia.eu

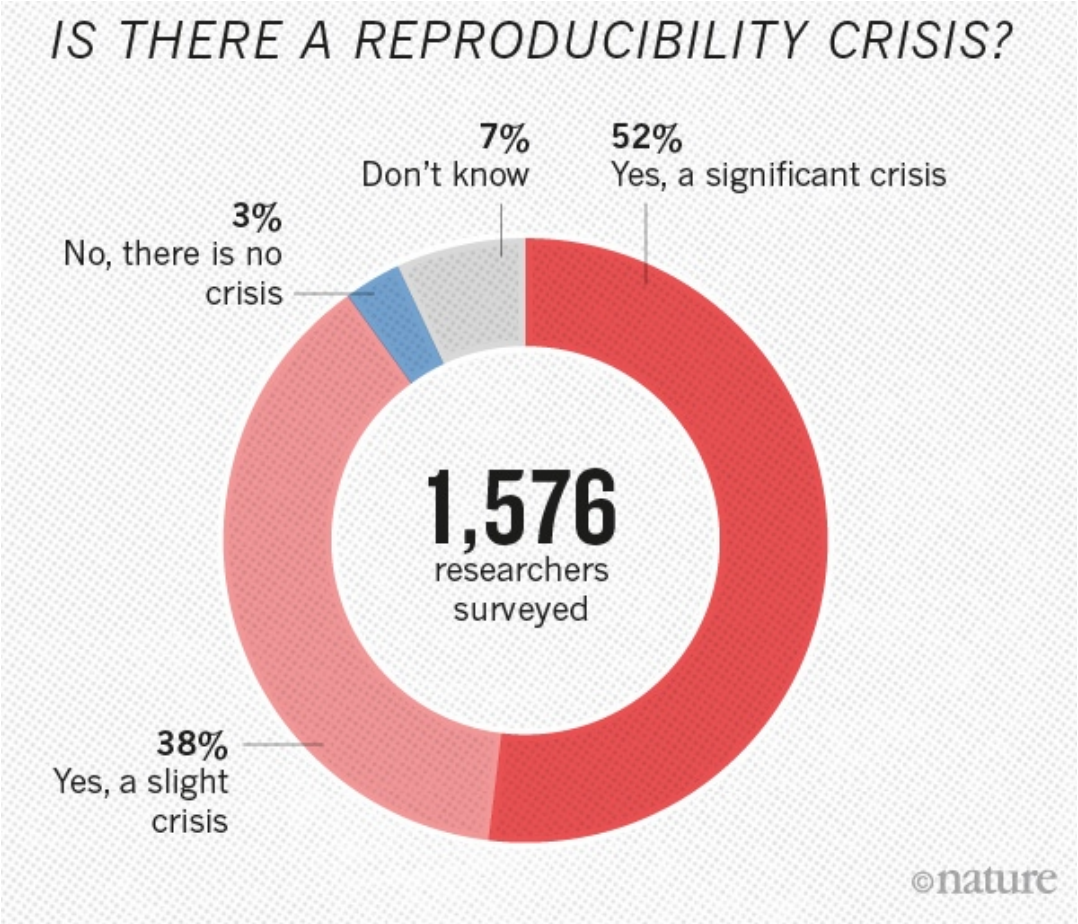
"We evaluated 18 articles published in *Nature Genetics* in 2005 or 2006 that presented data from comparative analyses of microarrays experiments that had not been previously published elsewhere."

Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses. *Nature Genetics* 41: 14



"More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research."

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



A Reproducibility Crisis in M&S?

- Rahmandad and Sterman (2012) sampled one year of articles from System Dynamics Review and found that out of **27 models 16 (59%) included no equations at all while 2 (7%) reported 'some' equations**.
- Kurkowski, Camp, and Colagrosso (2005) reviewed **114** discrete-event simulation models of Mobile Ad Hoc Networks (MANETS) and **found that all papers had issues in reporting key aspects of the method used**… terminating/steady state, no pseudo random number generator details; initialization details, warm-up. 25% of studies did not state the simulation software in which the model was implemented.
- Grimm et al. (2006) focused on agent-based simulation models in ecology and drew similar conclusions that as the modelling becomes more complex, the potential and flexibility increases but the reproducibility decreases. **This means that the results of agent-based simulation models are rarely reproducible**.
- Janssen (2017) investigated the reproducibility of 2367 agent-based models returned from a search of ISI Web of Science. The study found that 50% of publications report complete or 'some' equations. **Source code for the models was only available for 10% of the publications; there was a general lack of transparency in how models work.**

Open Science

- "… **the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods**."

  The FOSTER project (Facilitate Open Science Training for European Research) (www.fosteropenscience.eu)

- Open Science therefore refers to **efforts to make the output of research more widely accessible** to scientific communities, business sectors, and society more generally

  (OECD 2015)

Potential Benefits of Open Science

(OECD 2015, p.18)

- **Improving efficiency** in science by reducing duplication and the costs of creating, transferring and reusing data; allowing more research from the same data; and multiplying opportunities for domestic and global participation in the research process.
- **Increasing transparency and quality** in the research validation process by allowing a greater extent of replication and validation of scientific results.
- **Speeding the transfer of knowledge** from research to innovation.
- Increasing knowledge spillovers to the economy and increasing awareness and conscious choices among consumers.
- **Addressing global challenges** more effectively by globally coordinated international actions.
- **Promoting citizens' engagement in science and research** – Open Science and open data initiatives may promote awareness and trust in science among citizens. In some cases, greater citizen engagement may lead to active participation in scientific experiments and data collection

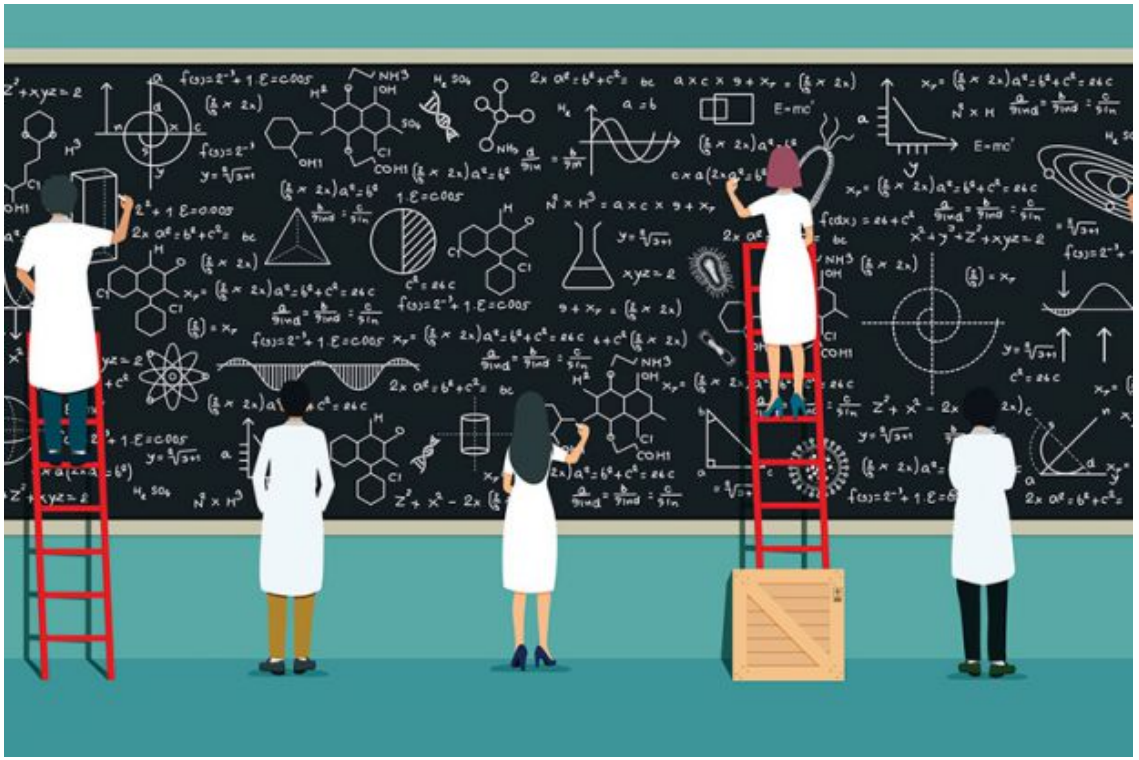
Many Funders are Moving Towards OS

- **European Commission** is promoting open (free of charge) access to scientific publications and research data as a core strategy for H2020-funded research projects (ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access)
- The **National Science Foundation** (NSF) has developed an outline framework for activities to increase public access to scientific publications and digital scientific data resulting from funded research (www.nsf.gov/news/special_reports/public_access).
- The Concordat on Open Research Data (www.rcuk.ac.uk/media/news/120621) has been produced by **HEFCE, Research Councils UK, Universities UK and the Welcome Trust** to guide the development of Open Science.

All emphasize the impact of Open Science on scientific progress

So how do we become more open? How do we openly share?

- Published articles?
- Data?
- Results?
- Models?
- Software?
- Computing Infrastructure?



https://www.timeshighereducation.com/blog/jisc-futures-research-age-open-science#survey-answer

How can we share?

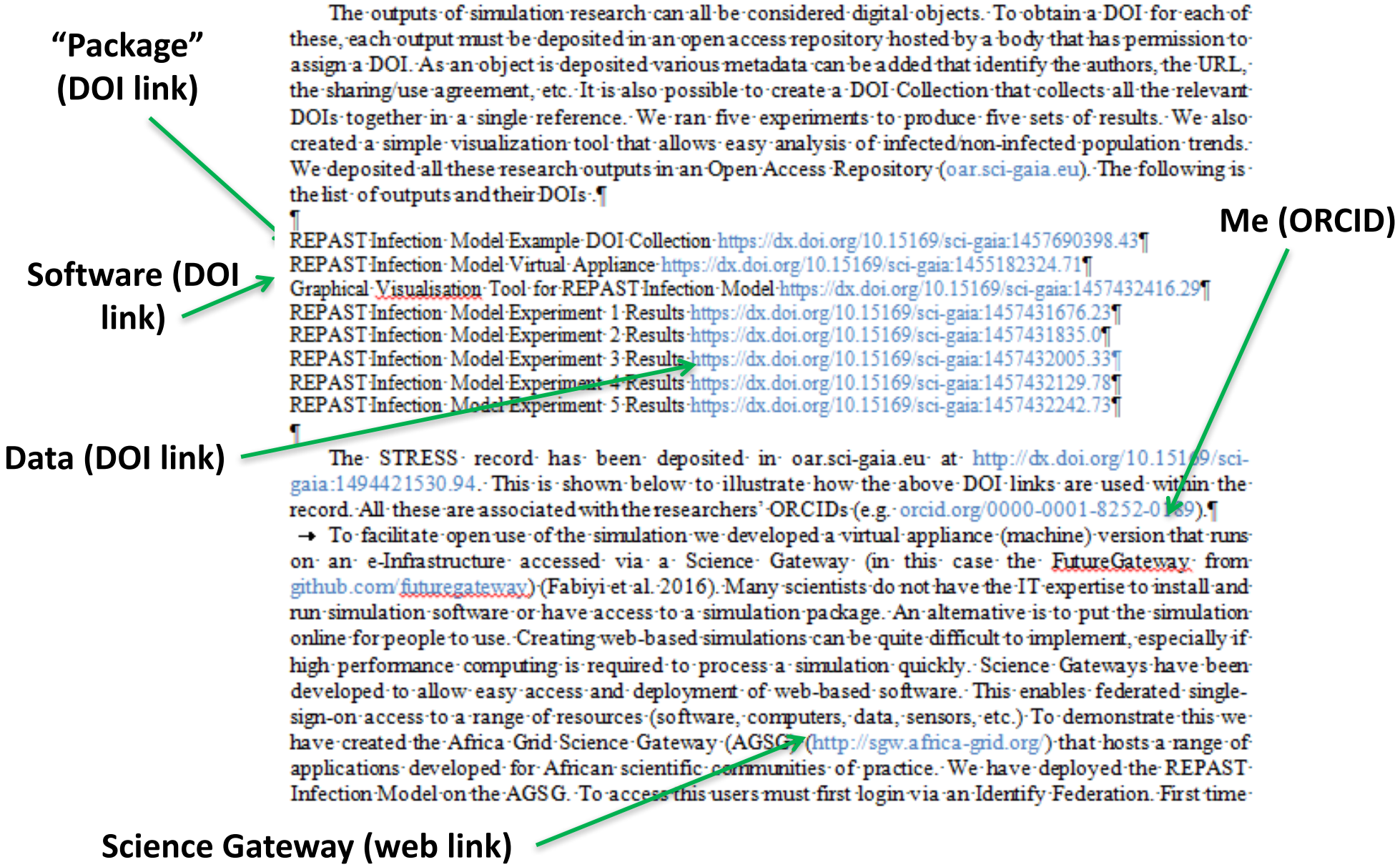
- Publish articles using green or gold open access
- Deposit data, results, model, software in an open access institutional repository with DOIs and using FAIR open data principles
- Use independent review or simulation reporting guidelines (e.g. STRESS)
- Develop web-based science gateways to facilitate access to the software and computational infrastructure (e.g. cloud, grid, etc.)
- Use one of the emerging open science cloud initiatives (e.g. European Open Science Cloud)
- Obtain an ORCID and link DOIs to ORCID
- Bring it all together in your published work

ORCID and DOIs
– uniquely identify yourself and your works

http://orcid.org/0000-0001-8252-0189

… but you can call me 0189

Bring it all together in your published work…

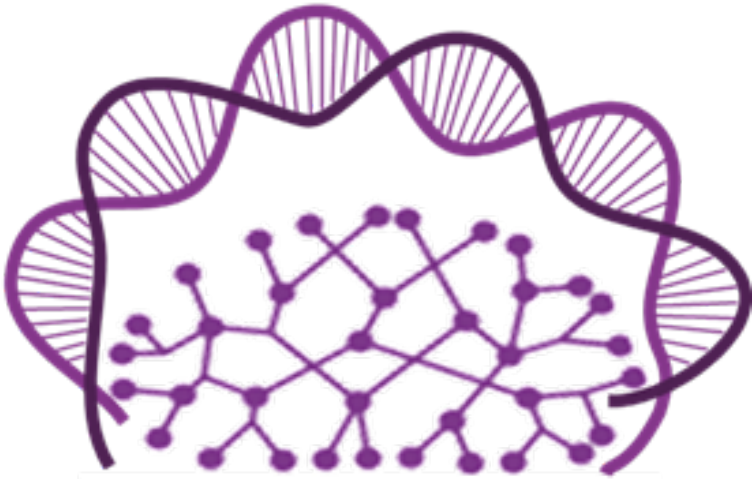


(Grand) Challenge
Can we as a community become open?

Work together, share together, build together… Simulation Everywhere!

open science

G. Emmerich



open m&s

A. Anagnostou