

The Claims Network: Collecting Research, Education, Impact, and Leadership Claims on the Decentralized Web

Patrick Hochstenbach^{1,2}[0000–0001–8390–6171],
Herbert Van de Sompel^{2,3}[0000–0002–0715–6126], and
Ruben Verborgh²[0000–0002–8596–222X]

¹ Ghent University Library, Ghent, Belgium

² IDLab, Ghent University - imec, Ghent, Belgium
`patrick.hochstenbach,ruben.verborgh@ugent.be`

³ DANS, Data Archiving and Networked Services, The Hague, The Netherlands
`hvdcomp@gmail.com`

Abstract. Common research assessment approaches rely heavily on data pertaining to traditional contributions to the research endeavor (R), including the number of journal articles published, the journals in which those articles are published, and the citations the articles or journals receive. Dissatisfaction with these approaches has led to a global call for alternatives that take a much wider range of contributions into account, including those in areas such as Education (E), societal impact (I), and academic leadership (L). While traditional research outputs are relatively easy to track, thanks to their publication in a finite set of dedicated scholarly venues, non-traditional REIL contributions are typically disseminated across the broader web. This includes outputs such as research data, software, reports, courses, live performances, interviews, and committee leadership, making their discovery and collection far from straightforward. In an attempt to devise a scalable, automated approach to tackle this “find the REIL needles in the web haystack” problem, we introduce the Claims Network approach, which collects information about REIL contributions by inviting researchers to post about them on the distributed Mastodon social media environment. Through a prototype developed around the academic Mastodon instance operated by SURF in the Netherlands, we demonstrate that the approach is feasible and can be extended to include other Mastodon instances, both academic and general-purpose.

Keywords: decentralized web · research assessment · research communication.

1 Introduction

Research assessment is an integral part of academic life. Established approaches utilize data extracted from the research communication system to gain insights

into the nature, volume, and quality of contributions to the scholarly endeavor. It provides system-wide insights and, through cumulative evaluations, also contributes to the assessment of individual researchers. In the latter realm, it faces numerous challenges, including a focus on a narrow range of contributions (i.e., paper publications), dependency on closed proprietary data sources (i.e., commercial citation databases), and use of a limited set of quantitative indicators (i.e., impact factor and h-index) [50] of which the appropriateness has been repeatedly questioned [25,26]. Alternative metrics (Altmetrics) [23] complement traditional indicators by expanding the range of data sources used for the derivation of indicators to include references to contributions on social media. However, the contributions that are taken into account remain limited (i.e., identified by a DOI), and assessing the relevance of a reference to a contribution is challenging when it is impossible to verify the creator of the reference in a trustworthy manner.

Global initiatives aimed at changing the status quo consider a broad range of activities and contributions to the scholarly endeavor, enriching the basis for researcher evaluations. Prominent examples are the Leiden Manifesto [31], the Declaration on Research Assessment (DORA) [7], the Coalition for Advancing Research Assessment (COARA) [5], and the Hong Kong Principles for assessing researchers [8]. The 2020 position paper “Room for everyone’s talent” published by Dutch public knowledge institutions and research funders (VSNU, NFWO, KNAW, NWO, and ZonMw) [22] expresses a similar desire to enhance research assessment. The position paper advocates for a modernized awards system that should reflect academic excellence across key areas, including Research (R), Education (E), Impact (I), academic Leadership (L), and patient care at university medical centers. We will refer to these types of academic excellence indicators as **REIL claims** in our paper.

The report “Open Infrastructures for Responsible Research Assessment” (OI4RRA) presents a metric for qualitatively evaluating open infrastructures. Key indicators include broadening the diversity of research outputs, linking qualitative narratives with verifiable and transparent data, enhancing standards to create an interoperable open research infrastructure, and reducing the administrative burden through automation [50]. These qualitative indicators will guide us in our research. However, effective ways to collect REIL information are still needed. Since many REIL contributions, both traditional and non-traditional, are published across the web (e.g., source code, newspaper interviews) or leave digital traces (e.g., conference programs), the core challenge is locating these REIL “needles” in the vast web “haystack.”

In this paper, we introduce the *Claims Network* approach that allows researchers to claim and organize their REIL contributions by mentioning them on social media. The approach was prototyped in the course of 2024 in a collaboration between SURF (the ICT cooperative of Dutch education and research institutions) and IMEC-IDLab (the Flemish Internet Technology and Data Science Lab).

The remainder of this paper is structured as follows: In Section 2, we examine the current state of the art to collect information about REIL contributions. In Section 3, we provide an architectural overview of the Claims Network approach, and in Section 4, we describe a prototype implementation of a Claims Network. In Section 5, we provide the results of a quantitative experiment that tests our approach to obtain metadata about REIL contributions for toots posted by researchers on Mastodon instances worldwide. In Section 6, we generalize our results and present a first draft of a Claims Network specification. In Section 7, we discuss our results and possible routes for further research and development.

2 Related Work

Collecting information about non-traditional REIL contributions is a nontrivial challenge because contributions – or traces of them – can reside anywhere on the web at large. Many services exist that monitor a finite set of dedicated scholarly venues for traditional research contributions (the R in REIL), such as journal articles. For example, CORE monitors institutional and subject repositories as well as journal publishers [6], OpenAlex monitors a wide range of data sources including DOI registration agencies and preprint repositories [9], and Web of Science monitors a selection of about 34k journals [17]. But monitoring the entire web to discover non-traditional REIL contributions remains a very significant challenge.

A prominent example is the Crossref/DataCite Event Data effort [28], which monitors a dozen web portals – including Wikipedia and social media aggregators like Reddit – for references to DOIs or their corresponding landing page URLs. The number of events aggregated by Crossref in this manner is substantial: in 2021, it was reported that 1.2 billion events were collected that mention DOIs in some way [52]. While many R contributions have DOIs, others use different PIDs or none at all [30,51]. It is particularly unusual for DOIs – or PIDs more broadly – to be assigned to E, I, and L contributions. Furthermore, reliably attributing a REIL contribution to a specific researcher is challenging, though for R contributions, discovering the ORCID(s) associated with a DOI is a possibility. However, ORCID adoption is not universal [46,57], and a registered ORCID does not confirm identity or affiliation.⁴ ORCID is also rarely used for E, I, and L contributions.

The OpenAIRE GraspOS effort [24] acknowledges the challenges related to discovering REIL contributions on the web at large. To address this, a federated cataloging infrastructure [56] is proposed, in which information pertaining to non-traditional REIL contributions is entered into distributed data nodes and subsequently aggregated into a central portal. How exactly the distributed nodes acquire their respective REIL contribution information remains unclear. Manual entry by researchers or research administrators is being considered, although it

⁴ See, for example, the two ORCID identifiers the author created without requiring any proof of identity or affiliation: <https://orcid.org/0000-0001-8390-6171> and <https://orcid.org/0009-0000-4241-5317>

would predictably place a significant burden on researchers and staff who are already resource-constrained.

The myresearch.institute experiment [47] tracked researchers in 11 web portals, including Wikipedia, GitHub, Blogger, Slideshare, and hypotheses. The experiment avoided attribution challenges by tracking unambiguous researcher identifiers in the portals and did not rely on DOIs, as E, I, and L contributions were identified by URLs, as all things on the web. The experiment had a limited scope but tracked over 17,000 contributions by 16 researchers in just 15 months. Aware that E, I, and L contributions often vanish or change over time [48,44], it enabled post hoc verification by archiving contributions in a dedicated web archive. To maximize long-term accessibility, the experiment adopted standards and best practices such as WARC [21], Memento [54], and Robust Links [45].

The Claims Network draws inspiration from both the myresearch.institute and Event Data approaches, and builds on interoperability and standards frameworks developed in our earlier work [41]. As in myresearch.institute, it detects REIL contributions on the open web by monitoring the activities of researchers whose identities are verified by trusted services. However, instead of tracking multiple web portals, the Claims Network focuses solely on Mastodon instances. Researchers themselves post toots referencing the URLs of their REIL contributions, which then trigger a bot to automatically retrieve metadata from the linked content. This collected information can be used to generate aggregated overviews of a researcher’s contributions. Similar to Event Data, the Claims Network monitors references to contributions, but with a broader scope: references can point to any REIL contribution URL, not just DOIs of traditional research outputs. Additionally, our Claims Network explores the use of the Event Notifications protocol as an interoperability layer [41].

3 The Claims Network

In 2024, SURF and IMEC-IDLab designed a Claims Network to collect traditional and non-traditional REIL contributions using Mastodon. In this network, depicted in Figure 1, a **national academic Mastodon service** provides Dutch researchers with a safe communication medium that supports informal researcher interactions focusing on science, popularizing science, and education. A researcher registered with this service can create a **REIL claim**: a toot (a message on Mastodon) that references a web resource that indicates a REIL contribution by the researcher. To make these REIL claims easy discoverable, the researcher includes a reference to a **claim bot** in each claim. Such a reference triggers a claim bot in the network to process the REIL claim. The claim bot uses a network of **summarizer services** to extract descriptive metadata from the web resources referenced in these toots. The extracted metadata is then used to update the researcher’s preferred **claims overview service**, which provides a consolidated view of all claims and allows researchers and administrators to consult, edit, and validate entries. While manual entry is possible, the Claims Network aims to automate much of this process. The extracted metadata is also

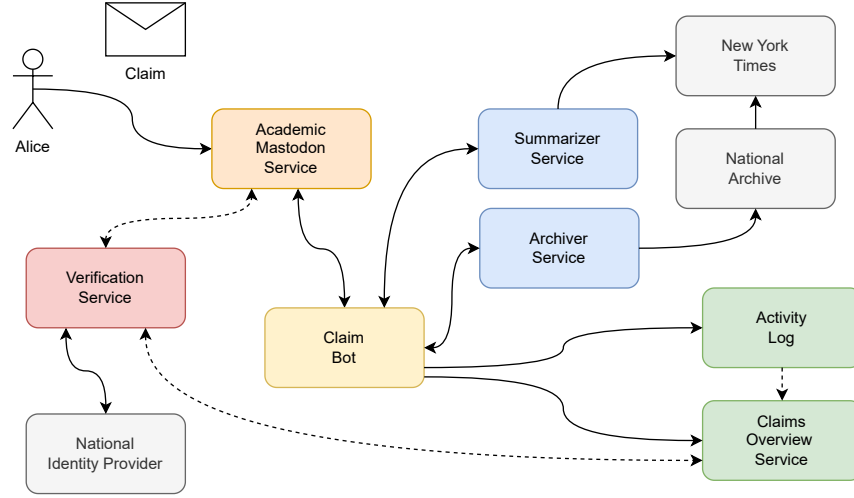


Fig. 1. The architecture of the SURF-IMEC-IDLab “Claims Network”

published as a personal **activity log** for each researcher and provides a transparent, machine-readable overview of all REIL claims made on Mastodon and their provenance in RDF format.

To unambiguously attribute REIL claims to a researcher, the Claims Network employs bidirectional links between its various services. For example, if a bidirectional link exists between a Mastodon profile and a profile on the claims overview service, this serves as a trust indicator suggesting that both profiles are controlled by the same user. This trust can be further strengthened through the implementation of a well-known **verification service**. Such a service is integrated with a national identity provider and provides researcher profiles where they can publish bidirectional links to Claims Network profiles. Verification services can also profile machine-readable metadata such as the role of the researcher in an institution (i.e., staff, student, faculty). This level of trust is similar to a verification of ownership of a website (e.g., by publishing a token on a well-known address).

To maximize long-term accessibility and versioning of web resources, an **archiver service** can be introduced in the network. This service, when provided with the URL of a web resource, generates a Memento [54] in a trusted web archive.

In addition to supporting the dissemination of REIL claims, Mastodon is envisioned as a platform for informal interactions among researchers, creating fertile ground for a new form of scholarly communication. A secondary function of the claims overview service is to provide a quality control step within the REIL claim process. While researchers can submit REIL claims, the Claims Network itself does not inherently guarantee their validity. We view this vulnerability as an integral aspect of academic trust and a core principle of good

scholarly conduct. Submitting false claims will ultimately damage a researcher’s reputation. To support transparency, the network offers services to discover REIL claims, archive the associated web resources, and maintain detailed activity logs. Within such a framework, the claims overview service acts as an additional layer of human-curated quality assurance.

4 SURF-IDLab Prototype

To test the feasibility of the Claims Network, between September 2024 and January 2025 a prototype was developed using off-the-shelf components and open-source technologies. Within the scope of the prototype, we defined the following goals:

1. Installation of an academic scholarly Mastodon service, which researchers use to disseminate claims about personal REIL contributions.
2. Installation of a claims overview service, which contains the researchers’ bibliography containing REIL claims along with their metadata and provenance.
3. Creation of a verification service to verify accounts on the claims overview service and academic scholarly Mastodon service.
4. Installation of a summarizer service that extracts descriptive metadata from a wide variety of websites, such as news agencies, blogs, and code/data repositories.
5. Creation of a claim bot service that acts as an orchestrator to listen to Mastodon toots, verify the Mastodon account using the verification service, resolve metadata about the URL posted in the toot using the summarizer service, update the claims overview service and activity log.

Academic scholarly Mastodon service We do not rely on a specific Mastodon installation in our prototype: any installation that is federating with other Mastodon instances is sufficient. Federation allows a Claims Network to have researchers and claim bots ‘live’ on different Mastodon hosts.

To attribute REIL claims to a particular researcher, bidirectional `rel="me"` links [13] are used in the various profiles of the researcher in the Claims Network, such as the claims overview service and the verification service, as discussed later in this paper.

Claims overview service To create a prototype for a claims overview service, we searched for an off-the-shelf tool that offered at least the following functionalities:

- Provide an environment where a researcher can view and edit the overview of REIL claims.
- Provide an API to add new entries to the overview of REIL claims.
- Provide a version history of claims that were added and edited in the service.
- Provide SAML [14] or OpenID Connect [12] integration with the SURF-conext [15] identity provider.

The Wiki.JS [18] software was our tool of choice for the prototype. The software runs nearly unchanged, requiring only two minor patches to improve rendering references to REIL claims in HTML pages. The GraphQL API allows the claim bot to upload directly to the claims overview service. A complete CRIS system might better serve the claims overview service in a production system. As we only created a prototype, the basic functionality of Wiki.js serves as a template for the features required in a possible CRIS system.

The Wiki.JS installation includes on each researcher’s bibliography page a similar `rel="me"` link-back mechanism for the Mastodon service to connect the identity of the claims overview service identity as part of the Dutch research community. This extra link-back can seem to be superfluous in our environment, where we can trust that a CRIS system only contains accounts for real researchers. However, in a decentralized setting, with many autonomous agents serving potentially many types of CRIS systems worldwide, trust becomes a scalability concern. Rather than trusting every potential CRIS system globally, we rely on a limited number of national verification providers.

Verification service At the time of writing, the verification service is still in development. Its purpose is to serve as a trusted endpoint within the Claims Network, allowing researchers to create a public profile page containing `rel="me"` links to their academic scholarly Mastodon account and the claims overview service of their choice. The researcher, Carol, can provide as much or as little information about herself, such as a verification profile. Typically, the verification service page mints a pseudo-anonymous URL for the researcher that does not reference a well-known personal identifier (such as an email address, login account, social account, ORCID). For example, for the researcher Carol, the profile URL <https://verify.edu.nl/profile/f7b7a81f> is minted. At the URL, an HTML profile is available, which contains the personal identifiers Carol wants to share with the world. E.g.:

- Her Mastodon identifier (in the form of a `rel="me"` link to her Mastodon profile).
- Her claims overview service identifier (in the form of a `rel="me"` link to her claims overview service profile)

This information should also be available in a machine-readable format such as JSON-LD. A prototype is available, and SURF will collaborate with GEANT implementing such a verification service with InAcademia⁵, a real-time validation service for student affiliation.

Summarizer service A summarizer service was implemented using the Zotero translation server [20], part of the Zotero project [19] that creates an open-source reference management system. Given the URL of a REIL claim, a Zotero translator service provides plugins for many types of websites to generate serialized metadata summaries using the Citation Style Language (CSL) [3].⁶

⁵ <https://inacademia.org>

⁶ At the time of writing, Zotero has translators for 727 websites.

In our prototype, we adhere to the architectural requirements to support a network of independent summarizer services that could operate within the Claims Network. To enable such a configuration, an application-neutral communication protocol is essential. The protocol of our choice was the Event Notification in Value-Adding Networks (EN protocol) [40,41], which is a profile of Linked Data Notifications (LDN) [27] and forms also the basis of the COAR Notify protocol [4]. Based on asynchronous messaging, using Event Notification messages (EN notifications), the EN protocol enables the summarizing service to queue incoming requests and process them sequentially. This batch-processing style provides an opportunity to scale services by optimizing computing resources.

A web service named “metadata-server” [34] provides an implementation of the summarizer service. The metadata-server publishes the Zotero generated summary as a new web resource and notifies the requester using an EN notification about the generated summary. Publishing the summary result as a web resource provides a public provenance trail that can be used as proof that a trustworthy summarizer was involved in the process.

Claim bot The “mastodon-bot” [33] is a simplified version of the “Orchestrator” software described by Hochstenbach [43] in 2023. Our claim bot implementation includes a Mastodon-to-EN-protocol gateway, ensuring forward compatibility for the backend orchestrator service to manage a wide network of social media platforms (e.g., including Bluesky [2] and LinkedIn [11]). The gateway monitors the claim bot account on a Mastodon instance for new ‘toot’ messages directed at the claim bot. These messages are then converted into EN notifications, which activate the claim bot. This design allows the claim bot and other services in the Claims Network to use the EN protocol as a communication bus.

The claim bot will only process an EN notification if a verified `rel=" me"` link can be found in the Mastodon user’s profile that references a trusted verification service. The verification service must provide a backlink to the Mastodon user’s profile and a link to the user’s profile on the claims overview service. The claims overview service profile should also link back to the trusted verification service. Upon successful verification, the claim bot issues an EN protocol request to a summarizer service to generate metadata for the claim. Once the metadata is prepared, the summarizer service contacts the claim bot again using an EN notification. This generated metadata is used by the claim bot to update an activity log. In the Claim Network architecture, this activity log is the input for the claims overview service. In our prototype, the claim bot can update the claims overview service directly using the GraphQL API of the WikiJS service. We are currently experimenting with providing the activity log and updates of the claims overview service via the EN protocol.

The SURF/IMEC-IDLab prototype does not include an archiver service to ensure long-term accessibility and verification of the artifacts forming the core of REIL claims. If such a service becomes available in the project’s next phase, the claim bot should send an EN notification to the archiver and wait for the response.

Activity Log To create an activity log, a web service [32,35] was developed as part of the claim bot, which stores all EN notifications sent and received by the claim bot. The content of these messages provides provenance information about REIL claims and includes a link to a summary of the claimed resources in CSL format. In our prototype, we chose to transform the CSL and the provenance information of each REIL claim into an RDF model using JSON-LD and N-Quads serialization.⁷ A numerical analysis of the generated metadata using real-world data will be presented in the next section.

A local setup of our implementation is available for download as Docker containers from GitHub [36]. The Docker demonstrator provides the service components, with the exception of the Mastodon component, which requires an external provider. However, during our experiments, we used separate installations of all software components. A demonstration video of the Claims Network in action is available at Zenodo [37].

5 SURF Experiment

To illustrate the feasibility of the SURF/IMEC-IDLab prototype, we conducted two experiments that process toots for researchers that are already active on Mastodon worldwide. The focus of our experiments is on quantitative studies concerning the generation of metadata using a Claims Network, rather than on surveying user satisfaction with Mastodon as a REIL claims platform or conducting a comparative analysis to assess the impact of a Claims Network on the real-world discovery of REIL contributions. As input for our experiments, we used websites such as “Academics on Mastodon” [1] that provide lists of researchers active on Mastodon. We choose to *pull* toots from existing Mastodon accounts as input for our quantitative study. This approach is in contrast with the Claims Network architecture as presented in section Section 4 that requires a researcher to *push* toot notifications to a claim bot. However, whether toots are pulled or pushed does not affect the dynamics of metadata generation. A push-based architecture is more scalable than a pull-based architecture for two reasons. First, it eliminates the need to enumerate all possible Mastodon accounts within a given research community (e.g., Dutch researchers). Second, it avoids the necessity of processing all toots from those accounts.

To retrieve toots from Mastodon, we developed an EN generator that converts the public timeline for the selected Mastodon accounts into EN notifications. This transformation ensures that within a Claims Network, the EN protocol is used as a communication bus. The EN generator analyzes only toots that contain references to web resources. These web resources will be treated as “REIL claims.” While this experimental design is not intended to replicate actual REIL claims realistically, it provides a pool of web resources that researchers could potentially use as examples of REIL claims. We assume that the web resources referenced

⁷ For an example of a JSON-LD version of a REIL claim, see the Claims Network specification section 5.4.1 <https://spec.knows.idlab.ugent.be/claims-network/latest/#claim-summary-format>.

by researchers on Mastodon are representative of the types of REIL claims that could exist. However, to a lesser extent, they may also reference web resources outside these categories, such as *faits divers*. The rest of the Claims Network treats this input as if genuine claims were made.

An example of a typical researcher toot from our sample is represented below:⁸

ICYMI: We recently interviewed C... L... of the Norwegian Humanist Association on the Beyond Atheism podcast. Check it out! Link here: <https://some.url.on.the.web> #atheism #humanism #secularism

In this example `https://some.url.on.the.web` would be the digital trace of the REIL claim.

We developed a specialized workflow for the Zotero summarizer service to optimize the recall of finding metadata for web resources. If access to a web resource is blocked (e.g., due to a robot block), the summarizer service attempts to retrieve an archived version of the web resource from the Internet Archive’s Wayback Machine [16] and uses the archived version to generate metadata.

Our goal was not to execute the fastest claim processing method but to analyze the range of metadata produced within a Claims Network. We processed REIL claims in batches of five requests per minute to avoid overloading our small experimental server.

5.1 Results

Experiment 1 involved 19 handpicked researchers from 11 Mastodon hosts for which 348 toots were collected that contained URLs to web resources. Most researchers (17) had enough toot messages to collect a sample of 20 toots containing a URL. Only two researchers were not active enough on the Mastodon platform to provide a maximum of 20 toots with a URL reference. Of the 348 toot+URL combinations collected, 327 claims (=94%) were unique. When feeding the 348 toot+URL combinations into the prototype Claims Network, 318 (=91%) resulted in a metadata summary that could be added to the claims overview service and in the activity log. No metadata summary was found for 30 (=9%) claims.

Experiment 2 used a random sample of 100 researchers from the “Academics on Mastodon” [1] list. These researchers were active on 42 Mastodon hosts, yielding 555 toots that contain URLs to web resources. Of the 555 toot+URL combinations collected, 531 (=96%) were unique. In our sample, 52 (=52%), researchers were not active enough to generate any claim. Feeding the toot+URL combinations into the experimental Claims Network 501 (=90%) resulted in a metadata summary that the claim bot could add to the claims overview service and the activity log. No metadata summary was found for 54 (=10%) claims.

The reasons for this 9% “no metadata summary” case in both experiments included:

⁸ We anonymized the toot for the purposes of this article.

Table 1. Metadata fields generated for REIL claims in Experiments 1 and 2.

Metadata property	Experiment 1 (N=348)	Experiment 2 (N=555)
<i>schema:title</i>	317 (99%)	497 (99%)
<i>schema:abstract</i>	214 (67%)	297 (59%)
<i>schema:language</i>	185 (58%)	275 (55%)
<i>schema:datePublished</i>	182 (57%)	266 (53%)
<i>schema:author</i>	108 (33%)	174 (35%)
<i>.....schema:familyName</i>	73	157
<i>.....schema:givenName</i>	73	157
<i>.....schema:name</i>	35	17
<i>schema:issn</i>	28 (9%)	79 (16%)
<i>schema:publisher</i>	10 (3%)	10 (2%)
<i>bibo:pages</i>	8 (3%)	51 (10%)
<i>schema:editor</i>	1 (<1%)	4 (<1%)

Table 2. Resource types for the generated metadata in Experiments 1 and 2.

Resource type	Experiment 1 (N=348)	Experiment 2 (N=555)
<i>csl:webpage</i>	200 (63%)	339 (68%)
<i>csl:post-weblog</i>	32 (10%)	40 (8%)
<i>csl:motion_picture</i>	21 (7%)	11 (2%)
<i>csl:article-newspaper</i>	30 (9%)	23 (5%)
<i>csl:article-journal</i>	15 (5%)	56 (11%)
<i>csl:entry-encyclopedia</i>	5 (2%)	5 (<1%)
<i>csl:article</i>	4 (1%)	6 (1%)
<i>csl:software</i>	4 (1%)	4 (<1%)
<i>csl:article-magazine</i>	3 (1%)	8 (2%)
<i>other</i>	34 (=10%)	9 (2%)

Table 3. Available metadata fields per resource type in Experiment 1.

Experiment 1	<i>title</i>	<i>abstract</i>	<i>language</i>	<i>author</i>	<i>datepublished</i>
<i>webpage</i>	199 (100%)	127 (64%)	131 (66%)	31 (15%)	89 (45%)
<i>post-weblog</i>	32 (100%)	29 (91%)	29 (91%)	18 (56%)	21 (66%)
<i>motion-picture</i>	21 (100%)	21 (100%)	1 (5%)	20 (95%)	19 (90%)
<i>article-newspaper</i>	30 (100%)	29 (97%)	29 (97%)	28 (93%)	30 (100%)
<i>article-journal</i>	15 (100%)	14 (93%)	11 (73%)	15 (100%)	15 (100%)

Table 4. Available metadata fields per resource type in Experiment 2.

Experiment 2	<i>title</i>	<i>abstract</i>	<i>language</i>	<i>author</i>	<i>datepublished</i>
<i>webpage</i>	335 (99%)	152 (45%)	144 (42%)	44 (13%)	118 (35%)
<i>post-weblog</i>	40 (100%)	34 (85%)	35 (88%)	25 (63%)	33 (83%)
<i>motion-picture</i>	11 (100%)	9 (82%)	0 (0%)	11 (100%)	11 (100%)
<i>article-newspaper</i>	23 (100%)	23 (100%)	23 (100%)	19 (83%)	23 (100%)
<i>article-journal</i>	56 (100%)	48 (86%)	47 (84%)	56 (100%)	56 (100%)

- Zotero does not support the format of the claimed web resource. (e.g., PDF files) (=50% of the cases).
- Zotero cannot access the claimed web resource, and the summary service can not find a copy in the Wayback Machine (=18% of the cases).
- The attempt to summarize the web resource resulted in an internal server error in the Zotero service: (=21% of the cases).
- The attempt to summarize the web resource resulted in an empty response (=11% of the cases).

We ran both experiments on a virtual Linux server with 2 x 2.80GHz CPUs and 2GB RAM. Experiment 1 required 71 minutes to process 348 toot+URL combinations. Experiment 2 required 113 minutes to process 555 toot+URL combinations. This processing speed is equal to the five requests per minute we provided to the Claims Network. The input data, results, and statistics of both experiments can be consulted at Zenodo [37].

Table 1 provides an overview of the metadata fields that were generated for the 348 claims of experiment 1 and the 555 claims of experiment 2. Metadata properties provide an overview of the metadata fields as they are mapped to RDF in the activity log. Because in a Claims Network, the identity of the researcher is known, at least a basic bibliographic reference can be created for each REIL claim, including the researcher’s identity, title, and URL of the claim.

Table 2 provides an overview of the resource types reported by the Zotero summarizer service for the claims. Table 3 and Table 4 present the metadata fields per resource type for Experiments 1 and 2. The results are particularly interesting for web resources classified as **article-newspaper**, which have an almost perfect score for identifying key metadata fields, including title, abstract, language, author, and publication data information. Newspaper resources will be the focus of a qualitative survey conducted in 2025 with a test group of researchers from Dutch universities of applied sciences.

6 Towards a Specification

Our prototype provides a proof of concept for a Claims Network *implementation*. On a practical level, the next goal is to scale out the proof of concept into a viable platform for the Dutch use case. However, from a strategic perspective, the more compelling option is to seek additional European partners who can be engaged to implement similar components of a Claims Network. Given its decentralized nature, the Claims Network does not rely on one central application platform or implementation. Achieving a broader adoption requires further generalization and a specification of the Claims Network as an *open architecture*, in line with the strategic goals outlined in the OI4RRA report [50]. The start of such an open architecture is outlined in the “Claims Network” specification [39].

The Claims Network specification generalizes the communication patterns between components within a Claims Network, aiming to maximize interoperability. This generalization, and the resulting interoperability, applies both vertically and horizontally. Vertically, our generalization eliminates local implementation dependencies from communication protocols and accommodates multiple implementations of the same functionality. For example, the summarizer service in our prototype is implemented using Zotero. However, alternative implementations of such summarizers could be developed using AI systems, aggregators, or similar technologies. Horizontally, the communication patterns for each service within the Claims Network can adopt the same one-way and request-response communication patterns as specified by the underlying EN protocol. The communication patterns to request a web resource summary implementation can also be used to request a memento copy from an archiver implementation.

One consideration that requires further attention is the source of REIL claims. In our prototype, Mastodon was used as a valid option within the Dutch SURF context as the trigger for REIL claims. However, the Claims Network architecture is not dependent on any specific social networking platform, provided that a gateway can be established to translate REIL claim triggers into EN notifications used within the Claims Network. The Mastodon-to-EN gateway in the prototype provided a starting point for such an abstraction. However, other gateways, such as a Bluesky-to-EN, LinkedIn-to-EN gateway, can be envisioned.

The verification service plays a crucial role in maintaining trust within the Claims Network. A careful balance must be maintained between the need to validate social media personas as part of a (national) research community and the protection of researchers’ true identities within the network. In our experiments, two directions were investigated: a verification service that acts as a public Web profile (containing many bidirectional links to online identities) and a verification service that acts as a type of proof where a membership to a community can be demonstrated without conveying any information beyond the mere fact of membership. The former case is currently specified in the Claims Network architecture; however, our choices are currently being debated within the SURF community, in light of further implementations using pan-European identity providers – such as InAcademia – which tend to favor the latter approach.

An open topic in our specification work concerns the design of the activity log, which provides machine-level transparency through an append-only resource of activities within the network. In the current specification, individual claims are published as **claim summaries**. Each claim summary is represented as an RDF resource utilizing the schema.org ontology, describing the metadata of the claimed resource, the initiator of the REIL claim (i.e., the researcher), and provenance information regarding the origin of the claim. In our prototype, we are currently experimenting with Linked Data Event Streams (LDES) [29] as a means to generalize activity logging. However, the specification of an activity log is still in an experimental stage, targeting Claims Network use cases. A survey of broader use cases is necessary to enable the generalization of this work. The authors have previously advocated for a logging framework to enhance transparency in the scholarly publication process [42,38]. These considerations can be incorporated into the Claims Network specification.

The objective of the Claims Network specification, as an open architecture, is not confined to specific use cases related to the discovery of REIL contributions. Each service should also function as an *independent* component, thereby eliminating tightly coupled interdependencies. The decision to use the EN protocol aligns well with the COAR Notify protocol [4], which is based on the same principles and offers an alternative approach to scholarly peer review and overlay journal publishing. The same approach is adopted by the NDE-Usable program of the Dutch Digital Heritage Network, which aims to redesign the exchange of data between cultural heritage institutions and applications [55].

This independence among components guides the design of decentralized networks. Each component should be valuable even when separated from the network. Thus, the claim bot acts more as a choreographer of loosely coupled services than as a rigid orchestrator essential for network function. While directly referencing a claim bot in a toot can accelerate the REIL discovery process, the core idea of a Claim Network could, in principle, be replicated simply by tagging a Mastodon toot as a REIL claim and employing a tag aggregation mechanism to collect those claims. What remains essential to a Claim Network is the transparency with which it operates.

7 Discussion and Conclusion

The Claims Network offers an open architecture and an initial implementation of a decentralized network for “finding the REIL needles in the web haystack” problem. The architecture is based on academic social networking services, such as SURF Mastodon, which offer researchers a secure communication platform to facilitate informal interactions, focusing on science, science popularization, and education. Some of these social interactions also serve as triggers for Claims Network agents – including the claim bot, summarizer service, activity log, and claims overview service – to generate metadata concerning REIL claims mentioned on Mastodon.

Our prototype demonstrates that many types of REIL contributions – including both traditional and non-traditional resources, many of which fall outside the scope of conventional PID systems – can be effectively discovered with minimal manual intervention required from researchers. The metadata generation proved particularly successful for non-traditional REIL contributions, such as news media, which will serve as input for further analysis in qualitative studies to be conducted in 2025 with Dutch universities of applied sciences.

The prototype relies on web scraping. Although the initial results are very promising, it is too early to declare success. On the positive side, many websites expose metadata for search engine optimization (SEO) purposes. A claim bot can collect information about REIL contributions at a very early stage, as soon as they are published and attract the attention of researchers on social media. This also provides the network with the opportunity to generate a trusted copy of the REIL artifact, preserved in a web archive. On the negative side, the emergence of generative AI and their often overzealous data-collecting activities present new challenges [49]. We foresee that many websites will take precautions to prevent robots from accessing any information. In future work, we will investigate the state-of-the-art in AI control. This presents a challenge for the web at large [10].

Further analysis is required to assess the quality and accuracy of the REIL claims. Adopting a generic solution based on web standards, instead of a web scraping implementation, would be a more promising approach. Promoting Signposting [53] could encourage its adoption in institutional and data repositories, archives, software repositories, and platforms that provide evidence of REIL claims. Automating metadata generation does not automate accuracy. Claims overview services still need manual intervention to assess the accuracy of the claims. The Claims Network facilitates the discovery of non-traditional REIL contributions and provides transparency throughout this process through activity logs. A national CRIS system could utilize these activity logs as input and provide value-added services, such as a published, trusted version of these logs.

The setup of the SURF Mastodon instance revealed several challenges concerning the trustworthiness of user profiles, including anonymous accounts and researchers with multiple accounts on multiple Mastodon instances. These challenges provided valuable input for our discussion about the verification service component, which could offer trust indicators across multiple Mastodon instances. Future research should explore the requirements for integrating trusted social networking services into a Claims Network. These challenges are not limited to Claims Network use cases. Interest in a general verification mechanism for social media accounts has led to a first cooperation with GEANT and In-Academia to provide such services.

The Claims Network specification can be viewed as an initial iteration toward a decentralized approach to collecting REIL information created by the scholarly community using an open architecture.

Acknowledgments. This research was conducted with support from SURF. The computational resources and assistance by Laurents Sesink, Thomas van Himbergen,

Niels van Dijk, and Thijs Kinkhorst were instrumental in conducting the experiments for this study.

References

1. Academics on Mastodon, <https://nathanlesage.github.io/academics-on-mastodon/>, Accessed 24 Februari 2025.
2. Bluesky, <https://bsky.app>, Accessed 20 February 2025.
3. Citation Style Language, <https://citationstyles.org/>, Accessed 20 February 2025.
4. COAR Notify Protocol, <https://coar-notify.net/specification/1.0.1/>, Accessed 12 February 2025.
5. CoARA – Coalition for Advancing Research Assessment, <https://coara.eu/>, Accessed 10 February 2025.
6. CORE data providers search, <https://core.ac.uk/data-providers/>, Accessed 23 April 2025.
7. Declaration on Research Assessment, <https://sfdora.org/>, Accessed 10 February 2025.
8. Hong Kong Principles, <https://www.wcrif.org/hong-kong-principles>, Accessed 22 April 2025.
9. How does OpenAlex work?, <https://help.openalex.org/hc/en-us/articles/28932712154391-How-does-OpenAlex-work>, Accessed 24 April 2025.
10. IAB workshop on AI-CONTROL (aicontrolws), <https://datatracker.ietf.org/group/aicontrolws/materials/>, Accessed 29 December 2024.
11. LinkedIn, <https://www.linkedin.com/>, Accessed 20 February 2025.
12. OpenID, <https://en.wikipedia.org/wiki/OpenID>, Accessed 16 June 2025.
13. rel="me" - microformats wiki, <https://microformats.org/wiki/rel-me>, Accessed 16 June 2025.
14. Security Assertion Markup Language, https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language, Accessed 16 June 2025.
15. SURFconext | SURF.nl, <https://www.surf.nl/en/services/identity-access-management/surfconext>, Accessed 16 June 2025.
16. Wayback Machine, <https://webcf.waybackmachine.org/>, Accessed 24 February 2025.
17. Web of Science | Clarivate, <https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/>, Accessed 23 April 2025.
18. Wiki.js, <https://js.wiki/>, Accessed 12 February 2025.
19. Zotero | Your personal research assistant, <https://www.zotero.org/>, Accessed 20 February 2025.
20. zotero/translation-server, <https://github.com/zotero/translation-server>, accessed 12 February 2025.
21. The WARC file format (ISO 28500) - information, maintenance, drafts (2017), <https://bibnum.bnf.fr/WARC/>, Accessed 17 April 2025.
22. Room for everyone's talent (2019), <https://recognitionrewards.nl/wp-content/uploads/2020/12/position-paper-room-for-everyones-talent.pdf>, Accessed 29 January 2025.
23. Altmetrics (2025), <https://en.wikipedia.org/wiki/Altmetrics>, Accessed 22 April 2025.

24. Administrator: GraspOS, <https://www.openaire.eu/graspos-project>, Accessed 23 April 2025.
25. Bollen, J., Sompel, H.V.d., Hagberg, A., Chute, R.: A principal component analysis of 39 scientific impact measures. *PLOS ONE* **4**(6), e6022 (2009). <https://doi.org/10.1371/journal.pone.0006022>
26. Brembs, B.: Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience* **12** (2018). <https://doi.org/10.3389/fnhum.2018.00037>
27. Capadisli, S., Guy, A.: Linked Data Notifications (2017), <https://www.w3.org/TR/1dn/>, Accessed 5 February 2025.
28. Clark, R.: Event data, <https://www.crossref.org/services/event-data/>, Accessed 10 March 2025.
29. Colpaert, P.: Linked Data Event Streams (2025), <https://semiceu.github.io/LinkedDataEventStreams/>, Accessed 16 June 2025.
30. Culbert, J., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., Mayr, P.: Reference coverage analysis of OpenAlex compared to web of science and scopus. <https://doi.org/10.48550/arXiv.2401.16359>
31. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., Rafols, I.: Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **520**(7548), 429–431 (2015). <https://doi.org/10.1038/520429a>
32. Hochstenbach, P.: eventlog-server: A tiny event log demonstration server, <https://github.com/MellonScholarlyCommunication/eventlog-server>, Accessed 16 June 2025.
33. Hochstenbach, P.: mastodon-bot: A Mastodon Service Bot, <https://github.com/MellonScholarlyCommunication/mastodon-bot>, Accessed 16 June 2025.
34. Hochstenbach, P.: metadata-server, <https://github.com/MellonScholarlyCommunication/metadata-server>, Accessed 16 June 2025.
35. Hochstenbach, P.: rdf-generator: An RDF generator for CSL (citation style language) files of event notifications containing links to CSL., <https://github.com/MellonScholarlyCommunication/rdf-generator>, Accessed 16 June 2025.
36. Hochstenbach, P.: Claims network (2025), <https://github.com/MellonScholarlyCommunication/claims-network>, Accessed 16 June 2025.
37. Hochstenbach, P.: The claims network: Collecting research, education, impact and leadership claims on the decentralized web (2025). <https://doi.org/10.5281/zenodo.15296698>, <https://zenodo.org/records/15296698>
38. Hochstenbach, P., Klein, M., Van de Sompel, H., Verborgh, R.: Transparent and trustworthy artifact life cycle. In: *iPRES 2024* (2024). <https://doi.org/10.21428/5676bf2d.a76a90cf>
39. Hochstenbach, P., Van de Sompel, H.: Claims network (2025), <https://spec.knows.idlab.ugent.be/claims-network/latest/>, Accessed 12 February 2025.
40. Hochstenbach, P., Van de Sompel, H., Vander Sande, M., Dedeker, R., Verborgh, R.: Event notifications in value-adding networks. In: Silvello, G., Corcho, O., Manghi, P., Di Nunzio, G.M., Golub, K., Ferro, N., Poggi, A. (eds.) *Linking Theory and Practice of Digital Libraries*. pp. 133–146. *Lecture Notes in Computer Science*, Springer International Publishing. https://doi.org/10.1007/978-3-031-16802-4_11
41. Hochstenbach, P., Van de Sompel, H., Vander Sande, M., Dedeker, R., Verborgh, R.: Event notifications in value-adding networks. In: Silvello, G., Corcho, O., Manghi, P., Di Nunzio, G.M., Golub, K., Ferro, N., Poggi, A. (eds.) *Linking*

- Theory and Practice of Digital Libraries. pp. 133–146. Lecture Notes in Computer Science, Springer International Publishing (2022). https://doi.org/10.1007/978-3-031-16802-4_11
42. Hochstenbach, P., Verborgh, R., Van de Sompel, H.: Event Notifications and Event Logs: Transparent Sharing of Artifact Life Cycle Data. *International Journal of Digital Curation* **18**(1) (2024)
 43. Hochstenbach, P., Verborgh, R., Sompel, H.V.d.: Using Event Notifications, Solid and Orchestration for Decentralizing and Decoupling Scholarly Communication. *The Code4Lib Journal* (58), <https://journal.code4lib.org/articles/17823>
 44. Jones, S.M., Van de Sompel, H., Shankar, H., Klein, M., Tobin, R., Grover, C.: Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content. *PLOS ONE* **11**(12), e0167475 (2016). <https://doi.org/10.1371/journal.pone.0167475>
 45. Klein, M., Jones, S.M., Shankar, H., Wincewicz, R., Nelson, M.L., Van de Sompel, H.: Robustifying Links (2020), <https://robustlinks.mementoweb.org/spec/>, Accessed 12 March 2025.
 46. Klein, M., Van de Sompel, H.: Discovering Scholarly Orphans Using ORCID. In: 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). pp. 1–10. <https://doi.org/10.1109/JCDL.2017.7991573>, <https://ieeexplore.ieee.org/abstract/document/7991573>
 47. Klein, M., Van de Sompel, H.: My Research Institute, <https://myresearch.institute/>, Archived through <https://archive.is/hcXw7> . Accessed 12 February 2025.
 48. Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., Tobin, R.: Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLOS ONE* **9**(12), e115253. <https://doi.org/10.1371/journal.pone.0115253>
 49. Longpre, S.: AI crawler wars threaten to make the web more closed for everyone (2025), <https://www.technologyreview.com/2025/02/11/1111518/ai-crawler-wars-closed-web/>, Accessed 13 February 2025.
 50. Manola, N., Tzouganatou, A., Kuchma, I., OI4RRA, C.W.o.: Open infrastructures for responsible research assessment: Principles and framework (2025), <https://zenodo.org/records/14844582>
 51. Porter, S.J.: Measuring Research Information Citizenship Across ORCID Practice. *Frontiers in Research Metrics and Analytics* **7** (2022). <https://doi.org/10.3389/frma.2022.779097>, publisher: Frontiers
 52. Rittman, M.: Event Data now with added references (2021), <https://www.crossref.org/blog/event-data-now-with-added-references/>, Accessed 16 March 2025.
 53. Van de Sompel, H., Klein, M., Jones, S., Nelson, M.L., Warner, S., Devaraju, A., Huber, R., Steinhoff, W., Tykhonov, V., Boruta, L., Meijers, E., Soiland-Reyes, S., Wilkinson, M.: FAIR Signposting Profile, <https://signposting.org/FAIR/>, Accessed 5 February 2025.
 54. Van de Sompel, H., Nelson, M., Sanderson, R.: HTTP Framework for Time-Based Access to Resource States – Memento. <https://doi.org/10.17487/RFC7089>, <https://datatracker.ietf.org/doc/rfc7089>
 55. Vander Sande, M.: Use Cases of Event Notifications in digital heritage networks (2023), https://netwerkdigitaalerfgoed.nl/wp-content/uploads/2023/07/20230614-Miel-Vander-Sande_PLDN-Solid-Use-Cases.pdf, Accessed 9 October 2023.

56. Vergoulis, T., Chatzopoulos, S., Malaguarnera, G., Mannocci, A., Dumouchel, S., Tatum, C., Peroni, S.: GraspOS Deliverable D4.2: Infrastructure Architecture (2024), <https://zenodo.org/records/14162659>
57. Youtie, J., Carley, S., Porter, A.L., Shapira, P.: Tracking researchers and their outputs: new insights from ORCIDs **113**(1), 437–453. <https://doi.org/10.1007/s11192-017-2473-0>