

OPEN RESEARCH INDICATORS & PURE CRIS RECORD PARITY

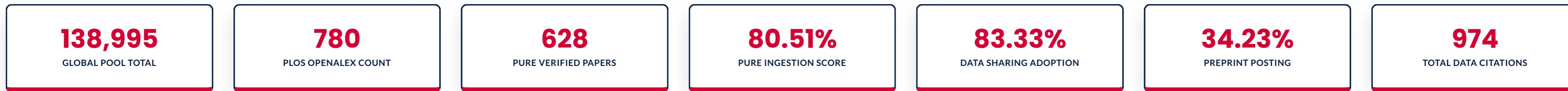
An Institutional Baseline Review of Output-Level Practices and Shared Tool Ecosystem Choice



THE UNIVERSITY of EDINBURGH

Anshita Priyadarshini Mekatotti

Open Research Data Analysis Intern | June 2026 | University of Edinburgh



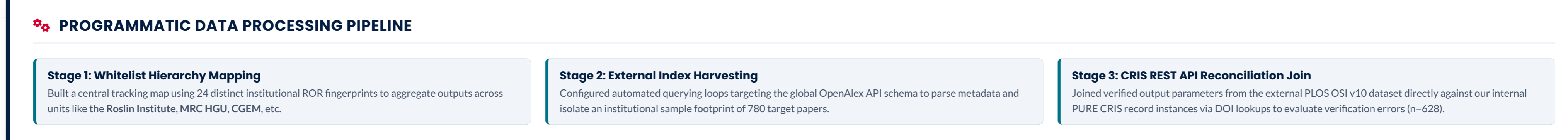
CONTEXT & RATIONALE

The Public Library of Science (PLOS) tracks global open science adoption through its Open Science Indicators (OSI) dataset in collaboration with DataSeer, an AI platform using natural language processing (NLP) and machine learning to identify data repository links within raw texts.

To move beyond self-reported feedback, I constructed this programmatic audit framework to track actual, evidence-verified repository execution. By evaluating real outputs, I cross-analyzed open behaviors directly against our internal PURE CRIS instance to locate systemic ingestion deficits.

5 OFFICIAL PLOS INDICATORS MONITORED

- Data Sharing**: Verification of raw dataset repository deposits or file attachments.
- Preprint Posting**: Locating early manuscript versions shared on public servers before indexation.
- Code Sharing**: Text-mined verification of functional computing scripts hosted on GitHub.
- Pre-Registration**: Tracing hypothesis maps registered in public hubs before collection.
- Protocol Sharing**: Verifying independent uploads of structured step-by-step wet-lab methodology steps on platforms like Protocols.io.



BENCHMARKING EDINBURGH AGAINST GLOBAL STANDARDS

OPEN SCIENCE INDICATOR	EDINBURGH RATE (%)	GLOBAL BASELINE (%)	INSTITUTIONAL ASSESSMENT
Data Sharing	83.33%	-76%	+7.3PP LEAD (STRONG ALIGNMENT)
Preprint Posting	34.23%	-21%	ELITE MULTIPLIER (1.6X GLOBAL)
Code Sharing	24.36%	-25%	PARITY BASELINE (GROWTH WINDOW)
Protocol Sharing	11.79%	-10%	ABOVE GLOBAL AVERAGE
Study Registration	9.36%	-10%	NEAR GLOBAL AVERAGE (PARITY ALIGNMENT)

Note: Weighted institutional baseline compiled from n=780 matched outputs (2018-2025) [UoE Master Audit 2026].

CORE AUDIT INSIGHTS & OPERATIONAL INTERPRETATIONS

- Funder Data Leadership**: An elite 83.33% Data Sharing Rate highlights how well regional RDM workflows successfully translate major national funder rules into practice.
- Preprint Posting Acceleration**: Our preprint rate (34.23%) yields a 1.6x global multiplier, driven by heavily embedded bioRxiv and arXiv traditions across divisions.
- Code Release Threshold**: Code sharing (24.36%) sits at global parity. Additional preparation overhead and sparse mandates identify this as an infrastructure support target.
- Protocol Sharing Emergence**: Standing above the baseline at 11.79%, protocol sharing indicates rising adoption of step-by-step methods on public platforms like Protocols.io.
- Study Registration Tracking**: Identified in 9.36% (73 papers) of the cohort. This near-parity alignment reflects the relative immaturity of study pre-registration frameworks.
- Artifact Generation Disparity**: While data generation is nearly universal at 97.05% across the cohort, an explicit behavioral gap still separates raw creation from public repository deposit.

SCHOOL-LEVEL PERFORMANCE TABLE (N=628)

VERIFIED SCHOOL / UNIT NAME	N	DATA%	CODE%	PREPRINT%	PROT%	REG%
** UNIVERSITY AVERAGE **						
Royal (Dick) School of Veterinary Studies	111	91.0%	18.9%	32.4%	15.3%	2.7%
School of Regeneration & Repair	92	81.5%	17.4%	38.0%	12.0%	28.3%
School of Population Health Sciences	92	80.4%	27.2%	32.6%	12.0%	10.9%
School of Biological Sciences	90	91.1%	33.3%	47.8%	26.7%	0.0%
School of Phil., Psych. & Lang. Sciences	27	96.3%	25.9%	33.3%	0.0%	0.0%
School of Health in Social Science (!)	17	70.6%	5.9%	11.8%	5.9%	11.8%
School of GeoSciences (!)	16	68.8%	0.0%	18.8%	0.0%	0.0%
School of Neuro. & Cardiovasc. Sciences (!)	16	75.0%	37.5%	43.8%	37.5%	0.0%
School of Social & Political Science (!)	16	75.0%	25.0%	18.8%	0.0%	12.5%
School of Informatics (!)	14	85.7%	71.4%	78.6%	7.1%	0.0%
School of Physics & Astronomy (!)	11	90.9%	18.2%	72.7%	18.2%	0.0%
Usher Institute (!)	10	40.0%	30.0%	20.0%	10.0%	40.0%

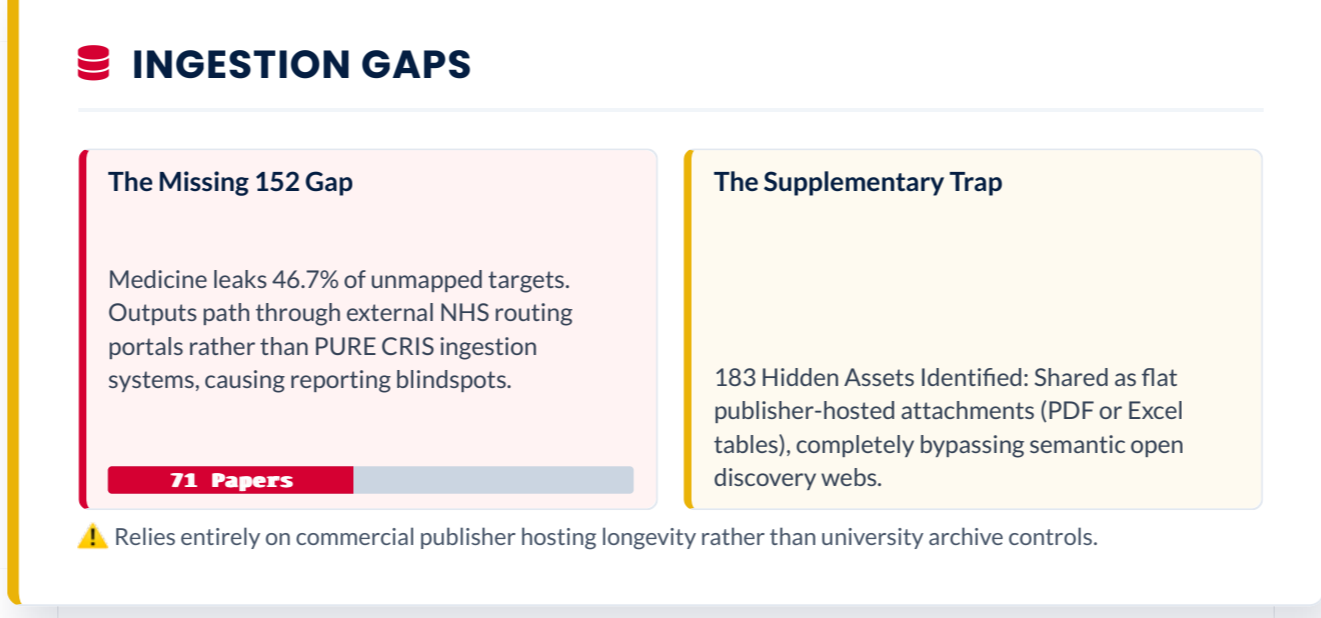
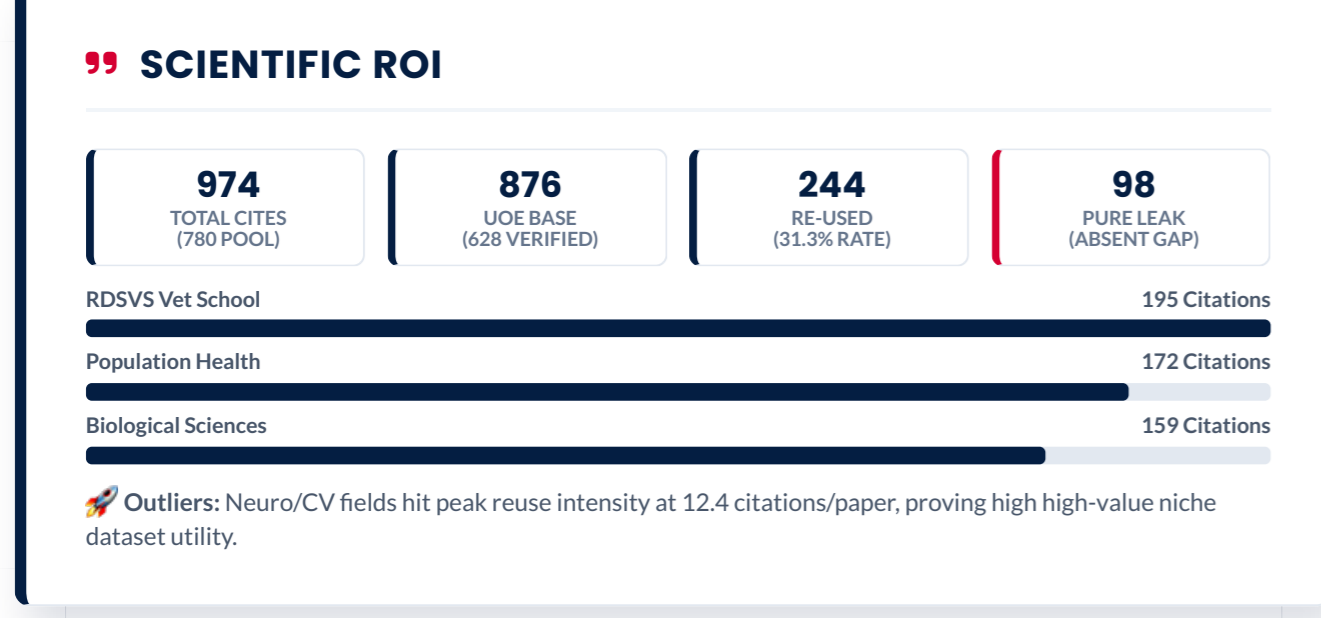
(!) Small sample (Sn < 205) — Interpret with caution. Prot% = Protocol Sharing; Reg% = Pre-Registration.

DISCIPLINARY PROFILES: EXCELLENCE & STRUCTURAL FRICTION

- School of Informatics** (Code Leader): 71.4% Code Sharing | 78.6% Preprints: Software acts as a primary research output; public GitHub code deployment functions as standard community practice.
- Usher Institute** (Governance Boundary): 40.0% Ingestion Base | 40.0% Pre-Reg: Health informatics utilizes strict NHS safe-haven patient data restrictions where open public repositories are blocked.
- Psychology / PPLS** (Data Ceiling): 96.3% Data Sharing: Driven by rigorous journal submission mandates and ethically clean, shareable experimental survey datasets.
- School of GeoSciences** (Pipeline Gap): 0.0% Code Sharing: Geospatial analysis scripts depend heavily on local file paths and proprietary environments, causing sharing friction.
- Biological Sciences** (Lifecycle Pioneer): 26.7% Protocols | 91.1% Data: Normalised use of domain repositories (GEO, Dryad) paired with wet-lab uploads via Nature/Springer Protocols.
- R(D)SVS & Roslin Institute** (Volume Anchor): 91.0% Data Sharing (n=111): Commands the highest absolute output volume while successfully maintaining elite open practices across animal sciences.

HIGHER-LEVEL EVALUATION: MACRO COLLEGE PERFORMANCE PROFILES

College	Data Sharing	Preprint Match	Study Pre-Reg	Technical Notes
CMVM (Medicine & Vet Med)	83.2% (n=339)	33.9%	13.3%	Clinical Registries: Leads in study pre-registration due to mandatory trial tracks and PROSPERO compliance loops across units.
CSE (Science & Engineering)	87.8% (n=131)	49.6%	32.1%	Technical Pioneers: Establishes peak indicators in code sharing and preprint server adoption via native, embedded arXiv workflows.
CAHSS (Arts, Humanities & Social Sciences)	83.3% (n=60)	23.3%	1.7%	Friction Boundaries: Lower marks match data confidentiality constraints; peak tracks are isolated to quantitative psychology sub-units.
Other (Independent Institutes)	80.6% (n=98)	38.8%	29.6%	Cross-Cutting Centers: Houses independent, interdisciplinary genomics pipelines and MRC units supporting cross-college safe havens.



PLATFORM ECOSYSTEMS

Evaluation of 38 tracked platforms reveals specialized choices based on local field work requirements over basic generic tools:

GitHub (Informatics Workflows)	78 Dep.
OSF Hub (CAHSS & Psych Pre-Reg)	45 Dep.
GEO / EMA (Genomics Silos)	31 Dep.
figshare (Generalist Buffers)	24 Dep.
Zenodo (Secondary Fallbacks)	22 Dep.

EXECUTIVE SUMMARY

80.51% PARITY SCORE	24.36% CODE BASE
----------------------------	-------------------------

- Informatics Excellence**: The School of Informatics leads overall institutional output vectors with an elite 71.4% code sharing rate.
- Volume Data Anchor**: Royal (Dick) Veterinary School maintains an outstanding 91.0% sharing baseline while commanding UoE's largest output volume.
- Clinical Study Maturity**: Clinical fields show high lifecycle integration, driven by a 28.3% study pre-registration track in Regeneration & Repair.
- Humanities Support Gap**: Qualitative barriers restrict CAHSS schools, causing History, Classics & Archaeology to trail with a low footprint.

AUDIT METHODOLOGICAL BOUNDARIES & SCOPE CONSIDERATIONS

- Funder & Governance Distortions**: Strict medical and clinical health data confidentiality restrictions or secure NHS patient records routinely restrict open public deposit. These constraints explain indices across healthcare fields.
- Publisher-Specific Capture Window**: Evaluation parameters are mapped explicitly against the verified target publisher subset (Sn=780\$). Practices may diverge across alternative closed access silos.
- Text-Mining Extraction Limits**: The parser loops trace text-mined data strings. Metrics indicate structured data statements; alternative repository schemas can generate minor tracking variances.

