



Open science e etica

17 Febbraio 2025

Prof. Margherita Lanz

Parliamo di

Le crisi che hanno dato origine all'open science

Unione europea e non solo

Le diverse Open

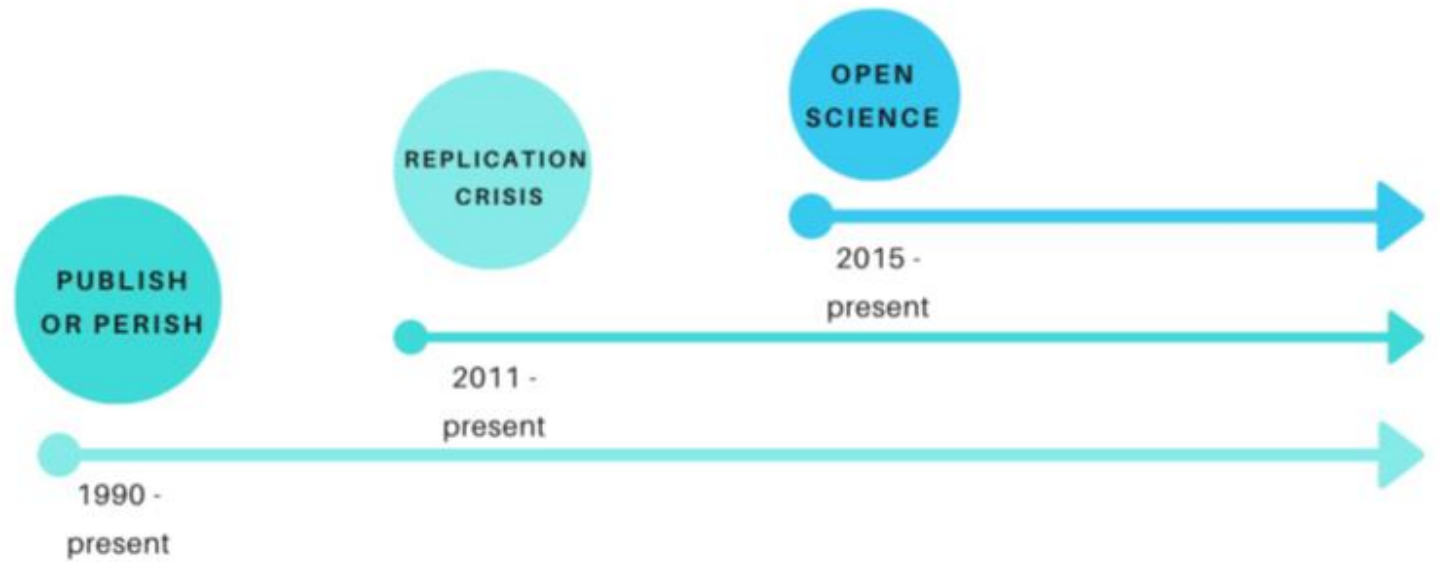
I dati FAIR

Le scelte etiche per la ricerca

Le scelte etiche per il partecipante

Open science

SCIENTIFIC TRANSITIONS



Publish or
perish



Crisi della replicabilità

- 2005

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; when there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p -value less than 0.05. Research is not most appropriately represented and summarized by p -values, but, unfortunately, there is a widespread notion that medical research articles

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α . Assuming that ϵ relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2

It can be proven that most claimed research findings are false.

should be interpreted based only on p -values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful.

[nature](#) > [news feature](#) > [article](#)

News Feature | Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

[Monya Baker](#)

[Nature](#) 533, 452–454 (2016) | [Cite this article](#)

224k Accesses | 5167 Altmetric | [Metrics](#)

i This article has been [updated](#)

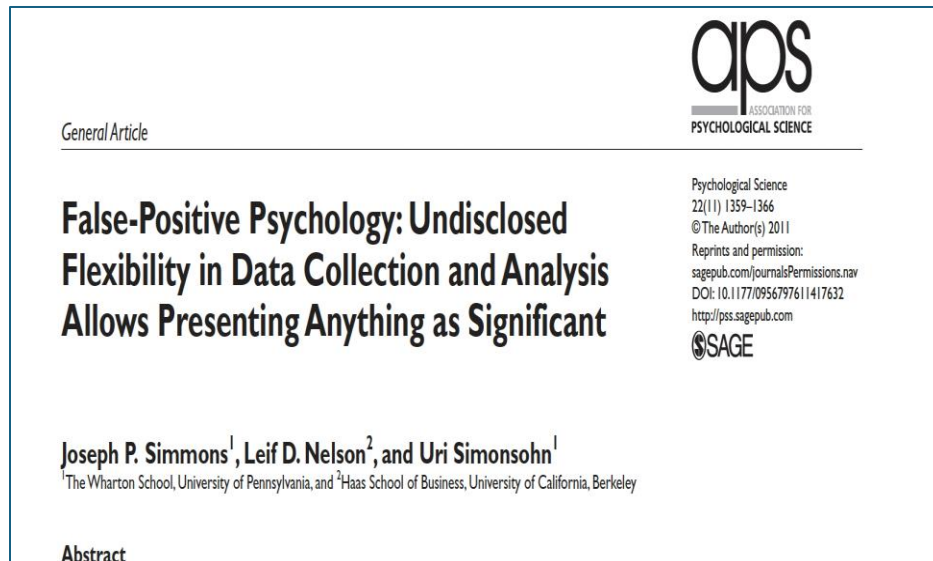
Survey sheds light on the ‘crisis’ rocking research.

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 questionnaire on reproducibility in research.

G Get citation

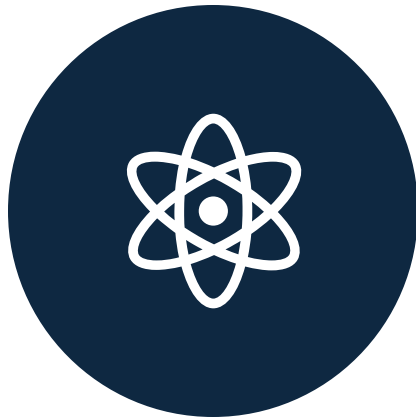
Crisi della replicabilità

2011



- **Gradi di libertà dei ricercatori** = scelte che compiono i ricercatori riguardo alla raccolta dei dati e alle analisi dei dati, anche non consigliate nei manuali!
- **Pratiche di ricerca discutibili (Questionable Research Practice)** = riguardano l'esclusione o la manipolazione di dati che non sono a favore dell'ipotesi di ricerca

CRITICITA' RISPETTO A ..



FIDUCIA NELLA
SCIENZA



BONTA' DEI RISULTATI
DELLA SCIENZA



FINANZIAMENTI



Open science

- «Open Science è l'idea che la conoscenza scientifica di ogni genere venga condivisa appena possibile» (Nielsen 2011).

L'Open Science è una pratica volta ad aumentare l'accessibilità e la trasparenza della ricerca scientifica. Il concetto di Scienza Aperta si basa su principi condivisi come inclusione, equità, giustizia e condivisione (Zee & Reich, 2018).

L'Open science comprende principi e comportamenti che promuovono una scienza trasparente, credibile, riproducibile e accessibile.

Open science

Riguarda come viene prodotta la scienza, come viene fatta la ricerca

- **Collaborazione**
- **Condivisione** della conoscenza
- **Condivisione** del processo di ricerca
- **Condivisione** dei dati



UNIONE EUROPEA

[Services](#)[Support](#)[Open Science In Europe](#)[About Us](#)[Sign In](#)

institutions or national clusters across ten European countries, including many core members of the continuing OpenAIRE consortium.



It enabled a production-quality digital repository infrastructure and led to the establishment of [COAR \(the Confederation of Open Access Repositories\)](#).

OpenAIRE (2009-2012), (2012-2014), (2015-2018), (2018-2020)

OpenAIRE has undergone **four** major phases via distinct **EC funded Projects**.



The **first phase**, the so-called “OPENAIRE Project” started in **1st December 2009** with a duration of **36 months**. The Project aim was to **support the implementation of Open Access in Europe**. It provided the means to promote and realize the widespread adoption of the Open Access Policy, as set out by the **ERC Scientific Council** Guidelines for Open Access and the Open Access pilot launched by the **European Commission** in FP7.

During the **second phase**, the “OPENAIREPLUS Project” which started in **1st December 2011** and lasted **30 months**, OpenAIRE extended the mission further to **facilitate access to the entire Open Access scientific production of the European Research Area, providing cross-links from publications to data and funding schemes**. This large-scale project brought together 41 pan-European partners, including three cross-disciplinary research communities. Find detailed results from [project executive report](#).



The **third phase** continued via the “OPENAIRE2020 Project” which started in **1st January 2015** and lasted for **42 months**. OpenAIRE2020 supported the H2020 full OA to publications implement and monitor these activities by promoting open scholarship and substantially improving the discoverability and reusability of research publications and data. The initiative brought together professionals from 50 institutions all over Europe, including research libraries, open scholarship organisations, national e-Infrastructure and data experts, as well as, IT and legal researchers. The network of National Open Access Desks (NOADs) has been extended expanded once again (to 33 countries), in order to collect H2020 project outputs, and support research data management. Backing this vast outreach, the OpenAIRE platform continued its extension.





PIANO NAZIONALE PER LA SCIENZA APERTA

PERCHÉ LA SCIENZA APERTA

La “scienza aperta” è un approccio al processo scientifico basato su collaborazione, condivisione aperta e tempestiva dei risultati, modalità di diffusione della conoscenza basate su tecnologie digitali in rete e metodi trasparenti di validazione e valutazione dei prodotti della ricerca.

La scienza aperta accresce l'efficacia della collaborazione e la riproducibilità dei risultati della ricerca. Essa aumenta il potenziale collaborativo con la possibilità di accesso ai dati e loro riuso per nuove analisi, anche di tipo interdisciplinare, e per l'insegnamento scientifico, nonché la fruibilità del sapere scientifico, in modo trasparente, a beneficio della società.

Per “accesso aperto” all'informazione scientifica si intende la possibilità di reperire in rete le pubblicazioni scientifiche, i dati e ogni altro risultato della ricerca e dell'insegnamento scientifico, assieme ai metadati che li rendono fruibili, senza costi per l'utente e senza barriere giuridiche e tecniche.


I principi della scienza aperta sono:

- la conoscenza come bene comune;
- la collaborazione e la solidarietà tra scienziati, l'equità e l'inclusione;
- la trasparenza del processo e dei contributi usati per la produzione e la validazione dei risultati scientifici;
- la messa a disposizione gratuita e con diritti di riuso, in rete, dei risultati della ricerca e dell'insegnamento per la società, l'innovazione e la citizen science;
- il rigore scientifico, la riproducibilità dei risultati sperimentali, la discussione critica dei dati, delle informazioni e della conoscenza resi accessibili in rete.

La scienza aperta favorisce un maggior ritorno degli investimenti nella ricerca pubblica con ampie ricadute sull'intera società, attraverso:

- l'accelerazione dei processi conoscitivi e di apprendimento. L'accesso gratuito e tempestivo ai risultati della ricerca, i diritti e gli strumenti per il riuso dei risultati della ricerca e della didattica (pubblicazioni, dati, software, licenze) e il sostegno alla collaborazione sono elementi fondamentali per progredire più rapidamente;
- le nuove possibilità per la ricerca interdisciplinare e multidisciplinare, necessaria per affrontare le sfide globali e le “Missioni” del programma quadro Horizon Europe (HEU)²;
- la verificabilità da parte della comunità scientifica dei nuovi risultati condivisi e della loro riproducibilità, in modo aperto e trasparente;
- l'integrità e qualità della comunicazione scientifica, basata su risultati accessibili, per una società più partecipe;
- il riuso di pubblicazioni, dati, software e protocolli disponibili;
- l'innovazione basata su tutte le acquisizioni, anche le più recenti, della scienza;
- lo sviluppo equo delle potenzialità di tutti i ricercatori, creando uguali opportunità di accesso a pubblicazioni, dati e altri risultati, indipendentemente da nazionalità o appartenenza istituzionale;
- l'allargamento dell'accesso alle risorse didattiche e formative.

Open Scholarship



Open Educational
Resources

Open Science

Open Access

Equity, Diversity,
Inclusion

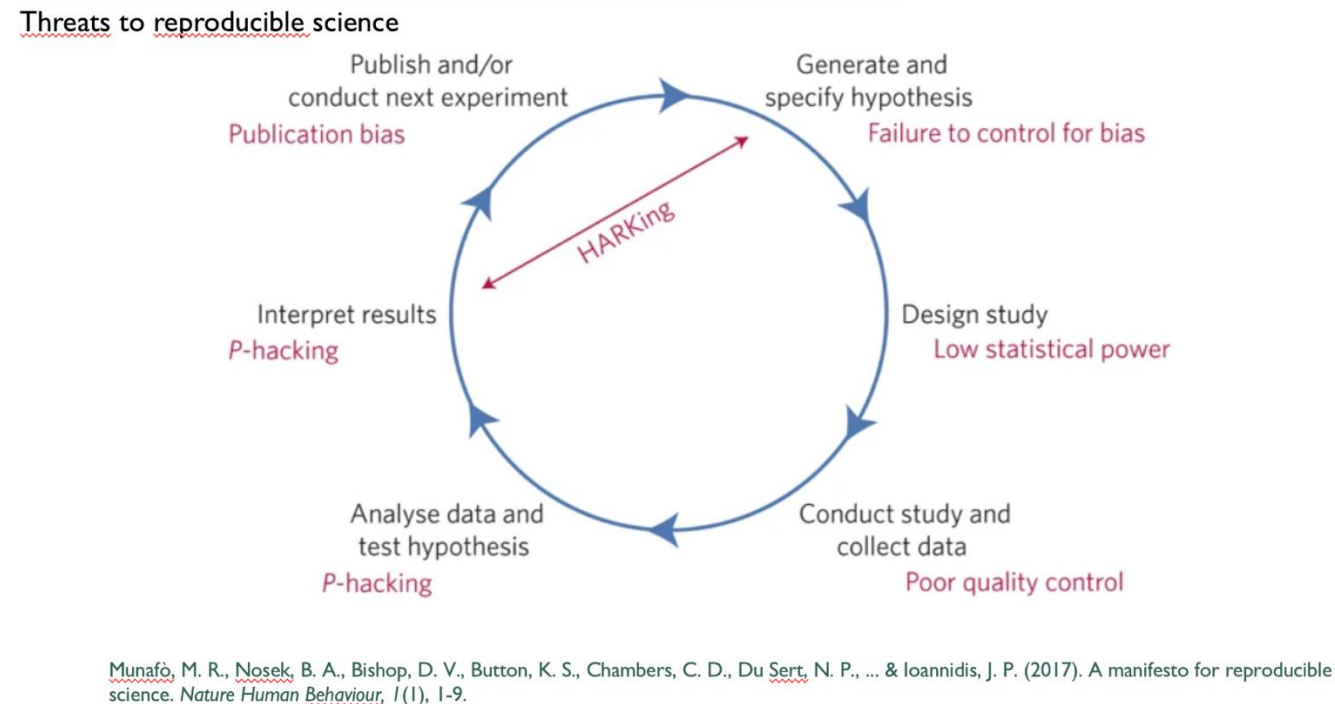
Open Data

Community Science

Open Source

LE QUESTIONI ETICHE NELLA PRATICHE DI RICERCA

- Le pratiche di ricerca discutibili o ingannevoli



CHE COSA SONO DISPOSTO A FARE PER ...

- 1. pubblicare
- 2. confermare le mie ipotesi
- 3. confermare l'ipotesi del committente



Come ottengo i dati della mia ricerca

Come analizzo i dati della mia ricerca



L'etica del ricercatore

- Competenza nella ricerca:
 - Teoria, metodi di ricerca, tecniche di analisi, interpretazioni dei risultati e applicazioni
- Responsabilità delle scelte metodologiche

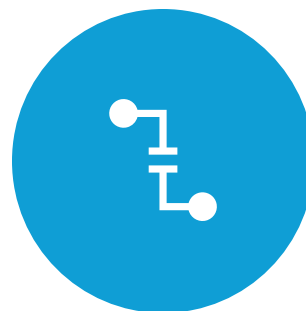
I DATI FAIR



FINDABLE (RINTRACIABILI) DALLA
COMUNITÀ ACCADEMICA E DAL
PUBBLICO GENERALE.



ACCESSIBLE (ACCESSIBILI) TRAMITE
IDENTIFICATORI PERSISTENTI,
METADATI E LINGUAGGIO CHIARO
COMPRESIBILE A PERSONE E
MACCHINE.



INTEROPERABLE (INTEROPERABILI)
TRAMITE L'APPLICAZIONE DI
STANDARD E PROTOCOLLI CHE
CONSENTANO LO SCAMBIO DI DATI
SENZA PERDITA DI INFORMAZIONI.



REUSABLE (RIUTILIZZABILI)
MASSIMIZZANDO IL LORO
POTENZIALE NELLA RICERCA.

Strumenti per il ricercatore



PRE-REGISTRAZIONE E
REGISTERED REPORT



CONDIVISIONE DEI DATI

DATI FAIR = DATI OPEN



DATI CHE VENGONO CONDIVISI **CON** LA
COMUNITÀ SCIENTIFICA



DATI CHE VENGONO **RESTITUITI** ALLA
COMUNITÀ SCIENTIFICA E ALLA
CITTADINANZA

Condivisione dei dati perche':

1. La rivista li richiede

2. attivare una rete di collaborazione per rispondere a domande complesse o indagare target «rari»

3. Aumento della trasparenza e della riproducibilità

4. Miglioramento della qualità e della validità scientifica (ridurre le pratiche di ricerca discutibili..)

5. Maggiore inclusione e collaborazione fra i ricercatori evitando ricerca MASKing" (Making Assumptions based on Skewed Knowledge)

Cosa intendiamo per dato della ricerca?



Ogni materiale che è stato raccolto, osservato, creato, elaborato, analizzato per validare o riprodurre i risultati della ricerca, con le informazioni e la documentazione di corredo necessaria per capire e gestire il dato stesso

[Questa foto](#) di Autore sconosciuto è concesso con licenza [CC BY-SA](#).

Data sharing

E' UNA SCELTA PROGETTUALE

Quali implicazioni etiche



Data sharing

- Quando progetto la mia ricerca:
 - 1. devo condividere tutte le operazioni che ho fatto per arrivare ai dati



Findable (rintracciabili) dalla comunità accademica e dal pubblico generale.



Accessible (accessibili) tramite identificatori persistenti, metadati e linguaggio chiaro comprensibile a persone e macchine.



Interoperable (interoperabili) tramite l'applicazione di standard e protocolli che consentano lo scambio di dati senza perdita di informazioni.



Reusable (riutilizzabili) massimizzando il loro potenziale nella ricerca.



Comitato etico

Data sharing:

- Quando progetto la mia ricerca:
 - 1. devo condividere tutte le operazioni che ho fatto per arrivare ai dati
 - 2. la scelta di condivisione dei dati **va indicato nel consenso** informato rendendo consapevoli le persone che i dati verranno condivisi

Il Data sharing consapevole

AS OPEN AS CAN POSSIBILE AS CLOSE AS
NECESSARY

Dipende dal tipo di dato

Dipende dall'infrastruttura che possiamo
avere

Dipende dal perché vogliamo mettere a
disposizione i dati e il livello di sintesi del
dato che vogliamo avere..

Sfide etiche nella condivisione dei dati

1. competenze di chi
utilizza i miei dati

2. uso diverso dei
miei dati rispetto a
quanto chiesto nel
consenso

3. Tutela dei
partecipanti

As open as can be as close as necessary

Dati open ma con licenze di utilizzo
: restringo utilizzo dei miei dati

Dove li metto

Anonimizzo e pseudoanonimizzo i
miei dati

Dati aggregati e di secondo ordine

Quale dato devo/posso condividere?

Il dato **anonimizzato** è un dato privo di elementi identificativi, reso irreversibilmente non riconducibile a una persona tramite un trattamento di anonimizzazione.

Il dato **pseudonimizzato** è un dato personale a cui è stata applicata una tecnica che impedisce l'attribuzione diretta a una persona senza l'uso di informazioni aggiuntive, che sono conservate separatamente e protette



Quale dati posso condividere?

K-anonymity: Garantisce che ogni combinazione di attributi sia condivisa da almeno k individui, riducendo la possibilità di identificare singoli partecipanti (ad es. donna, 1980, Milano, tumore al seno nel 2020... K volte)

Rimozione selettiva dei dati: Eliminare variabili altamente identificabili come luoghi specifici o date precise di nascita.

Responsible Data Sharing: Identifying and Remediating Possible Re-Identification of Human Participants

Kirsten N. Morehouse¹, Benedek Kurdi², and Brian A. Nosek^{3, 4}

¹ Department of Psychology, Harvard University

² Department of Psychology, University of Illinois Urbana-Champaign

³ Department of Psychology, University of Virginia

⁴ Center for Open Science, Charlottesville, Virginia, United States

Open data collected from research participants creates a tension between scholarly values of transparency and sharing, on the one hand, and privacy and security, on the other hand. A common solution is to make data sets anonymous by removing personally identifying information (e.g., names or worker IDs) before sharing. However, ostensibly anonymized data sets may be at risk of *re-identification* if they include demographic information. In the present article, we provide researchers with broadly applicable guidance and tangible tools so that they can engage in open science practices without jeopardizing participants' privacy. Specifically, we (a) review current privacy standards, (b) describe computer science data protection frameworks and their adaptability to the social sciences, (c) provide practical guidance for assessing and addressing re-identification risk, (d) introduce two open-source algorithms developed for psychological scientists—MinBlur and MinBlurLite—to increase privacy while maintaining the integrity of open data, and (e) highlight aspects of ethical data sharing that require further attention. Ultimately, the risk of re-identification should not dissuade engagement with open science practices. Instead, technical innovations should be developed and harnessed so that science can be as open as possible to promote transparency and sharing and as closed as necessary to maintain privacy and security.

Rischio di ri-identificazione dei partecipanti

MinBlur and MinBlurLite: Open-Source Algorithms to Identify and Remedy Risks to Privacy

Computer scientists have developed several techniques to increase the protection of data sets against the risk of re-identification. However, many of these techniques rely on opaque statistical methods that are difficult to interpret (e.g., recurrent neural networks; Dernoncourt et al., 2017; Liu et al., 2017) or on algorithms that apply arbitrary partitioning processes to quasi-identifiers (e.g., optimal *k*-anonymization; Bayardo & Agrawal, 2005). For example, if a data set has many participants under 40, the algorithm might produce a new age variable with the following levels, which are both highly unequal in their span and difficult to interpret substantively: <18–19>, <20–21>, <22–25>, <26–29>, <30–40>, and <41–79>.

Indeed, these techniques are often ill-suited for use by social scientists who seek to preserve the usefulness of demographic information in open data. The minimal blurring algorithms, MinBlur and MinBlurLite, are introduced here as innovations in algorithms that apply conceptual frameworks

Open science ...

- Ricerca rigorosa e trasparente attraverso pratiche di ricerca corrette
- OPEN strumento per garantire la replicabilità della ricerca



Per il comitato etico

- Sollecitare Ateneo per una repository affidabile per la conservazione dei dati che tuteli il ricercatore e i partecipanti
- Formazione per i ricercatori sulle diverse possibilità di «chiusura» dei dati (licenze)
- Formazione per i ricercatori sulle diverse possibilità di anonimizzazione
- Sostenere la condivisione dei dati anche attraverso il modulo
- Open science e etica



La ricerca scientifica ... come scelta responsabile



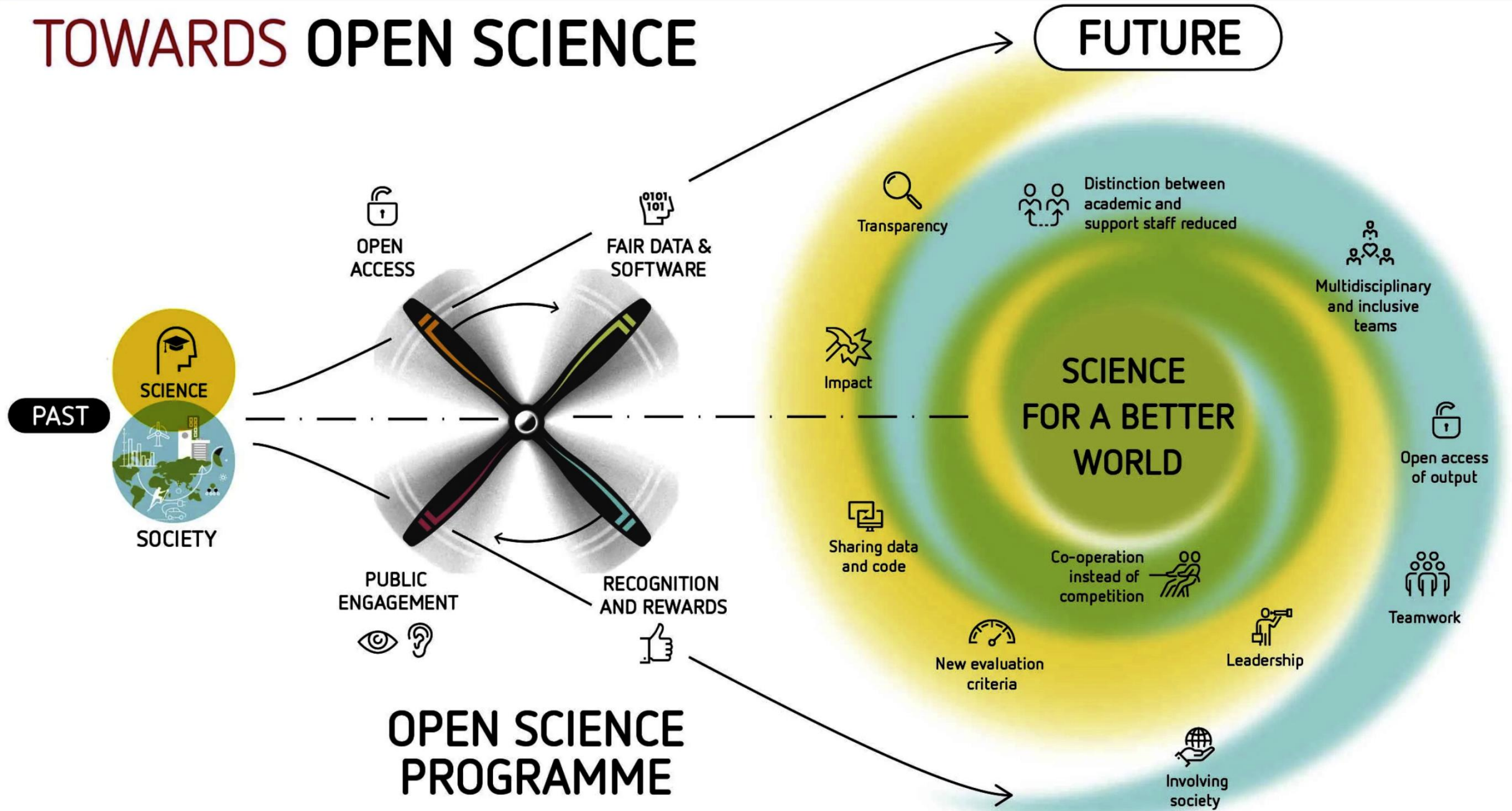
Ricercatore

The diagram consists of three blue rectangular boxes. One box labeled 'Ricercatore' is on the left. Two other boxes, labeled 'Partecipanti' and 'Cittadinanza', are stacked vertically on the right. There are no lines or arrows connecting the boxes.

Partecipanti

Cittadinanza

TOWARDS OPEN SCIENCE



Qualche riferimento..

- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131–144. <https://doi.org/10.1177/2515245917747656>
- Moussaoui, J. R., McCarley, J. S., Soicher, R. N., & McCarley, J. S. (2024). Ethical Considerations of Open Science. *Beyond the Code: Integrating Ethics into the Undergraduate Psychology Curriculum*; Pantesco, B., Beising, R., Thakkar, V., Naufel, K., Woolf, L., Eds.
- Morehouse, C. N., Kurdi, B., & Nosek, B. A. (2024). Responsible data sharing: Identifying and remedying possible re-identification of human participants. *American Psychologist*.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.
- Tourish, D., & Craig, R. (2020). Research misconduct in business and management studies: Causes, consequences, and possible remedies. *Journal of Management Inquiry*, 29(2), 174-187.