

Appendix 6

Open Research Indicators: Report of a pilot on the monitoring the use of CRedit in research outputs



Authors

Alastair Arthur¹, Nicola Barnett², Amanda Boll³, Ann Campbell⁵, Daryl O'Connor², Sally Dalton², Harry Dimitropoulos⁶, Johanna Groothuizen⁶, Serena Mitchell⁶, Jonathon T. Newton⁶.

Research institutions: ¹University of Glasgow, ²University of Leeds, ³Newcastle University, ⁴King's College London.

Suppliers: ⁵Digital Science & ⁶OpenAIRE AMKE

CRedit Statement

Conceptualization: Sally Dalton, Jonathon T. Newton, Amanda Boll, Nicola Barnett, Serena Mitchell, Daryl O'Connor, and Ann Campbell. **Data curation:** Sally Dalton, Amanda Boll, Serena Mitchell, Harry Dimitropoulos, and Ann Campbell. **Formal analysis:** Johanna Groothuizen. **Investigation:** Sally Dalton, Amanda Boll, Nicola Barnett, Serena Mitchell, Daryl O'Connor, Harry Dimitropoulos, and Ann Campbell. **Methodology:** Sally Dalton, Johanna Groothuizen, Jonathon T. Newton, Amanda Boll, Nicola Barnett, Serena Mitchell, Alastair Arthur, Daryl O'Connor, Harry Dimitropoulos, and Ann Campbell. **Project administration:** Sally Dalton, Jonathon T. Newton, Amanda Boll, Serena Mitchell, and Daryl O'Connor. **Software:** Harry Dimitropoulos and Ann Campbell. **Validation:** Sally Dalton, Johanna Groothuizen, Amanda Boll, Nicola Barnett, Serena Mitchell, Alastair Arthur, Daryl O'Connor, Harry Dimitropoulos, and Ann Campbell. **Writing - original draft:** Sally Dalton, Serena Mitchell, Harry Dimitropoulos, and Ann Campbell. **Writing - review & editing:** Sally Dalton, Johanna Groothuizen, Jonathon T. Newton, Amanda Boll, Nicola Barnett, Serena Mitchell, Alastair Arthur, Daryl O'Connor, Harry Dimitropoulos, and Ann Campbell.

Contents

Executive Summary	1
Pilot timeline	2
Introduction	3
The prior state of monitoring this aspect of open research	4
How pilot partners support and monitor the use of CRedit	6
Methodology	7
Supplier methodologies	9
Evaluation methodology	11
Findings	12
Comment from suppliers about Findings	12
OpenAire	12
Discussion and recommendations	13
Reference list	13

Executive Summary

The UKRN CRediT Indicators Pilot Report outlines the findings and recommendations from a pilot study aimed at improving the utility of the Contributor Role Taxonomy (CRediT) in research outputs. The pilot focused on exploring how its usage could be identified and measured across various institutions and publications.

Key Findings:

Current state of CRediT usage

The pilot found that there is a lack of standardisation in the utilisation of CRediT roles in journal articles, which leads to difficulties in accurately identifying and analysing these roles. Institutional guidance or policy related to CRediT and the preparedness of Current Research Information Systems (CRIS) within research institutions, to collect and share related metadata, varied among the pilot partners.

Methodologies and data analysis

Nine partner institutions provided datasets for analysis, which included open access articles published in 2023, extracted from institutional repositories. Two suppliers, OpenAIRE and Digital Science, utilised text mining techniques to analyse the datasets. Their findings indicated varying levels of accuracy in identifying the use of CRediT, including the specific CRediT roles used, where applicable.

Challenges identified:

- Currently without standardisation of the use of CRediT, text mining can be used to identify CRediT roles, however the pilot found significant issues with sensitivity and specificity.
- Individual CRediT roles were used in contribution statements by coincidence rather than an intentional use of the taxonomy resulting in false positives.
- CRediT roles were used consistently within statements but written in a narrative format not using the exact roles or using different tenses (e.g., XX conceptualized the study).
- There is a need for improved metadata schemas to facilitate easier and more accurate recording of CRediT roles.

Recommendations:

Several recommendations have emerged from the pilot that would improve the utility of CRediT.

1. Stakeholders should collaborate to create and adopt an open, machine-readable metadata schema for recording and displaying CRediT roles. Benefits include increased usage for authors due to ease-of-use, reduced errors in identifying CRediT roles, and the opportunity for institutions to monitor usage. If automated, this monitoring would have low cost and low administrative burden.
2. Identify publishers from pilot datasets that do not use the CRediT taxonomy and engage these publishers in conversations to increase adoption.
3. Work should be undertaken to automatically match extracted CRediT roles to individual contributors. This automation could feed into individual researchers' profiles, providing a granular overview of their contributions to research.
4. Institutions should consider updating wording of documents, policies and systems to include use of 'contributor' alongside author.
5. UKRN and institutions should carry out advocacy and engagement activities to raise awareness of CRediT, increase usage and to limit metric gaming.

Pilot timeline

Date	Action
February 2024	Pilot leads meet, set up recurring lead meetings
March 2024	Partner institutions identified
April 2024	Generic dataset template agreed
May 2024	Put call out to suppliers about involvement in pilot
May 2024	Arranged initial meetings with suppliers
June 2024	Finalised suppliers involved in the pilot
June 2024	Partner institutions begin adding datasets to Figshare
September 2024	First supplier dataset provided (Elsevier)
November 2024	Second supplier dataset provided (Digital Science)
December 2024	Third supplier dataset provided (OpenAIRE)
January/February 2025	Hand checking carried out on Digital Science and OpenAire datasets
March-May 2025	Analysis of results and writing final report

Introduction

The Contributor Role Taxonomy (CRediT) is a structured vocabulary for defining a range of contributions to a published work, providing additional information beyond authorship lists. Since its introduction in 2015, journals have increasingly required authors to use CRediT within author contributor statements. More recently, research institutions such as University of Glasgow and University of Exeter have outlined expectations for researchers to use CRediT within their institutional authorship policies. In 2024 CRediT was awarded a 'Highly Commended' status in the 'Practices' category of the UK Hidden REF Competition for 2024.

For researchers, the use of CRediT may encourage and facilitate early discussions between authors regarding contributions, which may help to avoid and reduce the number of authorship disputes. It can prompt researchers to consider a broader range of activities than are often recognised as having enabled a discovery, including for example, data visualisation or project management. CRediT may support career progression; by making contributions to published work explicit, researchers and research support staff can better evidence their contribution and impact. There is also a perception that CRediT can encourage collaboration by making contributions visible and trackable, meaning that researchers can more easily find potential collaborators.

Research integrity is also a beneficiary of the CRediT approach. CRediT can enhance reproducibility and transparency by explicitly defining contributions, making it easier to trace who was responsible for specific aspects of a study. This can help improve accountability when concerns are raised with the work.

Research institutions are interested in CRediT within the context of the Contribution to Knowledge and Understanding and the People, Culture and Environment elements of REF,

as a way of systematically documenting contributions from authors as well other contributors who may not have met the authorship criteria. Using a systematic framework like CRediT ensures institutions can more easily evidence and evaluate the range of contributions researchers make and feed these into strategies to recognise broader contributions.

Given the above, this pilot has assumed that employing CRediT within an authorship or contribution statement to describe the contributions to a piece of work is a good research practice that enhances transparency and accountability of published work. A pilot to explore how CRediT usage could be identified and measured was seen to be necessary because databases, suppliers and Current Research Information Systems (CRIS) are not making metadata on CRediT readily available. The pilot intended to provide institutions with better data on the use of CRediT to enable promotion and support of the use of CRediT as good research practice. There was a perception that use of CRediT was more prevalent in certain subject areas, therefore, an additional aim of the pilot was to identify if this was true at partner institutions. This pilot is a step towards demonstrating the value to institutions of surfacing the data in a reusable form. It was also hoped the pilot could help to make the case for amending the language in internal systems and policies to include the word 'contributor' to encourage the recognition of all contributions to research beyond those of authors.

The prior state of monitoring this aspect of open research

To date, most of the available information about CRediT consists of advocating for its value, announcing its implementation (by, for instance, specific journals or publishers), or highlighting possible next steps, such as incorporation into Crossref's metadata schema. There is less literature concerning its utility in apportioning research contributions, describing methods or developing tools for analysing CRediT usage. The list below outlines the most relevant recent initiatives that we have been able to find.

Beyond authorship: Analyzing disciplinary patterns of contribution statements using the CRediT taxonomy ([González-Salmón et al., 2024](#))

This study examined 714,732 articles published between 2017 and 2024 in Elsevier and PLOS journals, in combination with bibliometric data extracted from the Scopus database. It was multidisciplinary - including Health, Life, Physical and Social Sciences. The descriptive analysis of the dataset focused on the overall coverage of the merged data, the distribution of authorship and disciplines at paper level, and the interactions between contribution statements, author order and disciplines. The main findings were:

1. Analysing contributions and authorship order can enrich the way we understand science as a social endeavour, while delving deeper into contributorship differences by field is key.
2. There are considerable differences in behaviours and accreditations per field - not just when considering individual papers, but also when considering their activities across all the articles they have contributed to. Social Sciences stands out the most here because, despite having fewer authors and contributions per publication, each author performs more of them.

A systematic scoping review of the ethics of Contributor Role Ontologies and Taxonomies (CROT) ([Hosseini et al., 2023](#))

The authors identified eight themes and 20 specific issues related to the ethics of CROT - of which CRediT is the best known example - and provided four recommendations for CROT

developers, custodians, or others seeking to use CROTs in their workflows, policy and practice:

1. Compile comprehensive instructions that explain how CROTs should be used
2. Improve the coherence of used terms
3. Translate roles in languages other than English
4. Communicate a clear vision about future development plans and be transparent about CROTs' strengths and weaknesses.

They concluded that CROTs are not the panacea for unethical attributions and should be complemented with initiatives that support social and infrastructural transformation of scholarly publications.

CRediT reliability poster (Coles, 2024)

This was a psychology research project for which researchers independently read detailed descriptions of other researchers' contributions, classified them within the CRediT framework and then cross-checked each other's work. The results indicated that the independent researchers had low agreement about both the number and type that the contributions should be classified into. Discussions indicated that vagueness of categories contributed to unreliability of classification results, the language (e.g. sentiment) that the authors used often influenced how much credit the coders granted, and many of the categories were open to overlap interpretation (e.g., supervision and administration).

Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT) (Larivière, Pontille and Sugimoto, 2021)

The research team used data from 30,054 articles from Public Library of Science (PLOS) journals over the 2017–2018 period. They analysed how research contributions are divided across research teams, focusing on the association between division of labour and number of authors, and authors' position and specific contributions. The paper concluded that the most critical tasks (writing and conceptualising) was attributed in nearly 100% of cases so is viewed as essential. However, other tasks (resources, validation, visualisation, software) were the least likely to be specifically called out for attribution. There were also some distinct gender differences in allocated tasks (indicating that male academics tend to perform more of the more 'senior' tasks on average than females, such as supervision). In addition, there are distinct patterns in multi-author works, with some tasks being performed by increasing numbers of individuals as the team grows, while others plateau at a certain point. At the same time, there are correlations between tasks, so if a contributor has performed role A, they are more likely to have performed role B as well. The authors warn, however, that 'As criteria for authorship vary considerably across disciplines so too might the interpretation of contribution roles', and CRediT in its current form does not yet protect against this opacity.

Contributions Made by Undergraduates to Research Projects: Using the CREDIT Taxonomy to Assess Undergraduate Research Experiences (CREDIT URE) (Honoré et al, 2020)

This study developed a tool called the CREDIT URE to define and measure roles performed by undergraduate students in research projects. This tool utilised the CRediT taxonomy to document and assess the specific contributions of students and was administered longitudinally across three cohorts of undergraduate student-mentor pairs involved in a biomedical research training programme for students from diverse backgrounds. Students engaged most frequently in roles involving data curation, investigation, and writing and least

frequently in roles related to software development, supervision, and funding acquisition. Their roles changed over time as they gained experience, and agreement between students and mentors about responsibility for roles was high.

The findings from this study contrast with some of the findings from other studies. It may be that this was partly because this cohort was concentrated within a particular discipline and career stage, and both the students and mentors had well matched interpretations of the various contributorship categories in comparison.

How pilot partners support and monitor the use of CRediT

The use of CRediT and the readiness of systems to collect and share metadata vary across the pilot partners. The participating Universities of the wider Open Indicators project were asked about their current policies and systems, and their responses are summarised below.

At the University of Leeds, authors can include credit roles when adding research output records to their CRIS (Symplectic). However, this information is not mandatory and is not currently surfaced in the ePrints repository. There is no process to verify contributions, and currently uptake is not actively monitored. The Leeds [Publication's Policy](#) states that 'author's must specify authors' contributions in all research outputs to ensure individuals' roles are identifiable and duly recognised'. Additionally, the [Fair Attribution Guidelines](#) emphasise that contributions from technical and specialist staff and facilities to a research output should be appropriately attributed.

At the University of Glasgow, CRediT roles can be recorded in their CRIS ([ePrints](#)), but this is not actively monitored. They use a mediated deposit service model, which allows them to verify accuracy, and they have capability to report on the use of CRediT.

At the University of Liverpool, authors can add credit roles in their CRIS (Elements), but there is no requirement for authors to include this information.

The University of Exeter has an [Attribution Policy](#) that encourages the use of the CRediT Contributor Roles Taxonomy and the inclusion of contributorship statements alongside traditional attributions.

At Newcastle University there is currently no way of recording the use of CRediT roles in the current CRIS. The University is moving to a new CRIS in 2026 (Symplectic) and intends to automate data entry as much as possible. The University is interested in knowing where CRediT data is federated and how accurate and reliable it is. The pilot data collection and evaluation are therefore very helpful for this goal.

At King's College London the Research Culture team is developing an Authorship and Contributions Policy which recommends that researchers use CRediT in all of their research outputs. The team is also discussing how CRediT can be embedded into the Pure CRIS with colleagues in Research Information Systems. They have been in contact with Research Culture counterparts in several other institutions that use this CRIS and connected with the UK Pure user group to engage Elsevier regarding amending the overall Pure functionality to incorporate CRediT roles.

Methodology

This pilot was originally conceived of as two separate pilots, on the extent of use of CRediT in research related to the institution and on how CRediT are records being populated. Several institutions from the Open Research Programme volunteered to join each pilot and within the

initial meetings, decided on merging the two groups. The pilot group was formed of the below institutions. The lead institutions lead on administrative support for the pilot.

Lead institutions	Participating institutions
King's College London	University of Liverpool
University of Leeds	University of Surrey
Newcastle University	University Exeter
University of Glasgow	University of Reading

Initially the participating institutions considered the CRediT pilot within the INORMS SCOPE Framework for Research Evaluation to reflect on the institutional context in which indicators were being developed. Following this the team then reviewed the solution provider responses which had been collected at the broader project level. The team chose to work with Elsevier, Digital Science and OpenAIRE on the basis that they could analyse a wide range of publications. The providers were then sent the below questions and invited to initial meetings to discuss the scope of the task.

1. Whether each output contains a statement of author contributions, coded as a binary 'Yes' vs 'No' for existence of statement.
 - a. Indicator of 'Discipline' based on the journal
 - b. Year of publication, e.g., publications in 2023
2. What roles are populated in each statement?
 - a. How is this spread across disciplines, identified in this pilot or through institutional data?

In line with pilots across the broader project, the team decided to use institutional journal publications from 2023 as the dataset for providers to analyse. Each participating institution extracted outputs from their repository using the following criteria:

- Must be open access
- Must be published in 2023
- Type must be article only

Institutions uploaded their data into a Figshare area, including the following Mandatory and Optional Fields:

Mandatory Fields	Description
LocalID	A local repository id number
DOI	The publisher DOI
PageLink	The link to repository splash page or similar
Type	Item type
Year	Year of publication
OrgUnit	The university faculty/college/school/department. This is the minimum viable unit that each Research Organisation can provide.
Optional Fields	Description
FileURL	A link to a repository file - this could be the Author Accepted Manuscript (AAM) or the publisher Version of Record (VOR) depending on Open Access type
FileVersion	If there is a repository file (FileURL) then an explicit value flagging the version (publisher VOR or AAM)
DAS	Local system value for Data Access Statement (DAS) prevalence
OAType	The open access type
Funding	Funder acknowledgements recorded in the local system

OpenAIRE and Digital Science used the institutional data in the Figshare folder and carried out their analysis. Elsevier provided some initial analysis of their own publications but did not participate further in the pilot.

Figshare datasets for this pilot

Institution	# of pubs provided	Digital Science returned	OpenAIRE returned
Kings College London	3424	3315	2044
University of Liverpool	7600	7323	2884
University of Exeter	1372	1334	132
University of Surrey	1349	1051	499
University of Glasgow	5127	4902	2803
Newcastle University	2740	2624	1533
University of Reading	1456	2181	737
University of Leeds	3684	3524	1936

Total	26752	26254 (24,998 after deduplication)	12568
--------------	--------------	---	--------------

Supplier methodologies

OpenAIRE Methodology for Pilot- CRediT Statement Analysis

As part of UKRN's Pilot on CRediT, OpenAIRE developed and applied a bespoke text mining workflow to identify and extract CRediT authorship contribution statements from full-text Open Access publications. The process targeted outputs from UK institutions shared through the Pilot dataset, and OpenAIRE was able to retrieve and process a substantial subset of full-text PDFs (over 12.5K) to perform the analysis.

The algorithm was designed to search for commonly used headings that precede CRediT or author contribution statements (e.g., Author Contributions, CRediT Statement, Author Contribution.). Once a heading was identified, the subsequent text was scanned for the presence of any of the 14 standard CRediT contributor roles, including spelling variants and formatting alternatives (e.g., Conceptualization / Conceptualisation, or Writing – review & editing / Writing review and editing). A statement was validated if at least one such role was detected. To account for formatting inconsistencies due to PDF-to-text conversion—such as errant punctuation, line breaks, or author initials—the script applied a series of normalization steps and robust regex patterns.

Each extracted snippet was stored in structured format and later aggregated into spreadsheets for reporting and analysis. The final dataset covered 2,344 unique DOIs with at least one matched role. A follow-up analysis was also performed to determine the relative frequency of individual CRediT roles, providing insight into how contribution types are currently reported across UK institutions.

Digital Science Methodology for Pilot- CRediT Statement Analysis

To determine whether CRediT ontology has been used within a publication, outputs exhibiting Author Contribution sections were first identified using a pilot Research Integrity Trust Marker dataset in its Dimensions database.

Trust Markers are explicit statements within a research article, such as funding acknowledgements, data availability statements, conflict of interest disclosures, ethical approval statements, and author contribution statements, that signal adherence to accepted research integrity and transparency practices. In this pilot, Trust Markers were identified through automated analysis of full-text publications to detect the presence of these statements, using content that was openly available or for which permission had been granted for text analysis.

For further information - McIntosh, Leslie D.; Whittam, Ruth; Porter, Simon; Vitale, Cynthia Hudson; Kidambi, Misha; Science, Digital (2023). Dimensions Research Integrity White Paper. Digital Science. Report. <https://doi.org/10.6084/m9.figshare.21997385.v2>

The Trust Marker dataset was developed by converting available publication PDFs into plain text, isolating relevant segments of interest, and using those segments to train and validate machine-assisted extraction models. Building on this foundation, CRediT terminology was identified through a systematic multi-step process:

Selection of publications with available full text

The analysis was restricted to publications for which full text was accessible for text mining, including openly available articles, preprints, and conference papers. Because the extraction models operate on text content, inclusion required searchable full text either through open access or permission to analyse the content.

Conversion of PDFs into machine-readable text

Publication files (typically PDFs) were converted into text strings to allow automated parsing. This conversion enabled downstream processes to treat each publication as a sequence of text rather than as binary images, making it possible to detect relevant sections algorithmically.

Identification of author contribution sections

An adapted version of a previously validated research-integrity extraction algorithm was applied to these text strings to locate sections of the text where author contributions are described e.g. “Author Contributions”

Extraction of CRediT role terms

Once contribution statements were identified, natural language processing tools were used to match text against the standardised CRediT role vocabulary, extracting instances where specific CRediT roles were explicitly stated.

Where available, CRediT roles extracted from full text were validated against publisher-supplied XML metadata containing contributor role information, allowing estimation of extraction precision and recall, while recognising that such structured metadata was not universally present.

This project made use of an early iteration of the CRediT extraction workflow described above. Subsequent refinements and improvements to this approach were developed after the completion of the study. For further detail on the underlying methodology and its continued evolution, see Allen, L., Kiermer, V., Porter, S. and Whittam, R. (2025) *A ten-year drive to credit authors for their work — and why there’s still more to do*, *Nature*, 648(8092), pp. 33–34. <https://doi.org/10.1038/d41586-025-03860-5>.

This workflow was designed to address the practical challenge that many journals and publishers do not yet expose CRediT roles as consistent, machine-readable fields even when the roles are present in the article text. The approach therefore enables systematic broad-scale analysis of CRediT adoption and content even in the absence of standardised metadata.

For the purposes of this pilot the Dimensions Research Integrity pilot Dataset was accessed through Dimensions on Google Big Query. (This data is currently only available for collaborative research activity). Publications provided by the institutions were joined to the Dimensions Research Integrity pilot dataset using the DOI and Dimensions ID within the Google Big Query Console. In addition, several additional fields were added or derived from the main Dimensions dataset.

These included:

- Number of Authors
- Field of Research (ANZSRC)
- Collaboration Type* (National, International, Institutional, Single Authorship)
- Publication Type
- Journal
- Publisher
- First Author Academic Age*
- Last Author Academic Age*

* Author academic age and collaboration type were derived from the author/affiliation detail on the publication. Academic age was determined using disambiguated author/researcher data in Dimensions and the year of the author’s first publication.

A list of 26752 2023 publications were provided by 8 institutions. Within Dimensions, 26,254 of the 26,752 submitted institution–publication records (98%) could be matched to a publication record. Because the submitted lists included duplicate DOIs, both across institutions and, in some cases, within individual institutional submissions, these matched

records corresponded, after deduplication, to 22,069 distinct DOIs, representing the underlying set of unique publications analysed.

Out of the total 22,069 distinct publications, 17,679 (80%) were present within the Dimensions Research Integrity dataset (i.e. had been analysed for 'Trust Markers' including Author Contribution Statements, Data Availability Statements etc). Firstly, a count of those with an Author Contribution Statement was determined to satisfy requirement (a) A statement of author contributions, coded as a binary 'Yes' vs 'No' for existence of statement. A null value indicated that the publication had not been analysed for a Trust Marker and was not present in the DRI dataset. Next, the data was grouped by Field of Research / Discipline. It should be noted that in order to achieve accuracy especially with regard to interdisciplinary research, the Dimensions database does not assign discipline by journal, but instead by individual article. This means articles can be assigned one or more Fields of Research. Finally, to satisfy requirement (c) Specific CRediT roles within the author contribution statement, further analysis was performed on the Author Contribution text in order to extract specific CRediT ontology, namely conceptualiz(s)ation, Data Curation etc. Each publication was flagged accordingly.

Digital Science

This pilot represented Digital Science's first attempt at analysing CRediT roles using an early version of our Author Contribution detection approach, developed as part of a pilot Research Integrity Trust Marker dataset. Of the 22069 DOIs that were identified in Dimensions, 17679 (80%) were matched and processed through the Dimensions Trust Marker dataset - providing a comparatively broad and diverse source pool.

Out of these 17679 records 1507 were identified as having used CRediT terminology.

The initial version of the algorithm performed well in identifying when no Author Contribution Statement was present (with >94% accuracy), but had more limited success in extracting CRediT roles in varied formats, detecting verbatim CRediT roles in 35% of the manually checked cases. Since the pilot, a more advanced version of the algorithm has been developed independently. When retrospectively applied to the same validation sample, detection of Author Contribution Statements improved from 60% to 84%, and CRediT role detection rose from 35% to 83%. This however came with a 9% increase in false positives, reflecting the inherent trade-off between sensitivity and specificity when working with diverse, non-standardised contribution statements. It is also important to note that "accuracy" figures reported in the pilot were based on a relatively small manual validation sample (150 DOIs), with strict inclusion criteria requiring exact CRediT terminology. This narrow definition, while useful for consistency, may underrepresent tools designed to recognise semantically equivalent or narrative expressions of author roles. Additionally, differences in the size and composition of the source pools analysed by each provider introduce further complexity. For these reasons, headline percentages should be interpreted with caution, and viewed as indicative rather than definitive assessments of overall capability.

Evaluation methodology

To assess the accuracy and usefulness of the providers results, five participating institutions (Newcastle, Leeds, Glasgow, Reading, King's) completed an evaluation exercise.

The five institutions followed a project-level "Protocol for manual checks" (Overview Report Annex 4) and also discussed evaluation plans with colleagues from CWTS. Following an initial assessment of the providers results, the team decided that for evaluation purposes, a narrow criterion should be used for establishing whether CRediT was being used: at least one CRediT role used verbatim, and therefore the pilot didn't count variations on CRediT roles.

The five institutions manually checked for the presence of contributor or authorship statements, and whether CRediT roles had been used, for a random sample of DOIs within the Digital Science and OpenAIRE datasets. The team then compared this against the algorithms' return. Through online random number generators, each institution selected 30 of its own DOIs from the respective datasets for this process (5x30, N=150 for each provider).

The following accuracy percentages were then determined:

- Contributor statement present, % correctly identified by algorithm
- Contributor statement not present, % correctly identified by algorithm
- % of total cases where algorithm was correct
- CRediT roles present, % correctly identified by algorithm
- CRediT roles not present, % correctly identified by algorithm
- % of total cases where algorithm was correct

Where a CRediT statement was present, it was also assessed whether all the specified roles aligned with CRediT, and whether the standardised format had been used, or whether the record contained, for instance:

- Some roles that aligned with CRediT, but others that did not.
- CRediT-type role descriptions, but without complying with the standardised CRediT format

These observations were added as free-text comments.

Findings

CRediT operates as a highly controlled taxonomy and if its use within authorship statements was standardised, would allow for easy identification. Currently without standardisation, text mining can be used to identify CRediT roles, but this pilot found significant issues with sensitivity and specificity, which complicated the process. The pilot found that there was a large variation in how authorship statements were written and how CRediT taxonomy was used within them.

Through the evaluation exercise, common issues that resulted in the misidentification of the presence or absence of CRediT roles were identified and grouped into the following categories:

- Individual CRediT roles were used in contribution statements by coincidence rather than an intentional use of the taxonomy resulting in a false positive
- CRediT roles were used consistently within statements but written in a narrative format not using the exact roles or using different tenses (e.g., XX conceptualized the study)
- Spelling differences (e.g., conceptualization vs conceptualisation)

- Line breaks within contribution statements meant latter sections were not picked up by text mining

Within the Digital Science set, a number of Elsevier publications were not identified as having contributor statements due to the heading “CRediT authorship contribution statement” not being identified as a contributor statement. OpenAIRE also noted that unusual formatting or more than four letters in author initials also disrupted the algorithm extracting the CRediT role. Both suppliers returned a smaller subset of the data than was provided to them in Figshare (see note above, OpenAIRE dataset was also smaller than Digital Science - figures returned included in methodology).

Accuracy checks for both Digital Science and OpenAIRE showed a high level of accuracy in correctly identifying when no contributor statement — including no CRediT statement — was present. Across both datasets, accuracy exceeded 94%, with the OpenAIRE algorithm achieving 100% accuracy in these cases and reporting no false positives.

The accuracy check for identifying a CRediT statement when present was higher in the OpenAIRE dataset than the Digital Science dataset (93.5% and 34.8% respectively). Digital Science has identified how the accuracy of the algorithm can be improved.

Comment from suppliers about Pilot Findings

OpenAIRE

We greatly value the hand-curated analysis conducted by institutional partners, which confirmed that the algorithm achieved high precision and recall in identifying CRediT statements, with no false positives. Their feedback has also highlighted useful edge cases that will inform future improvements. One such issue relates to contributor statements being prematurely truncated during extraction — usually when author names included more than four initials. This occasionally confused the sentence segmentation process used in the Natural Language Toolkit (NLTK), leading the algorithm to misinterpret where the statement ended. Although the algorithm was tuned to handle up to four initials (e.g., “C.J.A.D.”), longer or irregular initial patterns still posed a challenge. To improve this, we are exploring a custom segmentation method that avoids standard sentence tokenization, in order to more reliably extract complete contributor statements.

Digital Science

This pilot represented Digital Science’s first attempt at analysing CRediT roles using an early version of our Author Contribution detection approach, developed as part of a pilot Research Integrity Trust Marker dataset. Of the 22069 DOIs that were identified in Dimensions, 17679 (80%) were matched and processed through the Dimensions Trust Marker dataset - providing a comparatively broad and diverse source pool. Out of these 17679 records 1507 were identified as having used CRediT terminology.

The initial version of the algorithm performed well in identifying when no Author Contribution Statement was present (with >94% accuracy), but had more limited success in extracting CRediT roles in varied formats, detecting verbatim CRediT roles in 35% of the manually checked cases. Since the pilot, a more advanced version of the algorithm has been developed independently. When retrospectively applied to the same validation sample, detection of Author Contribution Statements improved from 60% to 84%, and CRediT role detection rose from 35% to 83%. This however came with a 9% increase in false positives, reflecting the inherent trade-off between sensitivity and specificity when working with diverse, non-standardised contribution statements. It is also important to note that “accuracy” figures reported in the pilot were based on a relatively small manual validation sample (150 DOIs), with strict inclusion criteria requiring exact CRediT terminology. This narrow definition, while

useful for consistency, may underrepresent tools designed to recognise semantically equivalent or narrative expressions of author roles. Additionally, differences in the size and composition of the source pools analysed by each provider introduce further complexity. For these reasons, headline percentages should be interpreted with caution and viewed as indicative rather than definitive assessments of overall capability.

Discussion and recommendations

Text mining offers real possibilities for automatic extraction of CRediT verbs (and also, subsequently, automatic feeding into research information systems like Pure, etc.). Crossref are reportedly working to automate extraction of CRediT through the use of metadata, but this heavily depends on publishers. Text mining, if done well, could be a simpler and more universally applicable alternative for now. However, based on the pilot it will need a lot of fine-tuning.

Several recommendations have emerged from the pilot that would improve the utility of CRediT, by boosting its adoption, coverage, applicability, and impact:

1. Stakeholders should collaborate to create and adopt an open, machine-readable metadata schema for recording and displaying CRediT roles. Benefits include increased usage for authors due to ease-of-use, reduced errors in identifying CRediT roles, and the opportunity for institutions to monitor usage. If automated, this monitoring would have low cost and low administrative burden.
2. Identify publishers from pilot datasets that do not use the CRediT taxonomy and engage these publishers in conversations to increase adoption.
3. Work should be undertaken to automatically match extracted CRediT roles to individual contributors. This automation could feed into individual researchers' profiles, providing a granular overview of their contributions to research.
4. Institutions should consider updating wording of documents, policies and systems to include use of 'contributor' alongside author.
5. UKRN and institutions should carry out advocacy and engagement activities to raise awareness of CRediT, increase usage and to limit metric gaming.

Reference list

- Coles, N. A. 2024. CRediT reliability poster. OSF. Available at: osf.io/ctwd7. [Accessed 25 Apr. 2025].
- González-Salmón, E., Di Césare, V., Xiao, A. & Robinson-Garcia, N., 2024. Beyond authorship: Analyzing disciplinary patterns of contribution statements using the CRediT taxonomy. Zenodo. Available at: <https://doi.org/10.5281/zenodo.14168888> [Accessed 25 Apr. 2025].
- Honoré, M., Keller, T. E., Lindwall, J., Crist, R., Bienen, L., & Zell, A. 2020. Contributions Made by Undergraduates to Research Projects: Using the CREDIT Taxonomy to Assess Undergraduate Research Experiences. *Scholarship and Practice of Undergraduate Research Journal (SPUR)*, 4(1), <https://doi.org/10.18833/spur/4/1/3> [Accessed 25 Apr. 2025].
- Hosseini, M., Gordijn, B., Wafford, Q. E., & Holmes, K. L. 2023. A systematic scoping review of the ethics of Contributor Role Ontologies and Taxonomies. *Accountability in Research*, 31(6), 678–705. <https://doi.org/10.1080/08989621.2022.2161049> [Accessed 25 Apr. 2025].

Larivière, V., Pontille, D. and Sugimoto, C.R., 2021. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, 2(1), pp.111–128. Available at: https://doi.org/10.1162/qss_a_00097 [Accessed 25 Apr. 2025].

Appendix 6: Monitoring the use of CRediT in research outputs.

Annex 1: CWTS Review

Review of the “Open Research Indicators: Report of a pilot on the monitoring the use of CRediT in research outputs”.

Rodrigo Costas. Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands

Note: Sections of the report related to Digital Science have been updated since this review was written, and in response to it – see Annex 2.

General summary of the pilot

The UKRN CRediT Indicators Pilot aimed at exploring the possibilities of developing indicators on author contributor in research outputs. The pilot effectively focused on exploring how the use of author contributor statements, particularly CRediT statements, could be identified and measured across publications and institutions. The use of CRediT in scientific publications has the potential to enhance the reproducibility and transparency of author contributions, making it easier to trace responsibilities and accountability of scientific authors.

The pilot started with publication datasets extracted from the institutional repositories of the nine partner institutions participating in the pilot. Publications had to be open access, had a DOI and be published in 2023. These datasets were shared with two data providers (OpenAIRE and Digital Science), who utilized text mining techniques to analyze the datasets, identify contributor statements and CRediT statements. The results provided by the data providers were manually checked against a sample of 150 publications extracted from the different institutions.

Results showed that both data providers were able to accurately identify the lack/presence of contributor statements, although with varying levels of accuracy in identifying the use of CRediT, and particularly the use of specific CRediT roles.

The pilot concludes the relevance of text mining approaches for identifying contribution statements and CRediT, but it recommends a broader adoption of a “machine-readable metadata schema for recording and displaying CRediT roles”. Collaboration with publishers and other stakeholders to automate and systematize the use and adoption of the CRediT taxonomy is also recommended.

Rigor of the approaches used in the pilot

In general, the approaches used in the pilot can be considered rigorous and valid. The pilot carried out a brief literature review, providing evidence of the limited scientific production on the topic, which hints to the relative infancy of this type of research, like due to the limited metadata availability and analytics on contribution statements.

The pilot created a single joint dataset of scientific publications, which was shared with the data providers for their text mining in the search for contribution statements and CRediT statements. The use of the same dataset makes the results from the data providers comparable. An independent hand check of the results by the data providers was performed by the pilot, allowing for more in-depth considerations about the most common issues in the identification of contributor statements and CRediT. The pilot also exchanged feedback with the data providers, allowing data providers to learn and even improve their algorithms.

The rigor of the methodology would benefit from clarifying certain methodological aspects. In particular, it would be useful to specify whether the datasets extracted from the pilot participant organizations were complete or randomly selected, how the 150 publications were chosen (e.g., randomly), and to provide more details about the hand check process (for example, how many people were involved, and how the criteria for the hand check were established).

Finally, the pilot states that “participating institutions considered the CRediT pilot within the INORMS SCOPE Framework for Research Evaluation to reflect on the institutional context in which indicators were being developed” but it is not clear how the framework has been incorporated into the pilot and how the results reflect or relate to institutional contexts.

Relevance of results and recommendations

Given the scarcity of studies on the presence of contribution statements and, particularly, the use of CRediT in scientific publications, the pilot can be considered as highly relevant. The findings of the pilot suggest a varying accuracy in the text mining identification of the use of CRediT and its roles in scientific publications. This strongly supports the recommendation of the pilot to work towards the adoption of open, machine-readable solutions to more easily record this information and reduce errors. This will also require collaboration strategies among stakeholders, including publishers, to develop policies and systems that can facilitate the inclusion of author contributions in scientific publications. Until such policies and systems are available, the pilot suggests that text mining techniques can be used to identify such statements, but it warns of different levels of accuracy and the presence of false positives in the results.

Recommendations to increase the technical understandability of the pilot

Here, a few issues that remain somehow unclear in the report are highlighted. Raising these issues is meant to point to potential improvements in the clarity and relevance of the future dissemination of the results provided by the pilot.

Page 8, the table reporting the distribution of the dataset is not entirely clear. For example, what is meant with “returned” for each of the data providers is not explicitly mentioned. It is implied that these are the publication sets that the two data providers (Digital Science and OpenAIRE) were able to match in their data systems. However, being clearer in this aspect would be helpful.

Related with the previous, there is also some unclarity with the total composition of the dataset. In the table in page 8, there is a total of 26,752 publications reported, but for the case of Digital Science is reported that a total of 24,998 deduplicated publications were “submitted” (this is mentioned in p. 12). The reader may assume that this is the set of distinct DOIs from the joint dataset, and the value of 26,752 corresponds to the combinations of publications and institutions, with collaborative papers among the set of institutions causing the larger number of the non-deduplicated dataset presented in the table (over the deduplicated dataset). More clarity on this aspect would be helpful.

Clarification of some of the technical expressions in the report would be beneficial. For example, “regex patterns” (I assume regular expressions), or more information on how was the “Research Integrity Trust Marker dataset” by Digital Science created would be helpful. Also, a brief explanation of what are “Trust Markers” would be useful.

A summary table of the results would be beneficial for the reader, particularly to quickly understand the results. For example, it would be important to report how many publications were found to have an author contribution statement, and how many a CRediT statement, per data provider and type of dataset (e.g. deduplicated or not). Also, tabulated values of the manual evaluation process would be beneficial.

Appendix 6: Monitoring the use of CRediT in research outputs. Annex 2: Team response to CWTS Review

The partner institutions were not able to provide a response to the review. The responses below are provided by the two solution providers: OpenAIRE and Digital Science.

OpenAIRE response to CWTS review (CRediT pilot)

Scope and responsibilities

OpenAIRE's role in this pilot was limited to the **technical development and execution of a text-mining workflow** over the publication datasets provided by the participating institutions. Dataset construction (including decisions on completeness or sampling), selection of the 150 publications used for manual validation, and the hand-checking protocol itself were all designed and carried out by the pilot institutions. OpenAIRE did not participate in the manual labelling used for evaluation, nor in the interpretation of results within institutional or policy frameworks.

Clarification of dataset coverage and “returned” publications

As explained in the revised text provided by Digital Science, the term “returned” refers to the subset of submitted publication records that could be matched and processed within each provider's infrastructure. For OpenAIRE specifically, a publication could be processed only if:

- a resolvable identifier (e.g. DOI) could be matched in the OpenAIRE Graph¹, and
- **full text (typically a PDF)** was available and retrievable for text mining.

The lower proportion of publications processed by OpenAIRE therefore reflects **full-text availability constraints**, not differences in analytical approach or algorithmic capability. While the OpenAIRE Graph indexes metadata for a very large corpus of publications, text mining requires access to the publication full text², which depends on successful aggregation from repositories or publishers. Even for Open Access articles, full text may be unavailable for technical reasons (e.g. inaccessible landing pages, repository restrictions, non-standard formats, or harvesting failures). These factors are infrastructural in nature and independent of the CRediT extraction workflow itself.

Technical terminology and text-mining approach

CWTS notes that some technical terms used in the report would benefit from clarification. In the OpenAIRE methodology, references to “regex patterns” simply denote **regular expression patterns**, used to robustly match CRediT role names in text despite variations in spelling (e.g., conceptualization/conceptualisation), punctuation, or formatting introduced during PDF-to-text conversion.

At a high level, the OpenAIRE workflow involved:

¹ The OpenAIRE Graph is OpenAIRE's scholarly knowledge graph, aggregating publication metadata, links to research outputs, and—where available—full-text content harvested from repositories and publishers. See <https://graph.openaire.eu> and <https://graph.openaire.eu/docs/graph-production-workflow/> for an overview of data coverage and enrichment workflows.

² Full-text collection in OpenAIRE is handled via a dedicated PDF aggregation service, which retrieves openly accessible content from repositories and publishers, subject to technical accessibility and licensing conditions. See <https://graph.openaire.eu/docs/graph-production-workflow/enrichment-by-mining/>.

1. Identifying likely author contribution sections based on common headings; Normalising extracted text to mitigate PDF-to-text artefacts (line breaks, punctuation, initials); and
2. Matching against the standard CRediT role vocabulary using regular expressions.

Feedback from the institutional hand-checking exercise highlighted specific edge cases (for example, unusually long author initial strings affecting sentence segmentation), which are being used to further refine extraction logic. The underlying code and methodological details can be shared if useful for transparency or reuse.

On INORMS SCOPE and summary presentation

CWTS's observation regarding the use of the INORMS SCOPE framework and its relationship to the reported results is well taken. Interpretation of pilot findings within that framework sits with the participating institutions rather than the technical suppliers.

We also agree that **summary tables** aggregating key results across providers would improve readability for non-technical audiences. Such synthesis is most appropriately handled at the overall report level, drawing on the validated outputs already produced by the pilot.

Digital Science response to CWTS review (CRediT pilot)

Page 8, the table reporting the distribution of the dataset is not entirely clear. For example, what is meant with "returned" for each of the data providers is not explicitly mentioned. It is implied 2 that these are the publication sets that the two data providers (Digital Science and OpenAIRE) were able to match in their data systems. However, being clearer in this aspect would be helpful.

This is my explanation -

Eight participating institutions each supplied a list of research publications (identified by DOI) drawn from their institutional repositories. When combined, these lists comprised 26,752 institution - publication records. Because some publications involved co-authorship across multiple participating institutions and faculties, the same DOI could appear more than once in the combined dataset.

The submitted DOIs were independently matched by both providers against the content available within their respective infrastructures. The table below reports, for each institution, the number of submitted publication records and the number that could be matched within Dimensions and OpenAIRE, respectively.

Within Dimensions, 26,254 of the 26,752 submitted institution publication records (98%) could be matched to a publication record. After deduplication across all records, this corresponded to 22069 distinct DOIs, reflecting the underlying set of unique publications analysed.

OpenAIRE matched 12,568 records (47%) from the submitted lists. This lower proportion reflects differences in infrastructure scope and data availability, in particular, reliance on the presence of openly accessible full-text content within OpenAIRE's corpus rather than differences in analytical approach or performance.

It is important to note that these figures should not be interpreted as a comparison of provider quality or capability. Instead, they reflect structural differences in data coverage, DOI resolution, and full-text availability across the two infrastructures. All subsequent analyses were conducted only on the subset of publications successfully matched within each provider's environment.

Related with the previous, there is also some unclarity with the total composition of the dataset. In the table in page 8, there is a total of 26,752 publications reported, but for the case of Digital Science is reported that a total of 24,998 deduplicated publications

were “submitted” (this is mentioned in p. 12). The reader may assume that this is the set of distinct DOIs from the joint dataset, and the value of 26,752 corresponds to the combinations of publications and institutions, with collaborative papers among the set of institutions causing the larger number of the nonduplicated dataset presented in the table (over the deduplicated dataset). More clarity on this aspect would be helpful.

Covered above.

Clarification of some of the technical expressions in the report would be beneficial. For example, “regex patterns” (I assume regular expressions), or more information on how was the “Research Integrity Trust Marker dataset” by Digital Science created would be helpful. Also, a brief explanation of what are “Trust Markers” would be useful.

We have revised the text in page 10 to address these comments in addition to updating the aggregated counts.