OxFOS

Oxford Forum of Open Scholarship

AI versus open research: Exploring the evolving tensions

Monday 2 March 2026, 14:00-16:00

BODLEIAN LIBRARIES OFFICE FOR OPEN RESEARCH

go.glam.ox.ac.uk/OxFOS26

#OxFOS26

AI and Open Science: Reinforcing inequity or enabling a better world?

Sabina Leonelli

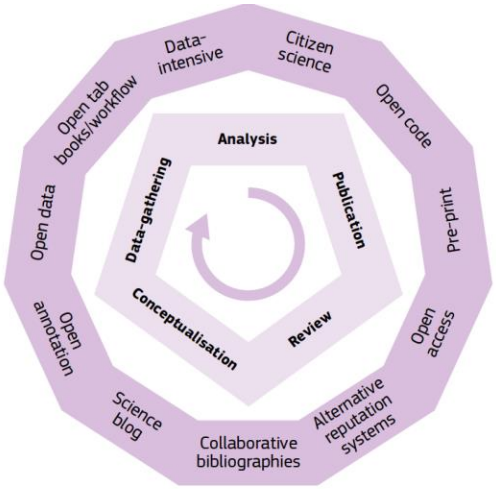
Technical University of Munich

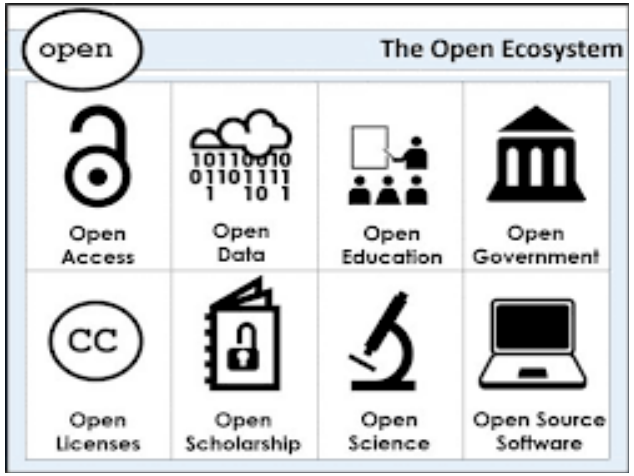
European Research Council

Established by the European Commission

Open Science today: Multiple, disputed, commercialized

- **Techno**-driven: reliance on new digital/computing tools and related online accessibility
- **IP**-driven: admissible forms of ownership and exchange (or lack thereof)
- **Value**-driven: openness defines research and its outputs
- **Practice**-driven: workflows and procedures of of collaboration and sharing [e.g. reproducibility]
- **Rule**-driven: conditions for knowledge production to count as scientific [e.g. venues for critical scrutiny, governance]
- **Policy**-driven: cosmopolitan aspirations for transnational science [e.g. disease surveillance]





Open Science today: Multiple, disputed, commercialized

- Failure of cosmopolitan ideal
- Triumph of corporate ownership over research tools and results
- Unclear role for pluralism and epistemic diversity: politically tainted, scientifically unfashionable vis-à-vis dominant repertoires & reproducibility and interoperability standards
- Problematic quality assessment procedures: rise of misinformation
- Deeply unequal, fragmented research environment
- Hyper-reliance on fragile digital systems: unsustainable ecosystem

In which sense is research open? Can and should it be, if at all? Why continue to talk about openness?

PHIL_OS (21-27): A Philosophy of Open Science for Diverse Research Environments

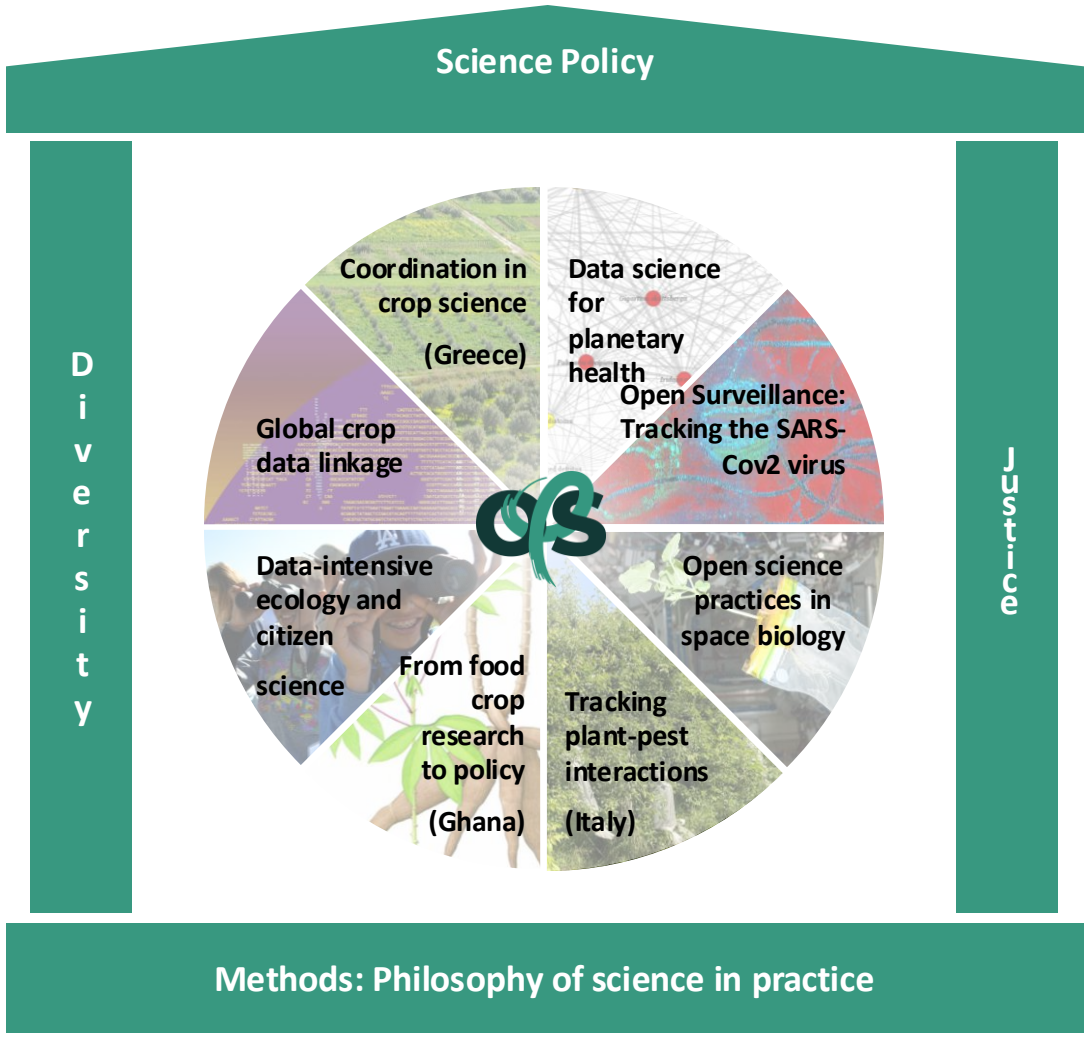
European Research Council
Established by the European Commission

Situating research processes

To understand how inferential practices relate to characteristics of research environments, epistemic diversity and (in)justice

- **Approach:** *co-produced* philosophy, history and social studies of science (with scientists, OS infrastructures and policy-makers)
- **Focus:** interpretations of openness as a window on the epistemic implications of

1. **Diversity** in research environments
   - Backgrounds and skills
   - Resourcing: material, human, conceptual, institutional, infrastructural
   - Grounds for reasoning around “best practice”
2. **Inequity** between research environments
   - Constraints on methods, resourcing and networks
   - Reputational cycles and epistemic injustice



Sabina Leonelli
Principal Investigator
homepage
sabina.leonelli@tum.de
CV

Rena Alcalay
Research Fellow (from Feb 2025)
homepage
RGoldstein@flagler.edu

Paola Castaño
Research Fellow
homepage
p.a.castano-rodriguez@exeter.ac.uk

Richard Williams
Research Fellow
homepage
r.williams2@sheffield.ac.uk

Emma Cavazzoni
PhD student
emma.cavazzoni@tum.de

Joyce Koranteng-Acquah
PhD student
jk677@exeter.ac.uk

Nathanael Sheehan
PhD student
homepage
ns651@exeter.ac.uk

Fotis Tsiroukis
PhD student
ft323@exeter.ac.uk

Rachel Ankeny
International Collaborator (Adelaide)
homepage
rachel.ankeny@adelaide.edu.au

Dasapta Erwin Irawan
Project illustrator
homepage
r-win@office.itb.ac.id

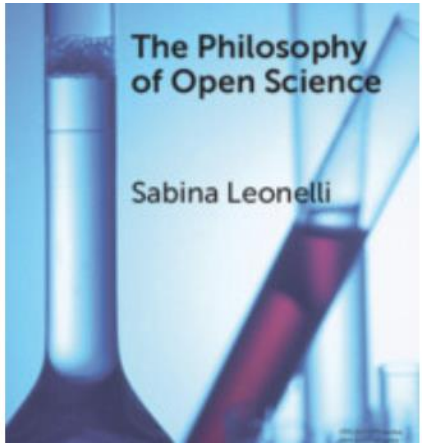
Desantila Hysa
Project administrator & webadmin
desantila.hysa@tum.de

Rose Trappes
Research Fellow 2021–2024
homepage
rose.trappes@uib.no


Global crop data linkage
Citizen science and data-intensive ecology (India)
Tracking plant-pest interactions (Italy)
Data science for planetary health
Open surveillance: tracking the SARS-Cov2 virus
Open science practices in space biology: NASA's GeneLab
From food crop research to policy (Ghana)
Coordination in crop science (Greece)

Openness as “sharing resources”: An object-oriented view

- about **unlimited access**: making any research element available at any time for everyone
- about the **digital transformation**: it is a novel phenomenon and completely dependent on ICTs
- always **good**: it automatically improves the content of science as well as researchers’ working conditions
- **global**: it can reach everybody with an interest in research, no matter where they are based
- facilitating **equity** in research production and consumption: it makes previously inaccessible resources available to those who may wish to use them



Three varieties of structural injustice

(1) Inequity in resourcing: allocation, access and deployment

(2) Misalignment between resourcing and scientific goals

(3) Misalignment between scientific goals and labor conditions

Three varieties of structural injustice

(1) Inequity in resource allocation, access and deployment

- Resourcing not grounded on merit or scientific relevance
  - whether and how science is recognised as worthy of investment, the availability of training in research and related digital skills, whether scientific investigations intersect with the activities of commercial-military enterprise (and how), national income, socio-economic disparity between rural and urban areas, institutional support
  - Problems with ’convenience AI” (Leonelli and Mussgnug under review, preprint on PhilSci Archive)
- Short-term availability privileged over capacity-building
- Post factum training and tech adoption – little if any involvement in design and governance

(2) Misalignment between resourcing and scientific goals

(3) Misalignment between scientific goals and labor conditions

Three varieties of structural injustice

(1) Inequity in resource allocation, access and deployment

(2) Misalignment between resourcing and scientific goals

- Local constraints, practical exigencies, socio-political agendas shape research directions, sometimes in tension with scientific interests
- Digital divide is also a *divide in epistemic power*: those who can shape the research agenda vs those who cannot

(3) Misalignment between scientific goals and labor conditions

Three varieties of structural injustice

(1) Inequity in resource allocation, access and deployment

(2) Misalignment between resourcing and scientific goals

(3) Misalignment between scientific goals and labor conditions

- Divide sharpened by constraints and expectations on what counts as credit for whom and within which system (local, national, global)
  - Publishing, patenting, outreach to various publics
- Goals shaped by expectations around future employment, including expected and valued outcomes (often NOT those associated with academic excellence)

Implications

1. shifts in research content: research directions picked to comply with existing constraints, resulting in lack of research spent on topics, domains and goals most relevant to vulnerable scientists and their contexts
2. exclusion of researchers: substantive portions of scientific community are unable or unwilling to contribute to internationally recognized body of knowledge to the best of their abilities
3. exclusion of methods/data: significant methods and sources of evidence are undermined or excluded from global scientific discourse
4. ineffective reproducibility guidelines: policing compliance with centralized rules instead of quality/trustworthiness of research design and processes
5. diminished research quality and resilience: loss of quality and reliability in processes and outcomes of inquiry
6. ineffective Open Science: sometimes implemented, yet no increase of visibility and findability of research

[Leonelli, S under review "Not All Research Environments Are Created Equal"]

The impact of AI: from OS..

- Failure of cosmopolitan ideal & triumph of corporate ownership over research tools and results → Amplified and strengthened by big tech AI platforms with global reach and secretive IP regimes
- Unclear role for pluralism and epistemic diversity -> Hype towards AGI strengthening hostility towards diversity as something to be stamped out through standardization
- Problematic quality assessment procedures: rise of misinformation → AI subverts information quality assessments by making data provenance/processing more difficult to track and disincentivizing investment in expert human assessment
- Deeply unequal, fragmented research environment → AI amplifies existing digital divides by broadening the gap between needs and abilities served by AI tools and those who are not
- Hyper-reliance on fragile digital systems: unsustainable ecosystem → AI disincentivizes investment in data curation and related venues and infrastructures in the long term, and takes attention away from environmental costs and prospects

The impact of AI: .. to structural injustice

- Failure of cosmopolitan ideal & triumph of corporate ownership over research tools and results → Amplified and strengthened by big tech AI platforms with global reach and secretive IP regimes → shifts in research content
- Unclear role for pluralism and epistemic diversity -> Hype towards AGI strengthening hostility towards diversity as something to be stamped out through standardization → exclusion of researchers, methods, data
- Problematic quality assessment procedures: rise of misinformation → AI subverts information quality assessments by making data provenance/processing more difficult to track and disincentivizing investment in expert human assessment → ineffective reproducibility guidelines and diminished research quality
- Deeply unequal, fragmented research environment → AI amplifies existing digital divides by broadening the gap between needs and abilities served by AI tools and those who are not → exclusion of researchers, methods, data
- Hyper-reliance on fragile digital systems: unsustainable ecosystem → AI disincentivizes investment in data curation and related venues and infrastructures in the long term, and takes attention away from environmental costs and prospects → diminished research quality and resilience

Beyond object-oriented sharing and standards-driven AI: Social engagement as central to good (open) science

- Beyond sharing research elements: standards, guidelines and computational tools are essential, but cannot be prioritized over conditions that make research elements *meaningful* for users
  - Slapping stuff on internet is NOT good OS practice NOR a viable form of ‘transparency’
  - Unsupervised data mining and modelling is NOT good science NOR a viable substitute for human research labor (creativity and judgement)
- Scientific inquiry as quintessential case of **collective agency:** critical to make it possible for people to work and reason together about the meaning of data analysis
  - Makes research scrutinizable as a common good
  - Anchors relationship between science and society
- How to build OS (and AI) with diversity and judgement as starting points and not obstacles?

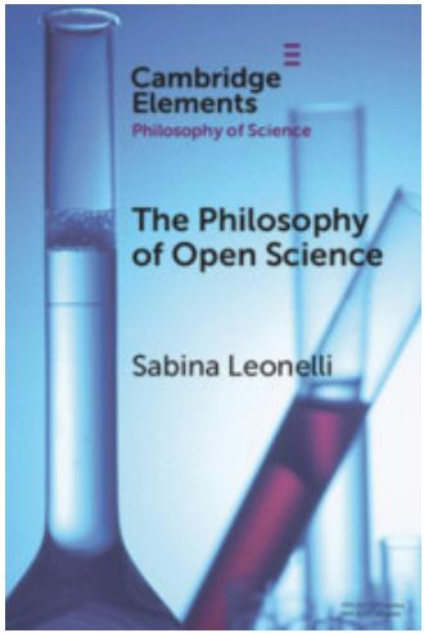
Openness as judicious connection: A process-oriented philosophy of OS

Discovery as skilled, distributed interaction with the world

- Focus on social agency: creating new intimacies, facilitating trust and collaboration among those who access research elements (data, code, papers..)
- Epistemic justice and diversity as crucial conditions for inquiry

Connection needs to be *judicious*:

- Situated and responsive to context
- Identifying what constitutes relevant context is key part of any investigation



Transparency

Inclusion

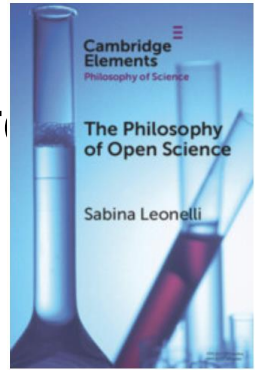
Inclusion

Quality

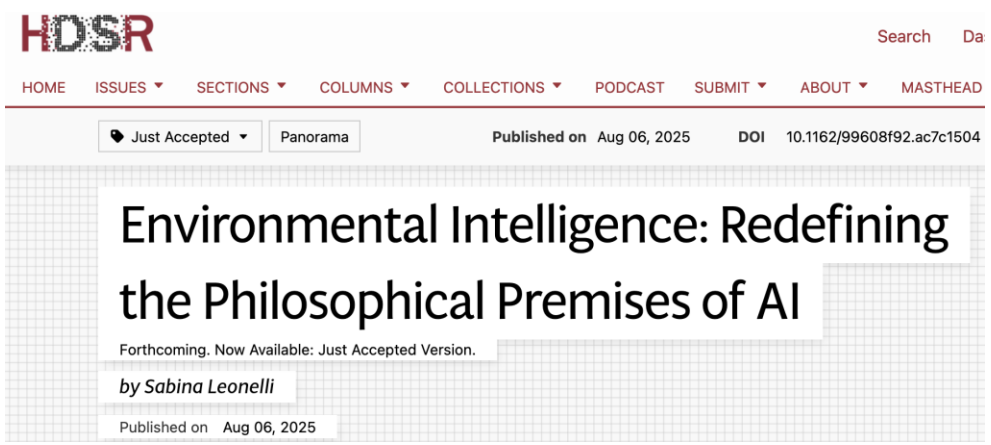
Transparency

Towards Engaged and Inclusive OS

- about **responsible use**
- about the critical and constructive scrutiny of how **digital platforms** can support existing and future work
  - Encouraging development of relationship that can sustain and nurtur scientific research in the long term
- **good for some and not others**: value-judgements and choices are unavoidable when developing open research and infrastructures
- **accessible to some and not others**: transparent criteria for which users are privileged can be a platform for trustworthiness
- facilitating **equity** in research production and consumption: it makes previously inaccessible resources *more easily* available to those who may wish to use them *for specific purposes* (whose social and scientific value has been explicitly evaluated)



Diversity as a starting point



- Acknowledging multiple perspectives and well-established (but diverse) cultures of openness: beware of centralized assessment criteria

- Support openness across publicly and privately funded institutions, taking care not to single out publicly funded institutions as the only conceivable target for OS policies and assessment

- Invest in understanding scientific motivations for specific habits and preferences, beyond conformity to problematic assessment / credit systems (a ‘culture problem’ is not necessarily a ‘people problem’)
  - Attention to ECRs is key, e.g. Global Young Academy activities since 2012

Diversity as a starting point

- Support researchers' transition to OS: cannot simply be delegated down, especially as researchers are already overwhelmed by admin and management

- Don't buy into 'novelty' narrative relating to OS: openness has long been a constitutive value for scientific research, with many different ways of operationalizing it over the last few centuries

- **Beware of attempts to interpret openness as disregard for expertise and know-how – and related hype towards automation**
  - **Build in methods to identify and value expert knowledge**

Example: Fostering Findability over (immediate) Accessibility

- Share metadata, require human contact for data sharing (including agreement around conditions for re-use) [e.g. some biomedicine-oriented data infrastructures]

- Foster direct contact between data creators/holders and users
  - Increase of trust
  - Opportunity for better contextualization and future collaboration
  - Visibility and credit for data creators

- Encourage long-term Communities of Practice

Sheehan, N. and Leonelli, S. (2024) Reconciling Data Actionability and Accountability in Global Health Research. *Global Health Research*.

Example: Situated versus norm-directed approaches to misinformation

- Norm-directed approach to misinformation parallels object-oriented view of OS
  - Means: building a rule-governed framework to censor / eliminate / flag problematic bits; information comes in commodified forms that can be traded and formalised, then routinised and scaled up through AI
  - Goal: find methods to address misinformation that are as generalisable as possible, so that their reliability, validity and reproducibility can be tested at scale, and t can be used to develop computational (and if possible, automated) tools to police and clean the public knowledge sphere

[Boumans, Goldenberg and Leonelli under contract, Understanding Misinformation. Cambridge University Press]

- Situated approach: misinformation is not defined by rules around what counts as truth or falsity, but rather by
- **localised** analysis;
- **qualitative** means for tracking the history and processing of information
- **social/functional** focus of analysis, where cases of misinformation are studied in terms of **their role within processes of inquiry**

Example: Qualitative Research as Reference Point

- Humanities, Arts and Social Sciences as alternative models of OS practice

- Shift of gears: emphasis on relations, situatedness of knowledge claims and research processes, contextualization
  - No reliance on ‘convenient tech’ and purely quantitative analysis
  - Triangulation with mixed methods
  - Hermeneutical effort at the core of knowledge-making
  - Reproducibility remains important but is not automated or universalized (e.g. citation checks in history are essentially an interpretative act, beyond being simply “wrong” – a matter for academic debate)

- Reflexivity at the heart of openness as engaged empirical inquiry

Addressing structural injustice means fostering structural change

Structural change requires more than mitigation strategies: **reframing the conceptual and institutional grounding of empirical inquiry**

Embracing vulnerability and community action as paths to open, reliable, equitable knowledge-making

Concrete pathways to reform:

- Engagement across local communities and beyond professional science **(Public Science Lab)**
- Reform of institutional, material, digital and computational infrastructures to serve widely diverse capabilities and goals **(Ethical Data Initiative)**
- Lobbying to reform education and labor markets, funding for transdisciplinary research **(PianetaLab)**

Thank you for your attention

pianoo lab

TransforM
Munich Center for Transformative Technologies and Societal Change
From more innovation to better innovation.

PUBLIC SCIENCE LAB

erc
European Research Council
Established by the European Commission

Philosophy of Open Science
OPEN?
OpenScienceStudies.eu

Join Us for the March 2026 Ethical Data Initiative Townhall

EDI The Ethical Data Initiative

Generative AI and Open Science

Mapping an Emerging Tension

Mcebisi Ntleki

Executable Pair

Weights · Architecture

Replication Pair

Data · Training Code

Transparency Pair

Documentation · Access Terms

Introduction & Generative AI

The Openness Paradox in Contemporary Research

The generative AI tools now embedded in research workflows are often trained on decades of open research – yet offer researchers little insight into their training, limited control over their operation, and no guarantee of stable access.

Open Science Has Come of Age – And Yet

Hard-won progress		Growing opacity
Major funders mandate open data		Training data undisclosed
Preprint culture – rapid transparent dissemination	PARADOX	Model weights often proprietary
Reproducibility: from aspiration to expectation		API access & behaviour change without notice
Open access publishing now the norm		No stable, inspectable research instrument

A Concrete Example

Reproducing a GPT-4 study – six months later

Early 2024 | Original Study Published

Team analyses gender bias in scientific text using GPT-4. Methods section documents the model and API call parameters.

Late 2024 | Reproduction Attempted

Second group follows the documented methodology precisely, using the same prompts and parameter settings.

Outcome | Reproduction Fails

The exact GPT-4 version is no longer accessible. Training data remains undisclosed. Model behaviour has changed. Results cannot be replicated.

This isn't researcher failure – they followed best practices. Those practices, built for stable tools, are inadequate for proprietary, evolving infrastructure.

Why It Matters – And Why Debate Fails

i EPISTEMIC · Science depends on independent verification – black-box AI resists direct inspection.

ii EQUITY · Proprietary access creates new stratification; research capacity depends on institutional resources.

iii DEVELOPMENT · Science advances cumulatively – but we cannot improve systems we cannot inspect.

'AI should be open' can mean:

- **Release the weights**
- **Disclose training data**
- **Ensure API access**

This paper's three contributions:

1. A taxonomy disaggregating 'openness' into six distinct dimensions
2. Mapping arguments for/against openness onto specific dimensions
3. A context-sensitive framework for when different forms of openness matter most

What is Generative AI?

Computational systems capable of producing novel outputs – text, images, audio, video, code, or structured data – in response to prompts, by learning statistical patterns from vast training corpora.

Language & Code	Vision & Audio	Structured Data	Key distinction
Essays · Reviews · Code Translation · Summaries	Images · Video · Music Speech synthesis	Tables · Graphs Molecules · JSON	Billions of learned parameters whose behaviour emerges from training – not explicit programming.

Systems like GPT-4, Claude, and Gemini now perform tasks previously requiring expert human judgement:

- Literature reviews
- Experimental design
- Qualitative analysis
- Code generation
- Manuscript preparation
- Research synthesis

Four Dimensions of a Generative AI Model

A model is simultaneously a dataset, an algorithm, a computational process, and an interface – each with distinct openness challenges

1	**Dataset** *Training corpus*	Embodies its training data – knowledge, biases, and blindspots.	⚠ *Often undisclosed; biases impossible to anticipate*
2	**Algorithm** *Architecture & weights*	Billions of learned parameters encode statistical patterns.	⚠ *Weights may be proprietary or too large to inspect*
3	**Computational Process** *Inference at query time*	Non-deterministic – same prompt can yield different outputs.	⚠ *Outputs unpredictable, not merely stochastic*
4	**Interface** *Access layer*	Researchers access models via APIs that change without notice.	⚠ *Prevents systematic documentation of limitations*

Layers of Opacity

Each layer of opacity compounds the others – proprietary data makes weights uninterpretable, which makes non-determinism unpredictable, which API access prevents documenting

Training Data

Undisclosed → biases undetectable

Model Weights

Proprietary → bias pathways opaque

Inference Process

Non-deterministic → unpredictable, not merely stochastic

API Interface

Changes without notice → documentation impossible

Result: compound opacity – each unknown amplifies the unknowability of every other layer

What Model Outputs Are – And What This Means for Science

Not observations

No evidential anchor to external reality

Not derivations

Inference path is inaccessible

Not documents

Generated reconstructions, not retrieved records

Artefacts of a socio-technical system whose internal evidential structure is largely inaccessible. Plausible by construction – optimised to be fluent, not to be true.

Scientific norms under pressure

Reproducibility

Same model at different times = systematically different behaviour

Traceability

Causal pathway from training data to output is invisible

Principled Disagreement

Disagreeing with a model requires access to its reasoning – unavailable

Instrument Stability

Updated without announcement = different instrument, same name

Next: Six Dimensions →

D1 Model Weights

PAIR I · EXECUTABLE

The trained parameters – billions of floating-point numbers in tensor arrays – that encode everything the model learned during training.

OPEN

Downloadable and modifiable. Users can run locally, fine-tune, and inspect.

Examples: Meta Llama, Mistral, Falcon, Qwen.

CLOSED

Access only via inference API – weights are never exposed.

Examples: GPT-4, Claude, Gemini. Users cannot inspect, modify, or reproduce behaviour.

WHY IT MATTERS

Reproducibility

Same weights + architecture + params → deterministic outputs. Minimal sufficient condition for replicating behaviour independently.

Derivative Research

Fine-tuning, mechanistic interpretability, and architectural analysis all require direct weight access that APIs cannot provide.

KEY TENSION

Open weights enable adversarial fine-tuning and safety mechanism removal. And 'open' weights are only meaningful to resource-rich actors – GPU requirements create de facto tiered access.

A 70B-parameter model is a point in an unimaginably high-dimensional space – weights encode, but do not explain, behaviour.

D2 Inference Code & Architecture

PAIR I · EXECUTABLE

The computational specification of how weights are organised and executed – layer types, attention mechanisms, activation functions, and the full control flow transforming inputs into outputs.

OPEN

Published architectural specs plus executable code.

Examples: HuggingFace Transformers, academic releases with PyTorch/JAX implementations.

CLOSED

Undisclosed implementation details; access via API only.

Reality: most models follow published transformer templates – but crucial production-level details remain hidden.

WHY IT MATTERS

Independent Verification

Researchers can verify the claimed structure produces reported behaviour, and identify numerical instabilities or computational bottlenecks.

Architectural Science

Systematic study of how depth, width, attention heads, and normalisation choices affect capabilities, efficiency, and failure modes.

KEY TENSION

The gap between 'architecture in principle' (paper equations) and 'architecture in practice' (production code with stability tricks) is significant. Secrecy is also largely illusory – templates are published — but undocumented details still matter enormously.

Architecture is the interpretive key to weights – the same numbers mean entirely different things in a transformer versus any other model family.

D3 Training Data

The corpus from which statistical patterns are extracted – text, images, code, multimodal collections spanning terabytes to petabytes. The epistemic foundation of the model.

OPEN

Comprehensive disclosure: enumerated sources, sampling procedures, filtering criteria – ideally the data itself or reproducible collection methods.

Examples: The Pile, RedPajama, ROOTS, C4.

CLOSED

Vague descriptions: 'diverse internet data' or 'proprietary datasets.' No mechanism to inspect what patterns the model was exposed to, what biases it absorbed, or whether benchmarks were contaminated.

WHY IT MATTERS

🔎 Provenance & Bias

Trace whether outputs reflect genuine regularities or collection artefacts. Understand demographic, linguistic, and temporal biases encoded in behaviour.

🧪 Contamination Detection

Test whether benchmarks leaked into training – a critical issue that can invalidate results and systematically overstate capabilities.

KEY TENSION

Full disclosure may itself cause harm: exposing scraped personal data, violating copyright at scale, or providing roadmaps for poisoning future training runs. The data is simultaneously essential scientific record and organisational liability.

⚡ *What matters is not the raw corpus but the effective distribution after filtering, deduplication, and sampling – and these curation processes are often proprietary.*

D4 Training Code & Methodology

The procedural knowledge transforming data and architecture into trained weights – optimisers, learning rate schedules, regularisation, curriculum design, and the myriad choices constituting a 'training recipe'.

OPEN

Published training scripts + comprehensive methodology docs: optimiser choices, hardware config, numerical precision, convergence criteria.

Examples: Llama training code, EleutherAI GPT-NeoX.

CLOSED

Minimal disclosure beyond 'we trained a transformer'. The entire causal path from initialisation to final weights is undocumented. Each group must independently rediscover the same numerical stability tricks.

WHY IT MATTERS

Process Reproducibility

Independent verification that following the documented procedure actually produces models with the reported characteristics – the foundation of scientific validity.

Failure Mode Transfer

Reveals gradient clipping tricks, warm-up schedules, mixed-precision strategies – accumulated practitioner knowledge that shouldn't need rediscovering from scratch.

KEY TENSION

Even with complete code, reproduction is practically impossible for most: frontier training requires infrastructure costing hundreds of millions. Training is stochastic and sensitive to hardware and software versions – exact replication may be theoretically possible but practically unattainable.

The same architecture and data can yield dramatically different models depending on how training proceeds – methodology is where potential becomes realised capability.

D5 Documentation

Meta-level knowledge about model behaviour – capabilities, limitations, failure modes, biases, evaluation benchmarks, and known risks. The 'methods section' for the model itself.

OPEN

Comprehensive model cards, detailed evaluation reports across diverse tasks and demographics, disclosed safety testing results, honest failure characterisation.

Examples: Anthropic model cards, Google Model Reports.

CLOSED

Marketing materials emphasising capabilities, cherry-picked benchmarks showcasing strengths, vague disclaimers with no actionable guidance. Users form mental models from narratives rather than technical reality.

WHY IT MATTERS

🎯 Appropriateness Judgements

Determine whether a model suits a specific research question; understand which biases might contaminate results; know when outputs should be treated as exploratory vs. reliable.

⚖️ Accountability

Without stated limitations, failures cannot be distinguished from unexpected bugs vs. known constraints users ignored. Documentation enables both praise and blame.

KEY TENSION

Honest documentation creates liability: explicitly stating failure modes provides evidence for negligence claims. Many behaviours are emergent and only discovered through use – comprehensive documentation may be impossible at release.

⚡ *Documentation quality correlates inversely with competitive pressure – organisations racing to release have strong incentives to emphasise strengths and under-investigate limitations.*

D6 Access Terms

The legal, economic, and procedural conditions governing who can use a system, for what purposes, and under what constraints – licensing, pricing, quotas, approval processes, acceptable use policies.

OPEN

Permissive licences (Apache 2.0, MIT, OpenRAIL), free or cost-recovery pricing, minimal usage restrictions, no gatekeeping beyond basic account creation.

Enables any researcher to audit, evaluate, probe, and build.

CLOSED

Expensive API access, institutional approval requirements, prohibitions on adversarial testing or bias auditing, geographic restrictions, opaque access decisions.

Critical evaluation restricted to those the provider approves.

WHY IT MATTERS

Epistemic Justice

Access terms determine who can participate in research. Restrictive terms consolidate capacity in well-funded institutions, systematically excluding under-resourced researchers.

Reproducibility at Risk

If pricing changes, terms tighten, or systems are deprecated, published research becomes unreproducible – not due to methodological opacity, but access revocation.

KEY TENSION

Free access at scale is economically infeasible without subsidy – but subsidies introduce their own equity questions. ToS restrictions often prohibit precisely the adversarial testing and bias auditing most needed for scientific rigour.

'Free' access funded by venture capital isn't neutral – it shapes research agendas towards questions that generate data or prestige valuable to providers.

The Tension is Real

The tension between generative AI and open science reflects genuine conflicts between legitimate values – not merely misunderstanding or strategic positioning.

Pulls Towards Openness

Reproducibility · Scientific scrutiny
Epistemic equity · Accountability

vs.

Pulls Towards Closure

Safety · Sustainability
Competitive dynamics · Practical constraints

These are structural features of this technological moment – not temporary problems awaiting simple solutions.

But we can navigate them better with conceptual clarity.

What the Six-Dimensional Taxonomy Gives Us

'Openness' is not monolithic. Each dimension serves different purposes and faces different constraints.

PAIR I · EXECUTABLE

D1 Weights · D2 Architecture

Arguments for weight release address capability transfer, adversarial fine-tuning, and resource asymmetry – quite distinct from architectural openness debates about paper-vs-practice gaps.

PAIR II · REPLICATION

D3 Data · D4 Training Code

Demands for training data disclosure target provenance, bias, and contamination; methodology transparency addresses reproducibility and craft knowledge – different problems requiring different solutions.

PAIR III · TRANSPARENCY

D5 Documentation · D6 Access Terms

Documentation failures are about incentive structures and liability; access term restrictions are about epistemic justice and who participates – conflating them obscures what's actually at stake.

The taxonomy enables proportional reasoning: matching transparency requirements to actual functional needs rather than ideological commitments.

Three Key Insights for Research & Policy

01

Context Determines Appropriate Transparency

High-stakes confirmatory medical research (where errors harm patients) requires different openness from exploratory humanities work (where interpretive pluralism is valued). No single transparency model fits all contexts. The taxonomy enables proportional matching of requirements to actual functional needs.

02

Current Practices Involve More Closure Than Legitimate Concerns Justify

Documentation remains poor not because evaluation is infeasible – but because it's expensive and reveals limitations. Training data stays closed not only due to legal complexity – but because opacity prevents accountability. Many restrictions framed as essential safety measures are better understood as choices that could be made differently.

03

The Burden of Proof Should Rest with Closure Advocates

Given open science's demonstrated value across centuries, departures from transparency require positive justification. This burden should be ongoing: closure should be provisional and regularly reassessed as contexts evolve, technologies mature, and mitigations develop – not default, and not permanent.

The Present Moment is Critical

Norms are being established through practice — not deliberate policy.

Researchers

Making ad hoc tool choices driven by convenience, not methodological appropriateness

Institutions

Improvising policies that may become entrenched precedents without adequate deliberation

Funders

Demanding openness performatively while funding proprietary systems – inconsistent signals

Publishers

Struggling to articulate standards for AI-assisted research with no settled frameworks

The choices being made now – often without recognising their long-term implications – determine the norms of tomorrow.

What's at stake

Whether open science principles extend into the AI era

– of or whether a new regime of opacity becomes normalised

These distributed, uncoordinated decisions collectively determine the answer.

The Goal: Productive Disagreement

Not premature consensus – but disagreement about actual values and trade-offs, fought with shared reference points.

✗ Naive Total Openness

Ignores real trade-offs in safety, sustainability, and the practical limits of documentation at scale.

✗ Convenient Closure

Treats organisational interests as scientific requirements, passing off strategic opacity as principled necessity.

✓ The Path Forward

Be maximally open where it matters most. Selectively closed where genuine concerns justify it. And honest about which is which.

What's Actually at Stake

Not simply how we use particular tools.

Whether the epistemological foundations of science – transparency, reproducibility, scrutiny – can be maintained as computational methods become central to knowledge production.

This framework aims to:

Make choices visible

Make trade-offs explicit

Make consequences harder to ignore

> *The answer depends on choices being made now, in laboratories and institutions and policy forums, often without recognising their long-term implications.*

Will evil AI scraping bots be the end of open access?

Niels Stern, OAPEN & Directory of Open Access Books (DOAB)

Oxford Forum of Open Scholarship (OxFOS), Monday 2 March, 14.00–16.00 | Online

oapen


KAREN SKOVGAARD-PETERSEN
Historiography at the Court of Christian IV
Naturens tankelæser
1
En biografi om
Hans Christian Ørsted
Dan Ch. Christensen
MUSEUM TUSCULANUMS FORLAG
Jørgen Hein
The Treasure Collection at Rosenborg Castle III
The Inventories of 1696 and 1718
Royal Heritage and Collecting in Denmark-Norway 1500-1900
Niels Erik Rosenfeldt
The "Special" World
MUSEUM TUSCULANUM PRESS
The Politics of Storytelling
Variations on a Theme by Hannah Arendt
MICHAEL JACKSON
John Pedersen
tvivl & tolerance
et skrift om Pierre Bayle
MUSEUM TUSCULANUMS FORLAG
Scholarly books
and AI

OAPEN Library - an open infrastructure for scholarly books

Quality curation & Open access

Hosting, distributing, preserving scholarly books

oapen

2010

Established by six
European university presses

40,000+

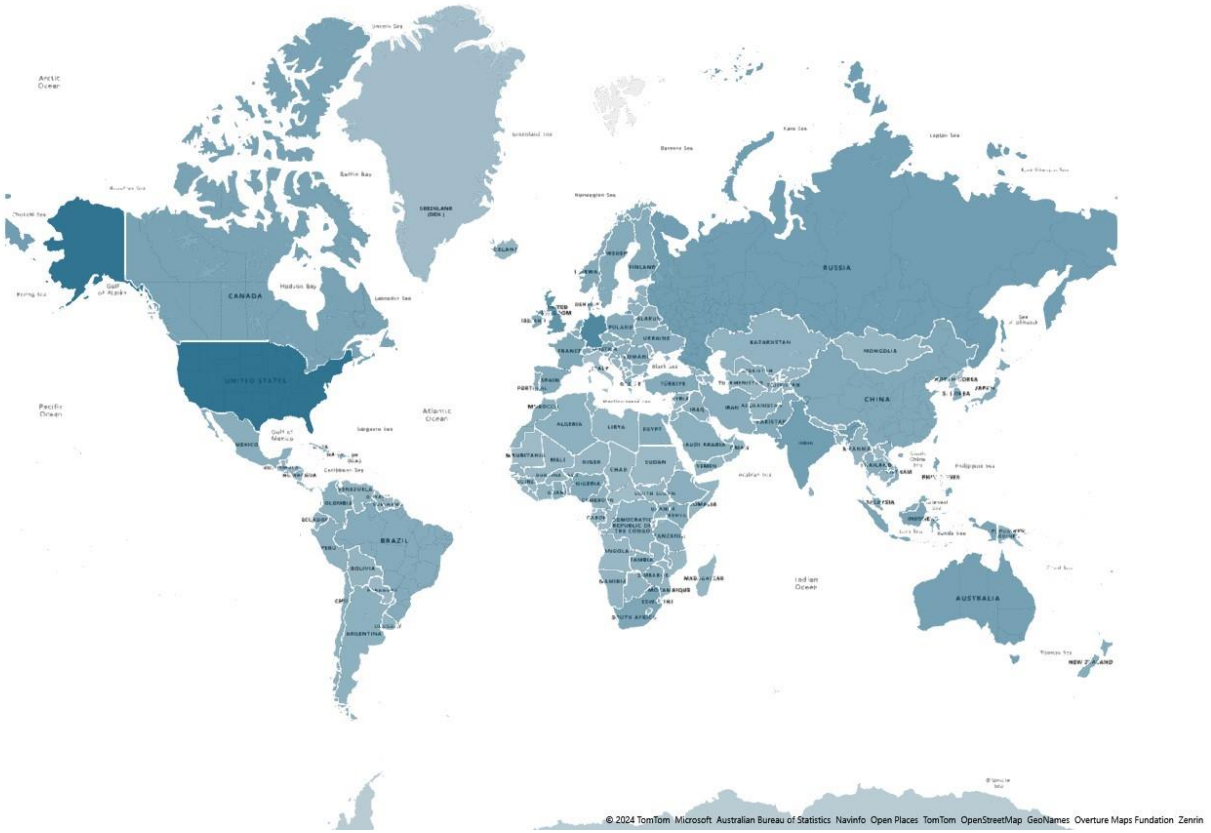
peer-reviewed open access books
from over 450 publishers

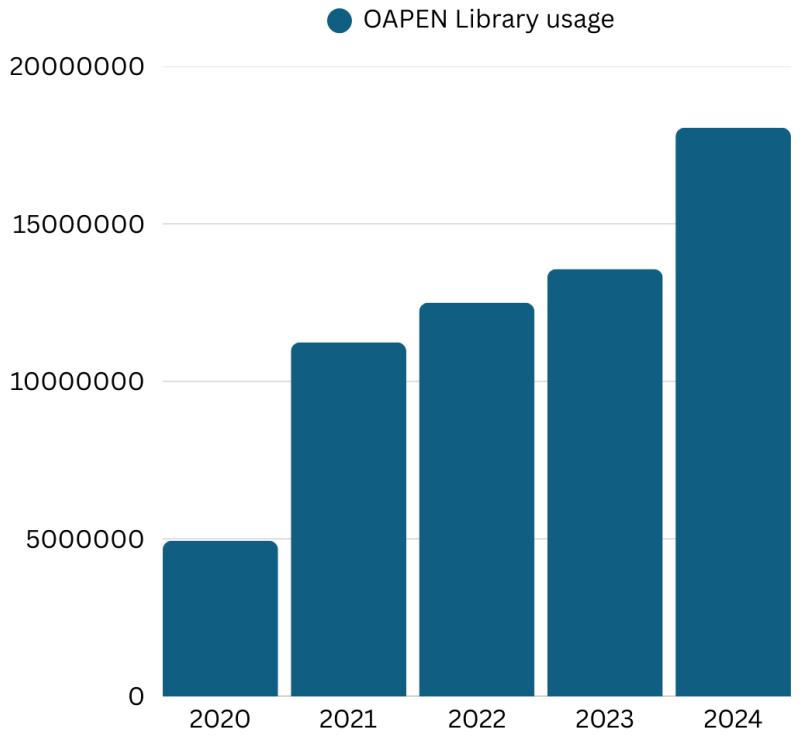
oapen

1.5 million

COUNTER downloads every month in 2024

Global & growing usage of the OAPEN Library

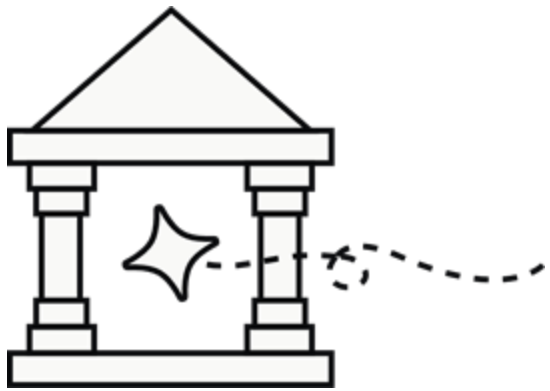




COUNTER conformant usage data of the OAPEN Library provided by IRUS-UK data for 2024

How we serve Libraries

OAPEN provides
seamless access to a curated and fast-growing collection
of thousands of peer reviewed OA books,
easy for libraries to integrate
into their catalogues in a structured and dynamic way.

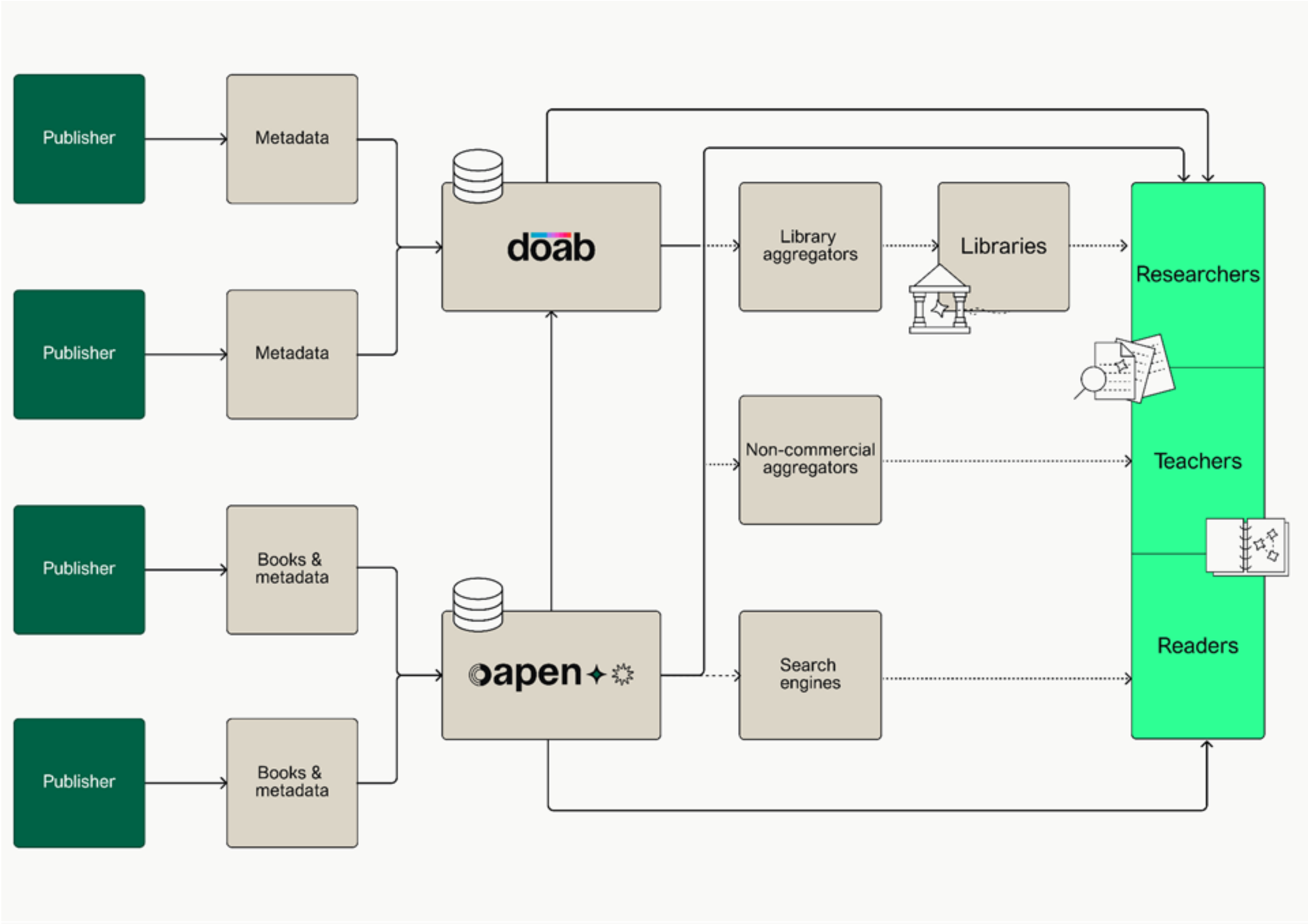


800+

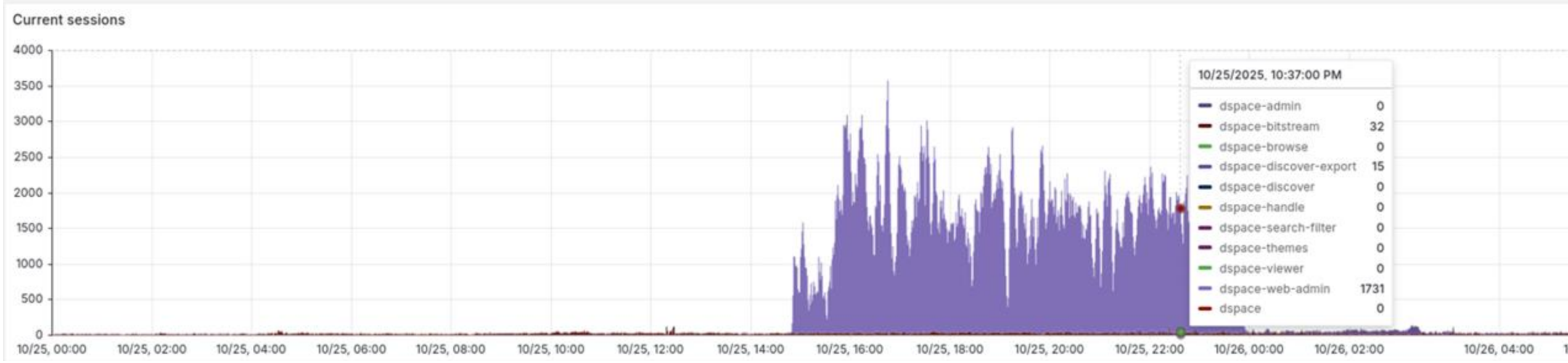
publishers

80+

languages

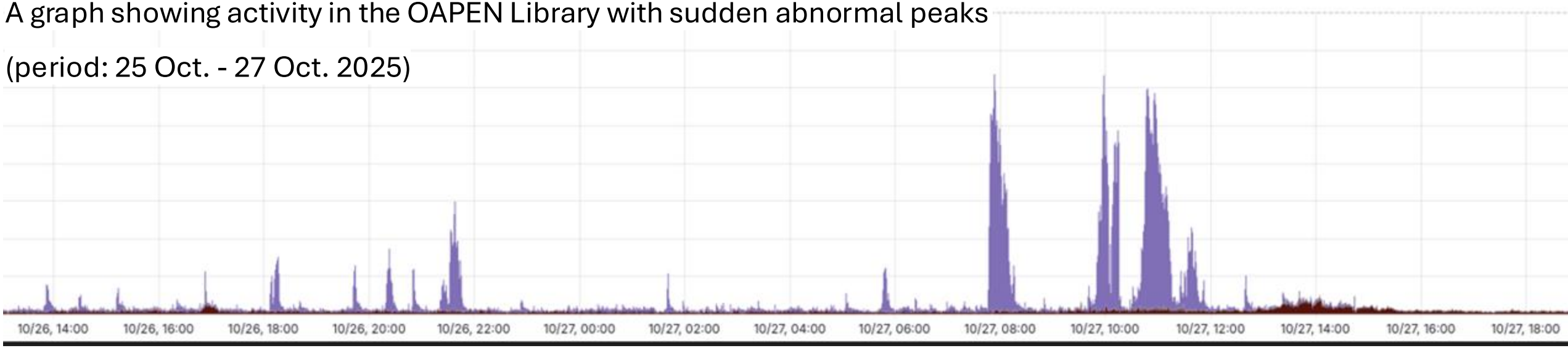


DSPACE



A graph showing activity in the OAPEN Library with sudden abnormal peaks

(period: 25 Oct. - 27 Oct. 2025)



go to hellman

Friday, March 21, 2025

AI bots are destroying Open Access

There's a war going on on the Internet. AI companies with billions to burn are hard at work destroying the websites of libraries, archives, non-profit organizations, and scholarly publishers, anyone who is working to make quality information universally available on the internet. And the technologists defending against this broad-based attack are doing everything they can to preserve their outlets while trying to remain true to the mission of providing the digital lifeblood of science and culture to the world.

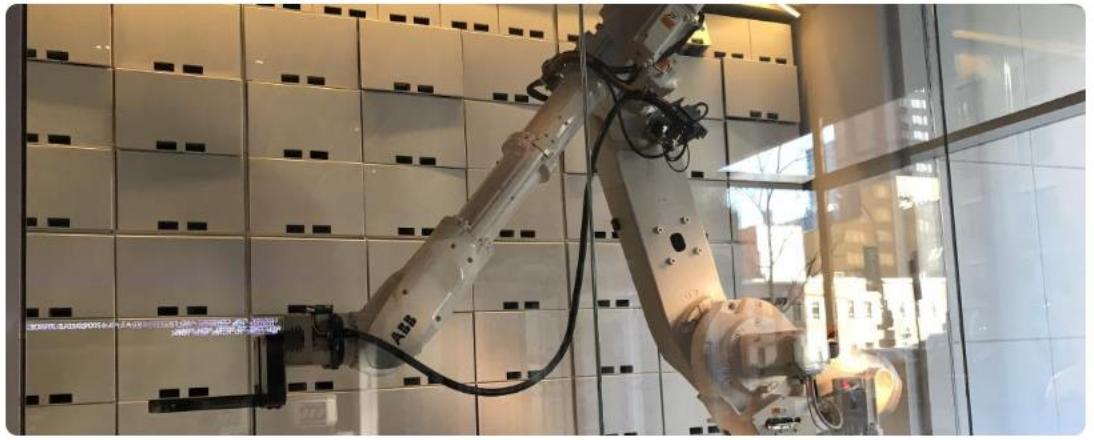
Yes, many of these beloved institutions are under financial pressures in the current political environment, but politics swings back and forth. The AI armies are only growing more aggressive, more rapacious, more deceitful and ever more numerous.

I'm talking about the voracious hunger of AI companies for good data to train Large Language Models (LLMs). These are the trillion-parameter sets of statistical weights that power things like Claude, ChatGPT and hundreds of systems you've never heard of. Good training data has lots of text, lots of metadata, is reliable and unbiased. It's unsullied by Search Engine Optimization (SEO) practitioners. It doesn't constantly

Bots Behaving Badly

Bots Behaving Badly

April 2, 2025 · 7 min · Geoffrey Bilder | Suggest Changes



Cite as:

Geoffrey Bilder (2025, April 2). *Bots Behaving Badly*. Retrieved from https://doi.org/10.59347/w366z-egx85

I've seen an uptick in people I respect concluding that the massive increase in badly-behaved web-scraping bots is intentional and just another example of sociopathic AI companies ripping-up long-respected web conventions for the sake of profit. Here are two recent examples:

Blog posts on the topic in spring 2025

OAPEN: A world of scholarly books: Open to all, built to last.

OAPEN connects libraries,

What was the problem?

The OAPEN Library and DOAB received so many requests from automated systems that they drowned out genuine users. On Tuesday 25 March, the problem reached a tipping point for the OAPEN Library when we saw this happening for several hours. From 8:30 up until 12:00 the number of sessions on the OAPEN Library exploded from 15 per second to 1,500 per second. As a result, the average response time of the OAPEN Library dropped from 0.2 seconds to 1.2 minutes, rendering the service unusable. Thanks to the support of our hosting partner CERN, the system did not crash.

OAPEN: A world of scholarly books: Open to all, built to last.

OAPEN connects libraries, publishers, researchers, and funders with a global infrastructure for peer-reviewed open access books to advance research and the public good.

Funnelling locusts – further reflections on the OAPEN Library and DOAB's response time

In my previous post on our less than optimal performance of the OAPEN Library and DOAB, I wrote about the effect of AI bots on our systems, and that it has become a common problem for many sites providing open access (OA) or open source software content. All of them – including the OAPEN Library and DOAB – see a new threat: not censorship or budget cuts, but bots. A new kind of digital swarm has arrived, and it's putting serious pressure on our in-

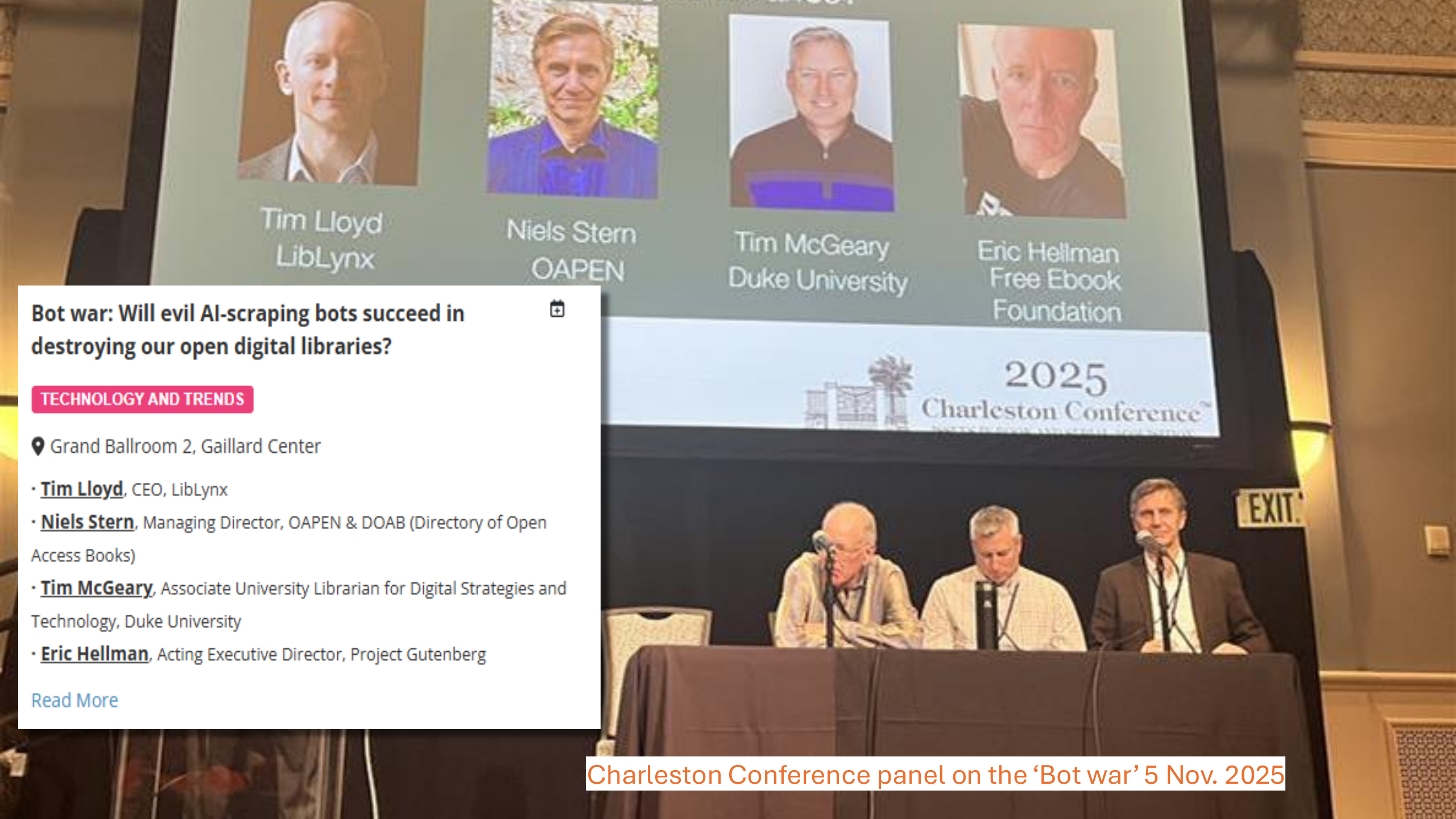
OAPEN: A world of scholarly books: Open to all, built to last.

OAPEN connects libraries, publishers, researchers, and funders with a global infrastructure for peer-reviewed open access books to advance research and the public good.

Protecting Stability and Access: How We're Safeguarding the OAPEN Library and DOAB Against Bot Traffic Surges

We want to share an important update regarding the temporary access issues that some of our community may have recently experienced on our platforms. We are fully aware of the problem and are working tirelessly to resolve it.

The Challenge: An Unprecedented Surge in Automated Traffic



Bot war: Will evil AI-scraping bots succeed in destroying our open digital libraries?

TECHNOLOGY AND TRENDS

Grand Ballroom 2, Gaillard Center

- **Tim Lloyd**, CEO, LibLynx
- **Niels Stern**, Managing Director, OAPEN & DOAB (Directory of Open Access Books)
- **Tim McGeary**, Associate University Librarian for Digital Strategies and Technology, Duke University
- **Eric Hellman**, Acting Executive Director, Project Gutenberg

Read More

Charleston Conference panel on the 'Bot war' 5 Nov. 2025

Challenges

- Aggressive AI scraping of open content
- Ignoring legal status & licences (including CC BY attribution requirements)
- Risk of undermining decades of open access efforts
- Scholarly content forced back behind paywalls?
- Is AI the perfect storm that could threaten open access?

Opportunities

- AI can enhance discoverability of peer-reviewed books
- Drive traffic and connect books with readers
- If performed responsibly, generative AI *could* support OAPEN's mission

Building trustworthy AI on Open Scholarship

- Addressing misinformation & disinformation
- Need for AI grounded in rigorously curated research
- Addressing North/South bias – and other biases

Example

- ~3,000 climate change titles in OAPEN
- Interdisciplinary (HSS + STEM)
- Potential to build reliable AI-driven knowledge resources

- For scholars, journalists, policymakers, students, wider public

oapen

The end of open access?

oapen

It depends!

Thank you!

Niels Stern, Managing Director
OAPEN Foundation

stern@oapen.org

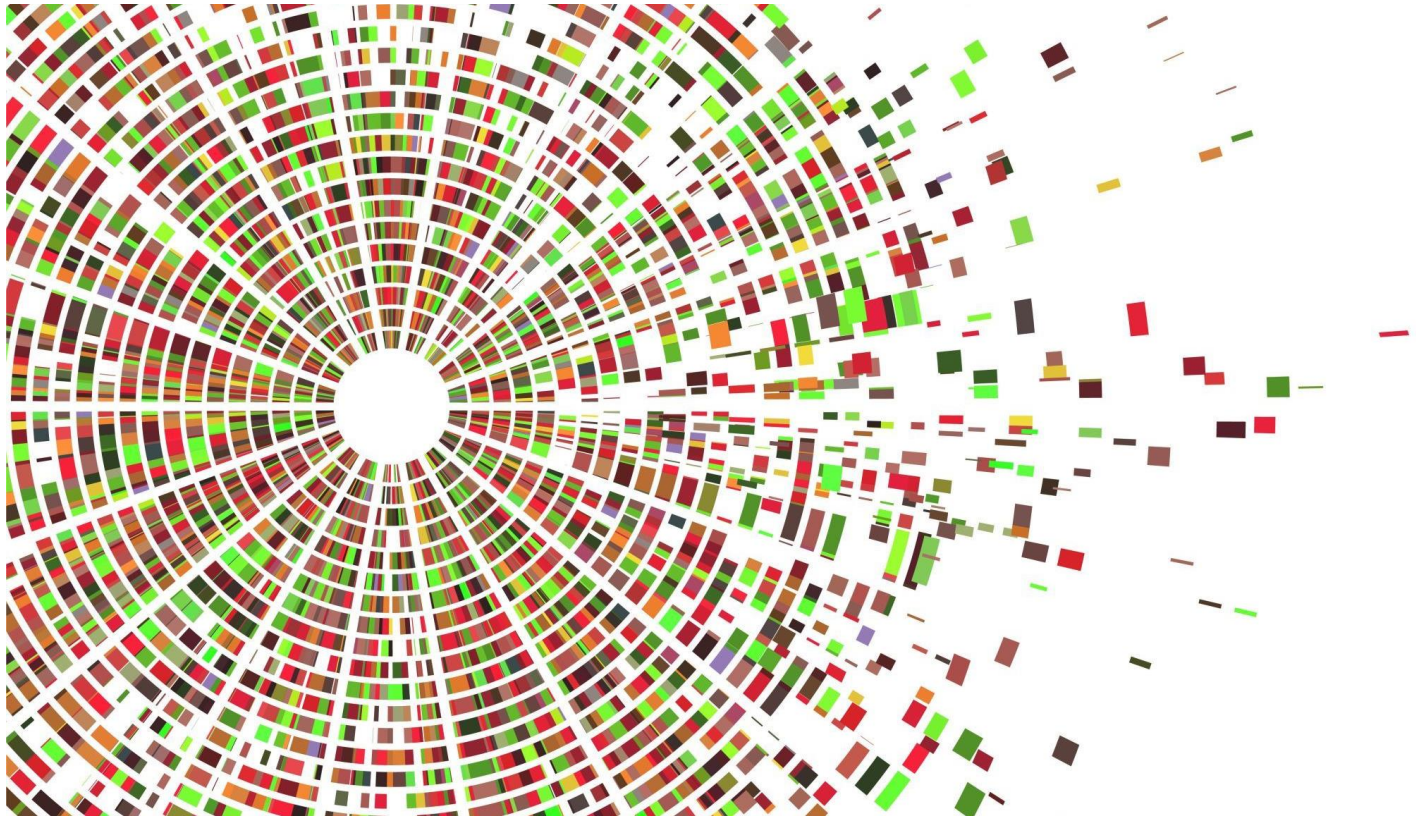
https://www.linkedin.com/in/nielsstern/

https://bsky.app/profile/nielsstern.bsky.social

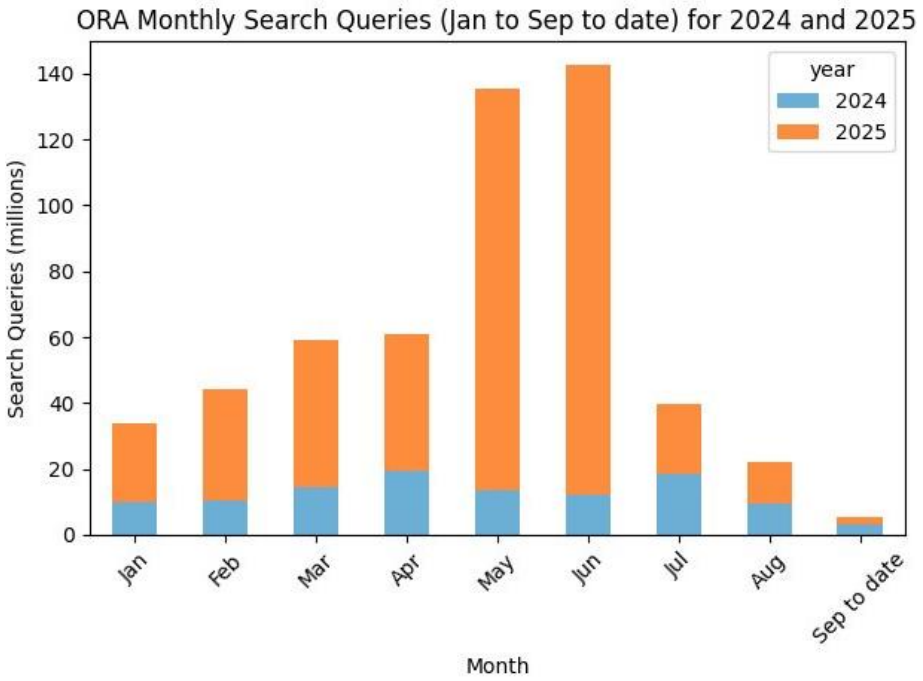
The Cost of Free

SUSTAINING ACCESS TO OPEN RESEARCH IN THE AGE OF AI

TOM WROBEL
ORA (OXFORD UNIVERSITY RESEARCH ARCHIVE)



The problem



In the last two years, open research infrastructure has faced an unsustainable increase in demand.

This demand has been unprecedented in scale. Requests for resources: search functionality, metadata, and binary files, have grown far beyond what our systems were built to handle.

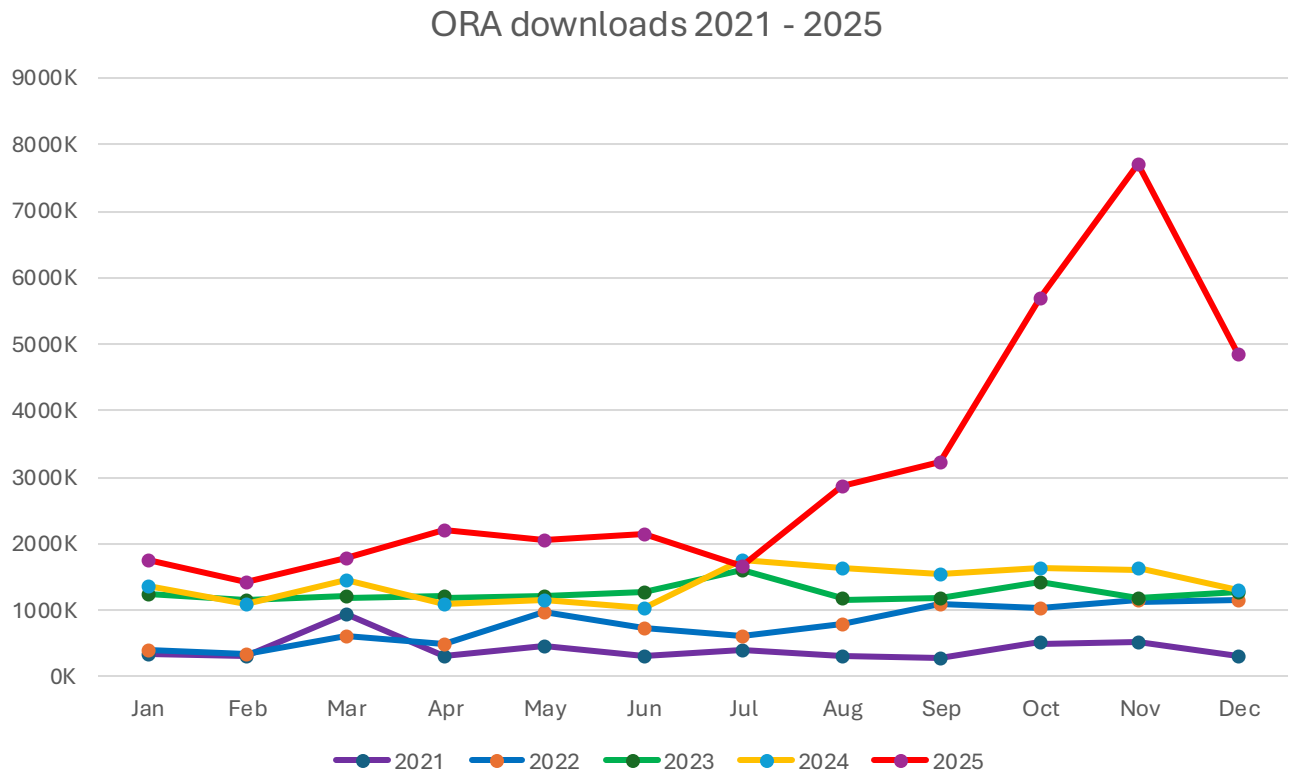
In addition, the techniques used to access our sites have evolved dramatically: the agents that are crawling our sites are able to defeat the methods we used to use to limit bot access.

Raw download requests

Year on year increases from the previous year:

- 2022: 87%
- 2023: 61%
- 2024: 10%*
- 2025: 124%

* DUE To service instability

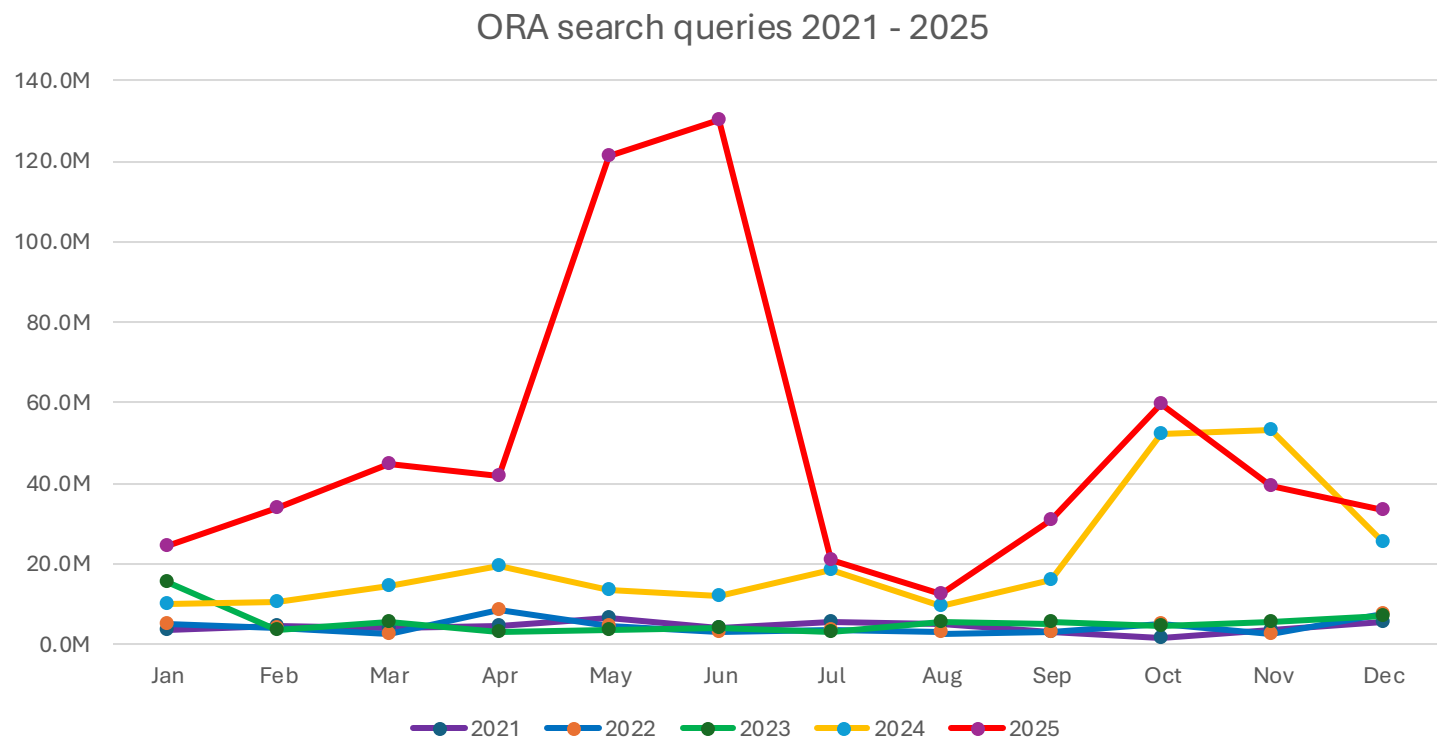


Raw search requests

Year on year increases from the previous year:

- 2022: 0%
- 2023: 29%
- 2024: 283%
- 2025: 133%*

* ORA introduced Cloudflare protection for search pages in July 2025



This is an old problem, at a new scale

- Bots have always placed more demand on our infrastructure than human users.
- Large harvesters, such as the search engine indexers for Bing or Google, have been troubling our systems for decades.
- The problem was mitigated by the fact that there were few agents motivated enough to want an - expensive – copy of all our metadata and files.
- Open Scholarship, described by human curated metadata and made available freely on open platforms, is extremely valuable training material for AI agents. This has dramatically widened the

It's not (just) who you think

The large corporate players make up only a fraction of this traffic.

ORA uses the Cloudflare service, which tells us that only 6% of our content comes from recognised AI agents such as OpenAI or Meta.

Most of our traffic is coming from anonymous harvesters.

The harvesters are using cyberattack techniques to hide their tracks

- The harvesters hide amid normal 'human' internet traffic
- They pretend to be a range of users, spreading their effort over multiple sub agents.
- Traditional methods, such as robots.txt, agent-based filtering, IP usage restrictions, are no longer effective

The tale of one harvester

- One bot, identifiable only because all the IPs were assigned from the same agency:
  - Used three data centres
  - In two continents
  - Employed sub agents spread across over 16,000 individual IP addresses
  - All using different IOS-based user agent strings
  - To make over 500,00 requests
  - In 7 minutes
  - With no one IP address making > 35 requests, none of them simultaneous

How does ORA meet this challenge?

The technological response must be policy led

ORA's decision: balance Open Access with sustainability

- The harvesters may be getting sneakier than ever, but they are trying to access our content, not trying to hurt us
- We want to remain as Open Access as possible, this includes harvesters
- We need to protect our service against excessive use
- We do not want to place restrictions on access to our metadata or our binary files
- We don't need to make human-only features available to bots

What we're doing: using managed challenges to

Restrict access to parts of the site that consume too many resources

- Bots can discover all of our content via our sitemaps.
- There are a lot of features bots don't need to see, like search result facets
- Many human-focused features, such as 'print this page', are unnecessary for harvesters
- Sometimes files are just too big for us to allow lots of them to be downloaded at once

Identify overly aggressive harvesters, and constrain them

- Put managed challenges in front of bots who are being too aggressive
- Ensure that these restrictions are reviewed to prevent excessive blocking, e.g., 'all IP addresses assigned by Vietnam Post and Telecommunications Group'
- Still allow human users within that space to access ORA content with minimal restrictions

Why this approach will not solve the problem for everyone

- It costs, both in cash and in staffing and other indirect costs
- It requires a well-resourced team who is actively engaged at identifying problem agents
- Other cultural heritage resources institutions will not have a dedicated development team and a dedicated infrastructure team to inform and implement technical responses
- Older resources that are not actively maintained are particularly vulnerable
- It's an arms race between us and the bots

What the community needs

The current demands are unsustainable for many institutions, and this may lead to research being made less open and accessible

What policy makers need to understand

- Open Access isn't free to sustain
- The demand generated by AI is beyond what we were designed for
- This demand is not going away, and the bots are only going to get more clever
- It isn't just big corporations who are doing this
- Institutions cannot handle this independently
- We need community solutions to this problem that are sustainable at scale

What are we already doing?

- Knowledge sharing and community engagement
- Experimenting with new solutions, such as new fingerprinting techniques and proof-of-work requirements
- Working on ways to help harvesters to access our content sustainably
- Spreading awareness of the technological problems

More resources

TOM WROBEL
DEVELOPMENT LEAD

ORA (OXFORD UNIVERSITY RESEARCH ARCHIVE)

THOMAS.WROBEL@BODLEIAN.OX.AC.UK

OFFICE FOR OPEN RESEARCH

BODLEIAN LIBRARIES

OxFOS

#OxFOS26

Thank you for coming

Still time to book!
go.glam.ox.ac.uk/OxFOS26

Tues, 3rd March, ONLINE, 11:00-13:00,
The future of Rights Retention to protect researchers' copyright

Tues, 3rd March, ONLINE, 15:00-16:30,
Future of open research in humanities and social sciences

Wednesday 4th March, 09:30-19:00,
In person conference day

Thurs, 5th March, ONLINE 13:00-15:00,
Cultivating FAIR data across disciplines

Fri, 6th March, ONLINE 11:00-12:00,
How to do open research safely

We'd love to hear what you thought of todays event

Join the mail list to get the recordings
Email: oxfos-subscribe@maillist.ox.ac.uk