Est. 1841 | YORK ST JOHN UNIVERSITY

Flack, Zoe M., Parsons, Sam, Elsherif, Mahmoud, Guay, Samuel, Shahim, Owen, Govaart, Gisela, Norris, Emma, O'Mahoney, Aoife, Parker, Adam, Todorovic, Ana, Pennington, Charlotte, Garcia-Pelegrin, Elias, Lazic, Aleksandra, Robertson, Olly, Middleton, Sara, Valentini, Beatrice, McCuaig, Joanne, Barker, Bradley, Collins, Elizabeth and Jane Fox, Kerry (2022) A Community-Sourced Glossary of Open Scholarship Terms. Nature Human Behaviour, 6. pp. 312-318.

Downloaded from: https://ray.yorksj.ac.uk/id/eprint/15513/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: https://www.nature.com/articles/s41562-021-01269-4

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repositories Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

Title: A Community-Sourced Glossary of Open Scholarship Terms

Authors:

Sam Parsons‡[1], Flávio Azevedo‡[2], Mahmoud M. Elsherif‡[3,4], Samuel Guay[5], Owen N. Shahim[6], Gisela H. Govaart[7,8], Emma Norris[9], Aoife O'Mahony[10], Adam J. Parker[1,11], Ana Todorovic[1], Charlotte R. Pennington[12], Elias Garcia-Pelegrin[13], Aleksandra Lazić[14], Olly M. Robertson[15,16], Sara L. Middleton[17,18], Beatrice Valentini[19], Joanne McCuaig[20,21], Bradley J. Baker[22,23], Elizabeth Collins[24], Adrien A. Fillon[25], Tina B. Lonsdorf[26], Michele C. Lim[1], Norbert Vanek[27,28], Marton Kovacs[29,30], Timo B. Roettger[31], Sonia Rishi[32], Jacob F. Miranda[33], Matt Jaquiery[1], Suzanne L. K. Stewart[34], Valeria Agostini[32], Andrew J. Stewart[35], Kamil Izydorczak[36], Sarah Ashcroft-Jones[1], Helena Hartmann[37,38], Madeleine Ingham[32], Yuki Yamada[39], Martin R. Vasilev[40], Filip Dechterenko[41], Nihan Albayrak-Aydemir[42], Yu-Fang Yang[43], LaPlume A. Annalise[44], Julia K. Wolska[45], Emma L. Henderson[46], Mirela Zaneva[1], Benjamin G. Farrar[13], Ross Mounce[47], Tamara Kalandadze[48], Wanyin Li[32], Qinyu Xiao[49], Robert M. Ross[50], Siu Kit Yeung[49], Meng Liu[51], Micah L. Vandegrift[52], Zoltan Kekecs[29,53], Marta K. Topor[54], Myriam A. Baum[55], Emily A. Williams[56], Asma A. Assaneea[57,21], Amélie Bret[58], Aidan G. Cashin[59,60], Nick Ballou[61], Tsvetomira Dumbalska[1], Bettina M. J. Kern[62], Claire R. Melia[63,64], Beatrix Arendt[65], Gerald H. Vineyard[66], Jade S. Pickering[67,68], Thomas R. Evans[69], Catherine Laverty[32], Eliza A. Woodward[70], David Moreau[71], Dominique G. Roche[72], Eike M. Rinke[73,74], Graham Reid[1], Eduardo Garcia-Garzon[75], Steven Verheyen[76], Halil E. Kocalar[77], Ashley R. Blake[21], Jamie P. Cockcroft[78], Leticia Micheli[79], Brice Beffara Bret[58], Zoe M. Flack[80], Barnabas Szaszi[81], Markus Weinmann[82,83], Oscar Lecuona[84,85], Birgit Schmidt[86], William X. Ngiam[87], Ana Barbosa Mendes[88], Shannon Francis[3], Brett J. Gall[89], Mariella Paul[90], Connor T. Keating[32], Magdalena Grose-Hodge[21], James E. Bartlett[91], Bethan J. Iley[92,93], Lisa Spitzer[94], Madeleine Pownall[56], Christopher J. Graham[95,96], Tobias Wingen[97], Jenny Terry[98], Catia M. F. d. Oliveira[78], Ryan A. Millager[99], Kerry Fox[80], Alaa AlDoh[98], Alexander Hart[55], Olmo R. van den Akker[100], Gilad Feldman[49], Dominik A. Kiersz[101], Christina Pomareda[32], Kai Krautter[55], Ali H. Al-Hoorie[102], Balazs Aczel[29]

Affiliations:

[1]Department of Experimental Psychology, University of Oxford, [2]Department of Communication, Friedrich Schiller University, [3]Department of Psychology, University of Birmingham, [4]Blavatnik School of Government, University of Oxford, [5]Department of Psychology, University of Montreal, [6]East Midlands School of Anaesthesia, [7]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, [8]Charité – Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin, [9]Department of Health Sciences, Brunel University, [10]School of Psychology, Cardiff University, [11]Department of Experimental Psychology, Division of Psychology and Language Sciences, University College

London, [12]School of Psychology, Aston University, [13]Department of Psychology, University of Cambridge, [14]Department of Psychology, University of Belgrade, [15]Departments of Psychiatry and Experimental Psychology, University of Oxford, [16]School of Psychology, Keele University, [17]Department of Zoology, University of Oxford, [18]Plant Sciences Department, University of Oxford, [19]Faculty of Psychology and Educational Sciences, University of Geneva, [20]Department of Foreign Languages, Hongik University, [21]Department of English Language and Applied Linguistics, University of Birmingham, [22]Department of Sport and Recreation Management, Temple University, [23]Mark H. McCormack Department of Sport Management, University of Massachusetts, [24]Department of Psychology, University of Stirling, [25]Department of Psychology, University of Aix-Marseille, [26]Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, [27]School of Cultures, Languages and Linguistics, University of Auckland, [28]Experimental Research on Central European Languages Lab, Charles University Prague, [29]Institute of Psychology, ELTE, Eotvos Lorand University, Budapest, Hungary, [30]Doctoral School of Psychology, ELTE Eotvos Lorand University, [31]Department of Linguistics and Scandinavian Studies, Universitetet i Oslo, [32]School of Psychology, University of Birmingham, [33]Department of Psychology, University of Alabama, [34]School of Psychology, University of Chester, [35]Division of Neuroscience and Experimental Psychology, University of Manchester, [36]Faculty of Psychology, SWPS University of Social Sciences and Humanities, [37]Department for Cognition, Emotion, and Methods in Psychology, University of Vienna, [38]Netherlands Institute for Neuroscience, [39]Faculty of Arts and Science, Kyushu University, [40]Department of Psychology, Bournemouth University, [41]Department of Cognitive Psychology, Institute of Psychology, Czech Academy of Sciences, [42]London School of Economics and Political Science, [43]Department of Psychology, University of Würzburg, [44]Rotman Research Institute, Baycrest Hospital (fully affiliated with the University of Toronto), [45]Department of Psychology, Manchester Metropolitan University, [46]School of Psychology, University of Surrey, [47]Arcadia Fund, [48]Department of Education, ICT and Learning, Østfold University College, [49]Department of Psychology, University of Hong Kong, [50]Department of Psychology, Macquarie University, [51]Faculty of Education, University of Cambridge, [52]Open Knowledge Librarian, NC State University, [53]Department of Psychology, Lund University, [54]Department of Psychology, University of Surrey, [55]Research Group Applied Statistical Modeling, Saarland University, [56]School of Psychology, University of Leeds, [57]Taibah University, [58]Department of Psychology, University of Nantes, [59]Centre for Pain IMPACT, Neuroscience Research, [60]Prince of Wales Clinical School, University of New South Wales, [61]Game AI Group, Queen Mary University of London, [62]Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, [63]Department of Mental health and Social Work, Middlesex University, [64]Department of Psychology, Keele University, [65]Center for Open Science, [66]Department of Psychology, University of East London, [67]School of Psychology, University of Southampton, [68]Division of Neuroscience & Experimental Psychology, University of Manchester, [69]School of Human Sciences, University of Greenwich, [70]Department of Psychology, Oxford Brooke University, [71]School of Psychology, University of Auckland, [72]Department of Biology, Carleton University,

[73]School of Politics and International Studies, University of Leeds, [74]Mannheim Centre for European Social Research (MZES), University of Mannheim, [75]Faculty of Health, Universidad Camilo José Cela, [76]Department of Psychology, Education and Child Studies; Erasmus University Rotterdam, [77]Department of Educational Science, Muğla Sıtkı Koçman University, [78]Department of Psychology, University of York, [79]Institute of Psychology, Leibniz University Hannover, [80]School of Humanities and Social Science, University of Brighton, [81]Institute of Psychology, ELTE, Eotvos Lorand University, [82]Faculty of Management, Economics and Social Sciences, University of Cologne, [83]Department of Technology and Operations Management, Erasmus University Rotterdam, [84]Faculty of Psychology, Universidad Rey Juan Carlos, Madrid, Spain, [85]Faculty of Psychology, Universidad Autónoma de Madrid, [86]University of Göttingen, [87]Department of Psychology, University of Chicago, [88]Erasmus School of Philosophy. Erasmus University Rotterdam, [89]Department of Political Sciences, Duke University, [90]Psychology of Language Department, University of Göttingen, [91]School of Psychology and Social Science, Arden University, [92]School of Psychology, Queen's University Belfast, [93]School of Psychology, University of Kent, [94]Leibniz Institute for Psychology (ZPID), [95]Division of Population Health, Health Services Research and Primary Care, University of Manchester, [96]Scottish Health Action on Alcohol Problems (SHAAP), [97]Department of Psychology, University of Cologne, [98]School of Psychology, University of Sussex, [99]Department of Hearing and Speech Sciences, Vanderbilt University, [100]Department of Methodology, Tilburg University, [101]Independent, [102]Jubail English Language and Preparatory Year Institute, Royal Commission for Jubail and Yanbu, Saudi Arabia

Correspondence: Flávio Azevedo (flavio.azevedo@uni-jena.de). ‡Shared co-first authorship.

Standfirst:

Open scholarship has transformed research, introducing a host of new terms in the lexicon of researchers. The Framework of Open and Reproducible Research Teaching (FORRT) community presents a crowdsourced glossary of open scholarship terms to facilitate education and effective communication between experts and newcomers.

Barriers to Open Scholarship Terminology

Open Scholarship is an umbrella term referring to the endeavour to improve openness, integrity, social justice, diversity, equity, inclusivity and accessibility in all areas of scholarly activities. Open Scholarship extends the more widely used terms "Open Science" and "Open Research" to include academic fields beyond the sciences and academic activities.

Over the last decade, Open Scholarship has radically changed the way we think and discuss research and higher education. New concepts, tools, and practices have been developed and promoted, introducing novel terms or repurposing existing ones. These changes have increased the breadth but also the ambiguity of terminology, creating barriers to effective understanding and communication for novices and experts. Presently, certain terms such as *replicability* or *reproducibility* are well known but frequently used interchangeably or differentially among fields and disciplines; other terms are less known beyond a small circle of researchers, such as *CARKing*, *PARKing*, or *paradata*. Terms that become conventional within a given field often reflect the preferences of those with the platforms and privileges to determine academic discourse and, consequently, can act as a barrier to participation by those without such platforms. A similar barrier is that much academic language is contained within a 'hidden curricula', meaning these terms and practices are often used under the misplaced assumption that students, or those unfamiliar with an area, understand them.

A Diversity of New Terminologies

Terms associated with Open Scholarship are diverse in many aspects. Some are neologisms (i.e., newly coined) while others are reclamations of older terms (e.g. *p*-hacking and adversarial collaboration). Furthermore, frequent use of acronyms can hinder immediate understanding. For example, ORCID iD[1] refers to a persistent unique identifier for individuals in their role as creator or contributor. It enables linking digital documents and other contributions to their digital records, attributes credit, and resolves name ambiguities. Other terms use metaphors, such as the *Garden of Forking Paths*[2], which highlights the many alternative paths researchers can embark on when analysing data. These examples highlight the complexity of Open Scholarship terminology, which often assumes prior knowledge, making it difficult for individuals not versed in these terms to engage in ongoing conversations about these topics, thus excluding them from joining the discourse.

Communication across disciplines, and across dramatically varying levels of subject and technical expertise, can be extremely difficult. Challenges can arise in understanding scientific

texts when words with a historical meaning gain a new one in certain contexts. The term *paper mills*[3], for example, typically refers to factories devoted to manufacturing papers but also, in the context of open scholarship, it denotes unethical for-profit organisations that create and publish on-demand fraudulent scientific papers based on techniques such as fabrication of data and plagiarism.

A similar challenge arises when the meaning and usage of certain terms differs between (sub)disciplines and methodologies. For example, in social science fields the term *preregistration* refers to an uneditable, timestamped version of a research protocol, whereas in healthcare fields it refers to an accelerated course that qualifies students to fast-track into a medical profession. As another example, social scientists understand the term *external validity* to mean that the findings can be generalised to other contexts (different measures, people, places, and times), while psychometricians regard it as the relationship of a psychological concept with theoretically relevant extrinsic variables. *Creative destruction*[4] is yet another example: in economics, it refers to revolutionising the economic structure from within by destroying the old system and replacing it with a new one. In psychology, a creative destruction approach to replication means that replications can —in particular circumstances— be leveraged to not only support or question the original findings, but also to replace weaker theories with stronger ones that have greater explanatory power (by preregistering different theoretical predictions and adding new measures, conditions and populations that facilitate competitive theory testing). As interdisciplinary collaboration is growing and often required by many funding agencies and stakeholders, this creates a potential for miscommunication and confusion when using such terminology.

The clarification of scholarly terminology is a challenging endeavour. It should be built on the insights of a community of experts with different perspectives and requires consensus among the members across disciplines. As the breadth of these initiatives can be overwhelming, digestible introductions to the language of Open Scholarship are needed[5–7]. In order to reduce barriers to entry and understanding of Open Scholarship terminology, as well as to foster the accessibility, inclusivity, and clarity of its language, a community-driven glossary using a consensus-based methodology could help clarify terminologies and aid in the mentoring and teaching of these concepts.

The Open Scholarship Glossary Project

The present glossary was developed by the FORRT[7] community; an educational initiative aiming to integrate open and reproducible research principles into higher education as well as supporting educators and mentors to address related pedagogical challenges. The work has been completed in three rounds. First, the lead team created an initial list of Open Scholarship related terms and a structure for each term. Each term was required to have (1) a concise *definition*; with supporting (2) *references;* (3) *related terms*; and (4) any applicable *alternative definitions*. The present glossary has been developed using a crowdsourced methodology with the involvement of over 100 contributors at various career stages and from a diverse range of disciplines, including

psychology, economics, neuroscience, information science, social science, biology, ecology, public health, and linguistics. In the second round, and in a dynamic and iterative crowdsourced process, members of the research community were invited through social media platforms or via organisations such as *ReproducibiliTea* to participate in the project. The community contributors suggested new terms, to which they provided main and alternative definitions, as well as reviewed and edited other terms iteratively throughout the project. We recorded these contributions, and this is reflected in our CRediT statement; all contributors were invited as authors on this manuscript. We considered definitions as ready for dissemination when they had been reviewed by a sufficient number of contributors (typically five or more) and reached consensus. Through this process, the community-driven glossary development procedure deliberately centred the Open Scholarship ethos of accessibility, diversity, equity, and inclusion.

The project resulted in the drafting of more than 250 terms. In Table 1, we present an abbreviated set of 30 terms that represent the plurality of terms for the broader Open Scholarship concepts (see https://forrt.org/glossary for the complete glossary, including key references and links to related terms, as well as a more detailed explanation of the project's mission and goals). The FORRT Glossary project is licensed under a CC BY NC SA 4.0 license. The present glossary is the 1.0 version. The version-controlled source code of the new releases of the complete Glossary is archived on FORRT's website, GitHub, OSF, and Zenodo, wherein new releases will also be stored. We set up a system allowing for continuous improvement, extension, and updating from community feedback and involvement. Versioning will also allow the study of the evolution of the terminologies.

Table 1 Open Scholarship Glossary 1.0 Examples (abbreviated version)

Note: The complete glossary, including key references and links to related terms, can be found in Supplementary Information and is available at https://forrt.org/glossary. Minor modifications were made to comply with editorial requirements.

Analytic Flexibility	Analytic flexibility is a type of researcher degrees of freedom that refers specifically to the large number of choices made during data preprocessing and statistical analysis. Analytic flexibility can be problematic as this variability in analytic strategies can translate into variability in research outcomes, particularly when several strategies are applied, but not transparently reported.
#bropenscience	A tongue-in-cheek expression intended to raise awareness of the lack of diverse voices in open science, in addition to the presence of behavior and communication styles that can be toxic or exclusionary. Importantly, not all bros are men; rather, they are individuals who demonstrate rigid thinking, lack self-awareness, and tend towards hostility, unkindness, and exclusion. They generally belong to dominant groups who benefit from structural

	privileges. To address #bropenscience, researchers should examine and address structural inequalities within academic systems and institutions.
CARKing	Critiquing After the Results are Known (CARKing) refers to presenting a criticism of a design as one that you would have made in advance of the results being known. It usually forms a reaction or criticism to unwelcome or unfavourable results, whether the critic is conscious of this fact or not.
Codebook	A codebook is a high-level summary that describes the contents, structure, nature and layout of a data set. A well-documented codebook contains information intended to be complete and self-explanatory for each variable in a data file, such as the wording and coding of the item, and the underlying construct. It provides transparency to researchers who may be unfamiliar with the data but wish to reproduce analyses or reuse the data.
Conceptual replication	A replication attempt whereby the primary effect of interest is the same but tested in a different sample and captured in a different way to that originally reported (i.e., using different operationalisations, data processing and statistical approaches and/or different constructs). The purpose of a conceptual replication is often to explore what conditions limit the extent to which an effect can be observed and generalised (e.g., only within certain contexts, with certain samples, using certain measurement approaches) towards evaluating and advancing theory.
Creative destruction approach	Replication efforts should seek not just to support or question the original findings, but also to replace them with revised, stronger theories with greater explanatory power. This approach therefore involves 'pruning' existing theories, comparing all the alternative theories, and making replication efforts more generative and engaged in theory-building
Credibility Revolution	The problems and the solutions resulting from a growing distrust in scientific findings, following concerns about the credibility of scientific claims (e.g., low replicability). The term has been proposed as a more positive alternative to the term replicability crisis, and includes the many solutions to improve the credibility of research, such as preregistration, transparency, and replication.
CRedIT	The Contributor Roles Taxonomy (CRediT; https://casrai.org/credit/) is a high-level taxonomy used to indicate the roles typically adopted by contributors to scientific scholarly output. There are currently 14 roles that describe each contributor's specific contribution to the scholarly output.

	They can be assigned multiple times to different authors and one author can also be assigned multiple roles. CRediT includes the following roles: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.
Decolonisation	Coloniality can be described as the naturalisation of concepts such as imperialism, capitalism, and nationalism. Together these concepts can be thought of as a matrix of power (and power relations) that can be traced to the colonial period. Decoloniality seeks to break down and decentralize those power relations, with the aim to understand their persistence and to reconstruct the norms and values of a given domain. In an academic setting, decolonisation refers to the rethinking of the lens through which we teach, research, and co-exist, so that the lens generalises beyond Western-centred and colonial perspectives. Decolonising academia involves reconstructing the historical and cultural frameworks being used, redistributing a sense of belonging in universities, and empowering and including voices and knowledge types that have historically been excluded from academia. This is done when people engage with their past, present, and future whilst holding a perspective that is separate from the socially dominant perspective. Also, by including, not rejecting, an individuals' internalised norms and taboos from the specific colony.
Direct replication	As 'direct replication' does not have a widely-agreed technical meaning nor there is no clear cut distinction between a direct and conceptual replication, below we list several contributions towards a consensus. Rather than debating the 'exactness' of a replication, it is more helpful to discuss the relevant differences between a replication and its target, and their implications for the reliability and generality of the target's results. Generally, direct replication refers to a new data collection that attempts to replicate original studies' methods as closely as possible. In this sense, direct replication is a replication attempt that aims to duplicate the needed elements that produced the original results.. The purpose of a direct replication can be to identify type 1 errors and/or experimenter effects, determine the replicability of an effect using the same or improved practices, or to create more specific estimates of effect size. Directness of replication is a continuum between repeating specific observations (data) and observing generalised effects (phenomena). How closely a replication replicates an original study is often a matter for debate, often with

	differences being cited as hidden moderators of effects. Furthermore, there can be debate over the relevant importance of technical equivalence (i.e., using identical materials) versus psychological equivalence (i.e., realizing the identical psychological conditions) to the original study. For example, consider a study on Trust in the US- President conducted in 2018. A technical equivalent replication would use Trump as stimulus (he was president in 2018) a psychological equivalent study would use Biden (he is the current president).
External Validity	Whether the findings of a scientific study can be generalized to other contexts outside the study context (different measures, settings, people, places, and times). Statistically, threats to external validity may reflect interactions whereby the effect of one factor (the independent variable) depends on another factor (a confounding variable). External validity may also be limited by the study design (e.g., an artificial laboratory setting or a non-representative sample). Alternative definition: In Psychometrics, the degree of evidence that confirms the relations of a tested psychological construct with external variables.
FAIR principles	Describes making scholarly materials Findable, Accessible, Interoperable and Reusable (FAIR). 'Findable' and 'Accessible' are concerned with where materials are stored (e.g. in data repositories), while 'Interoperable' and 'Reusable' focus on the importance of data formats and how such formats might change in the future.
Garden of forking paths	The typically-invisible decision tree traversed during operationalization and statistical analysis given that 'there is a one-to-many mapping from scientific to statistical hypotheses' (Gelman and Loken, 2013, p. 6). In other words, even in absence of p-hacking or fishing expeditions and when the research hypothesis was posited ahead of time, there can be a plethora of statistical results that can appear to be supported by theory given data. "The problem is there can be a large number of potential comparisons when the details of data analysis are highly contingent on data, without the researcher having to perform any conscious procedure of fishing or examining multiple p-values" (Gelman and Loken, 2013, p. 1). The term aims to highlight the uncertainty ensuing from idiosyncratic analytical and statistical choices in mapping theory-to-test, and contrasting intentional (and unethical) questionable research practices (e.g. p-hacking and fishing expeditions) versus non-intentional research practices that can, potentially, have the same effect despite not having intent to corrupt their results. The

	garden of forking paths refers to the decisions during the scientific process that inflate the false-positive rate as a consequence of the potential paths which could have been taken (had other decisions been made).
HARKING	A questionable research practice termed 'Hypothesizing After the Results are Known' (HARKing). HARKing has been defined as a post hoc hypothesis which is either based on —or informed by— a result in a research report as if it was, in fact, a priori . For example, performing subgroup analyses, finding an effect in one subgroup, and writing the introduction with a 'hypothesis' that matches these results.
Incentive Structure	The set of evaluation and reward mechanisms (explicit and implicit) for scientists and their work. Incentivised areas within the broader structure include hiring and promotion practices, track record for awarding funding, and prestige indicators such as publication in journals with high impact factors, invited presentations, editorships, and awards. It is commonly believed that these criteria are often misaligned with the telos of science, and therefore do not promote rigorous scientific output. Initiatives like DORA aim to reduce the field's dependency on evaluation criteria such as journal impact factors in favor of assessments based on the intrinsic quality of research outputs.
Inclusion	Inclusion, or inclusivity, refers to a sense of welcome and respect within a given collaborative project or environment (such as academia) where diversity simply indicates a wide range of backgrounds, perspectives, and experiences, efforts to increase inclusion go further to promote engagement and equal valuation among diverse individuals, who might otherwise be marginalized. Increasing inclusivity often involves minimising the impact of, or even removing, systemic barriers to accessibility and engagement.
Metadata	Structured data that describes and synthesises other data. Metadata can help find, organize, and understand data. Examples of metadata include creator, title, contributors, keywords, tags, as well as any kind of information necessary to verify and understand the results and conclusions of a study such as codebook on data labels, descriptions, the sample and data collection process. *Alternative definition*: data about data.
Multi-analyst studies	In typical empirical studies, a single researcher or research team conducts the analysis, which creates uncertainty about the extent to which the choice of analysis influences the results. In multi-analyst studies, two or more researchers independently analyse the same research question or hypothesis

	on the same dataset. A multi-analyst approach may be beneficial in increasing our confidence in a particular finding; uncovering the impact of analytical preferences across research teams; and highlighting the variability in such analytical approaches.
Open Scholarship	'Open scholarship' is often used synonymously with 'open science', but extends to all disciplines, drawing in those which might not traditionally identify as science-based. It reflects the idea that knowledge of all kinds should be openly shared, transparent, rigorous, reproducible, replicable, accumulative, and inclusive (allowing for all knowledge systems). Open scholarship includes all scholarly activities that are not solely limited to research such as teaching and pedagogy.
Open Science	An umbrella term reflecting the idea that scientific knowledge of all kinds, where appropriate, should be openly accessible, transparent, rigorous, reproducible, replicable, accumulative, and inclusive, all which are considered fundamental features of the scientific endeavour. Open science consists of principles and behaviors that promote transparent, credible, reproducible, and accessible science. Open science has six major aspects: open data, open methodology, open source, open access, open peer review, and open educational resources.
ORCID iD	An organisation that provides a registry of persistent unique identifiers (ORCID iDs) for researchers and scholars, allowing these users to link their digital research documents and other contributions to their ORCID record. This avoids the name ambiguity problem in scholarly communication. ORCID iDs provide unique, persistent identifiers connecting researchers and their scholarly work. It is free to register for an ORCID iD at https://orcid.org/register.
p-hacking	Exploiting techniques that may artificially increase the likelihood of obtaining a statistically significant result by meeting the standard statistical significance criterion (typically $\alpha = .05$). For example, performing multiple analyses and reporting only those at $p < .05$, selectively removing data until $p < .05$, selecting variables for use in analyses based on whether those parameters are statistically significant.
Papermill	An organization that is engaged in scientific misconduct wherein multiple papers are produced by falsifying or fabricating data, e.g. by editing figures or numerical data or plagiarizing written text. A papermill relates to the fast

	production and dissemination of multiple allegedly new papers. These are often not detected in the scientific publishing process and therefore either never found or retracted if discovered (e.g. through plagiarism software).
Paradata	Data that are captured about the characteristics and context of primary data collected from an individual - distinct from metadata. Paradata can be used to investigate a respondent's interaction with a survey or an experiment on a micro-level. They can be most easily collected during computer mediated surveys but are not limited to them. Examples include response times to survey questions, repeated patterns of responses such as choosing the same answer for all questions, contextual characteristics of the participant such as injuries that prevent good performance on tasks, the number of premature responses to stimuli in an experiment. Paradata have been used for the investigation and adjustment of measurement and sampling errors.
PARKing	PARKing (preregistering after results are known) is defined as the practice where researchers complete an experiment (possibly with infinite re-experimentation) before preregistering. This practice invalidates the purpose of preregistration, and is one of the QRPs (or, even scientific misconduct) that try to gain only "credibility that it has been preregistered.
Preregistration	The practice of publishing the plan for a study, including research questions/hypotheses, research design, data analysis before the data has been collected or examined. It is also possible to preregister secondary data analyses. A preregistration document is time-stamped and typically registered with an independent party (e.g., a repository) so that it can be publicly shared with others (possibly after an embargo period). Preregistration provides a transparent documentation of what was planned at a certain time point, and allows third parties to assess what changes may have occurred afterwards. The more detailed a preregistration is, the better third parties can assess these changes and with that the validity of the performed analyses. Preregistration aims to clearly distinguish confirmatory from exploratory research.
Replicability	An umbrella term, used differently across fields, covering concepts of: direct and conceptual replication, computational reproducibility/replicability, generalizability analysis and robustness analyses. Some of the definitions used previously include: a different team arriving at the same results using the original author's artifacts; a study arriving at the same conclusion after collecting new data; as well as studies for which any outcome would be considered diagnostic evidence about a

	claim from prior research.
Reproducibility	A minimum standard on a spectrum of activities ("reproducibility spectrum") for assessing the value or accuracy of scientific claims based on the original methods, data, and code. For instance, where the original researcher's data and computer codes are used to regenerate the results, often referred to as computational reproducibility. Reproducibility does not guarantee the quality, correctness, or validity of the published results. In some fields, this meaning is, instead, associated with the term "replicability" or 'repeatability'.
Registered Reports	A scientific publishing format that includes an initial round of peer review of the background and methods (study design, measurement, and analysis plan); sufficiently high quality manuscripts are accepted for in-principle acceptance (IPA) at this stage. Typically, this stage 1 review occurs before data collection, however secondary data analyses are possible in this publishing format. Following data analyses and write up of results and discussion sections, the stage 2 review assesses whether authors sufficiently followed their study plan and reported deviations from it (and remains indifferent to the results). This shifts the focus of the review to the study's proposed research question and methodology and away from the perceived interest in the study's results.
Under-representation	Not all voices, perspectives, and members of the community are adequately represented. Under-representation typically occurs when the voices or perspectives of one group dominate, resulting in the marginalization of another. This often affects groups who are a minority in relation to certain personal characteristics.
WEIRD	This acronym refers to Western, Educated, Industrialized, Rich and Democratic societies. Most research is conducted on, and conducted by, relatively homogeneous samples from WEIRD societies. This limits the generalizability of a large number of research findings, particularly given that WEIRD people are often psychological outliers. It has been argued that "WEIRD psychology " started to evolve culturally as a result of societal changes and religious beliefs in the Middle Ages in Europe. Critics of this term suggest it presents a binary view of the global population and erases variation that exists both between and within societies, and that other aspects of diversity are not captured.

Summary

This glossary is a first step in creating a common language for these concepts, facilitating discussions about the strengths and weaknesses of different Open Scholarship practices, and ultimately helping to build a stronger research community[8].

As with all terminologies, this glossary will be the subject of iterative improvement and updates. We encourage the scientific community to read the terms with critical eyes and to provide feedback and recommendations on FORRT's website, where instructions on how to contribute are provided and where the live version of all defined terms is publically available.

Competing interests

The other authors declare no competing interests.

Data and materials availability

All materials produced are available in its entirety at https://forrt.org/glossary

Acknowledgements

We would like to thank Dr. Eiko Fried and Professor Dr. Eric Uhlmann for helping steer the discussion and reviewing some Glossary terms.

Author contributions

Conceptualization: S.P., F.A., and B. Aczel.

Data curation: F.A. and S.G.

Formal analysis: F.A. and S.G.

Investigation: S.P. and F.A.

Methodology: S.P. and F.A.

Project administration: S.P. and F.A.

Software: F.A. and S.G.

Validation: S.P., F.A., M.M.E., and S.G.

Visualization: F.A. and S.G.

Writing - original draft: S.P., F.A., M.M.E., S.G., M.L.V., M.R.V., J.F.M., B. Arendt, J.T., M. Pownall, T.R.E., B. Szaszi, E.C., T.B.L., A.B.M., W.X.N., T.K., Y.Y., A.O.M., B.J.B., S.F., H.E.K., L.S., E.N., M.J., M. Paul, G.H.G., A.J.P., O.M.R., O.N.S., C.R.M., E.M.R., S.K.Y., A.B., E.A. Woodward, B.J.G., A.A.F., Q.X., A.G.C., C.L., C.T.K., F.D., R.M., B. Schmidt, B.G.F., M.Z., E.L.H., T.D., B.M.J.K., N.V., K.K., M.A.B., A.A., D.G.R., J.M., R.M.R., A.H., M.W., T.W., C.R.P., O.R.v.d.A., S.L.K.S., L.A.A., J.S.P., C.M.F.d.O., B.B.B., M.K.T., D.M., R.A.M., G.H.V., H.H., J.P.C., N.B., B.V., E.G.-G., Z.M.F., O.L., N.A.-A., Y.-F.Y., M.I., M.L., J.E.B., G.R., A.J.S., A.L., C.P., S.V., C.J.G., A.H.A.-H., and B. Aczel

Writing - review & editing: S.P., F.A., M.M.E., S.G., L.M., M.L.V., M.R.V., J.F.M., B. Arendt, M. Pownall, T.R.E., B. Szaszi, T.B.L., W.X.N., T.K., Y.Y., A.O.M., B.J.B., S.F., H.E.K., L.S., E.N., M.J., M. Paul, G.H.G., A.J.P., O.M.R., O.N.S., C.R.M., E.M.R., S.K.Y., A.B., A.R.B., B.J.G., A.A.F., Q.X., A.G.C., C.T.K., F.D., B. Schmidt, B.G.F., M.Z., E.L.H., T.D., B.M.J.K., G.F., K.K., M.A.B., A.A., D.G.R., J.M., R.M.R., A.H., M.W., T.W., C.R.P., O.R.v.d.A., S.L.K.S., L.A.A., J.S.P., C.M.F.d.O., B.B.B., M.K.T., D.M., R.A.M., H.H., J.P.C., N.B., E.A. Williams, T.B.R., S.R., B.V., B.J.I., M.K., Z.K., E.G.-G., Z.M.F., O.L., N.A.-A., K.I., Y.-F.Y., M.G.-H., M.I., K.F., M.L., V.A., J.E.B., A.A.A., S.A.-J., W.L., G.R., S.L.M., A.J.S., E.G.-P., A.L., C.P., A.T., S.V., M.C.L., C.J.G., J.K.W., A.H.A.-H., D.A.K., and B. Aczel

List of Supplementary Materials:

The unabbreviated version of the Open Scholarship Glossary can be found at the supplementary materials.

References:

1. Wilson, B. & Fenner, M. Open researcher & contributor ID (ORCID): solving the name ambiguity problem. *Educ. Rev* **47**, 1–4 (2012).
2. Gelman, A. & Loken, E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Dep. Stat. Columbia Univ.* **348**, (2013).
3. Byrne, J. A. & Christopher, J. Digital magic, or the dark arts of the 21st century—how can journals and peer reviewers detect manuscripts and publications from paper mills? *FEBS Lett.* **594**, 583–589 (2020).
4. Tierney, W. *et al.* Creative destruction in science. *Organ. Behav. Hum. Decis. Process.* **161**,

291–309 (2020).

5. Kathawalla, U.-K., Silverstein, P. & Syed, M. Easing Into Open Science: A Guide for Graduate Students and Their Advisors. *Collabra Psychol.* **7**, (2021).
6. Crüwell, S. *et al.* Seven Easy Steps to Open Science. *Z. Für Psychol.* **227**, 237–248 (2019).
7. Azevedo, F., *et al.* Introducing a Framework for Open and Reproducible Research Training (FORRT). (2019) doi:10.31219/osf.io/bnh7p.
8. Nosek, B. A. & Bar-Anan, Y. Scientific utopia: I. Opening scientific communication. *Psychol. Inq.* **23**, 217–243 (2012).