

New Open Science Contribution?

Toot it to Claim it!



claimbot



mastodon



summarizer



Carol

Herbert Van de Sompel

[@hvdsomp@w3c.social](mailto:hvdsomp@w3c.social) - <https://hvdsomp.info>



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

Claim It Project – UGent/IMEC - SURF



Thomas van Himbergen
Open Science Project manager
@ SURF



Patrick Hochstenbach
PhD candidate
@ UGent-IMEC IDLab



Laurents Sesink
Open Science Lead
@ SURF



Herbert Van de Sompel
Honorary Research Fellow
@ DANS



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

SURF Research Day Presentation & Live Demo

New Open Science Contribution? Toot It to Claim It! ☆

2025-05-20 11:30–12:15, On Air

Are you a researcher who wants to share an overview of your open science contributions, including papers, blogs, interviews, podcasts, teaching activities? Have you noticed that existing solutions only cover a subset? Have you no time to manually add the rest? If so, come test a prototype that lets you toot (Mastodon) the URL of your contribution to automatically add it to a trusted profile page.

A laptop is not required, a device is useful in order to participate

Research assessment approaches evolve towards evaluating your performance on the basis of a wide range of activities rather than merely relying on the traditional impact factor and citation count metrics. For this, existing efforts focus on the low hanging fruit, typically tracking contributions that have a DOI. But what about the rest? What about blogs, newspaper articles, podcasts, conference attendances, teaching activities? Obviously, as a researcher, you know when you made a new contribution. But manually adding it to your profile page is not realistic because you contribute so much. How about a solution that allows you to toot the URL of a contribution to automatically add it to your profile page? And, while you are at it, restart the informal scholarly conversation, lost when Twitter became X, on Mastodon?

If that all sounds appealing, join this session that introduces SURF's "Toot It to Claim It" prototype:

- Get an insight in the prototype's overall set-up and future plans
- Be among the first to compile a profile page by tooting about open science contributions
- Challenge the prototype by tooting different types of contributions and detect for which cases automation causes problems
- Learn about the prototype's success rate in controlled experiments
- Learn about aspects of trust provided by the prototype, including reliable attribution of toots to researchers, public logging of the claim process, archiving contribution snapshots, and profile page versioning



Thomas van Himbergen

Adviseur and coordinator at SURF



Patrick Hochstenbach

My professional career spans more than 25 years in academic libraries in Belgium (Ghent), USA (Los Alamos) and Sweden (Lund). I was always involved in library automation or library science related projects as software developer, architect and more generally innovator. Since 2021 I am doing a PhD in computer science and UGent-IMEC IDLab on scholarly communication on the decentralized web.

For more information: <https://patrickhochstenbach.net>

SURF Research Day, May 20 2025, Gooiland, Hilversum
<https://www.surf.nl/en/agenda/surf-research-day-2025>

(Hopefully) Forthcoming Paper

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
`hvdsomp@gmail.com`



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025





In February 2024, SURF organized a brainstorm to discuss potential innovative activities in the realm of scholarly communication and the distributed web.



@hvdcomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025



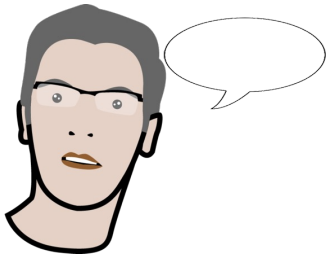
Claim It Project Context – SURF Meeting 02/2024

- Explore areas in which potential projects related to scholarly communication could be launched that embrace core principles:
 - Societal values
 - Decentralized web
 - Digital sovereignty



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

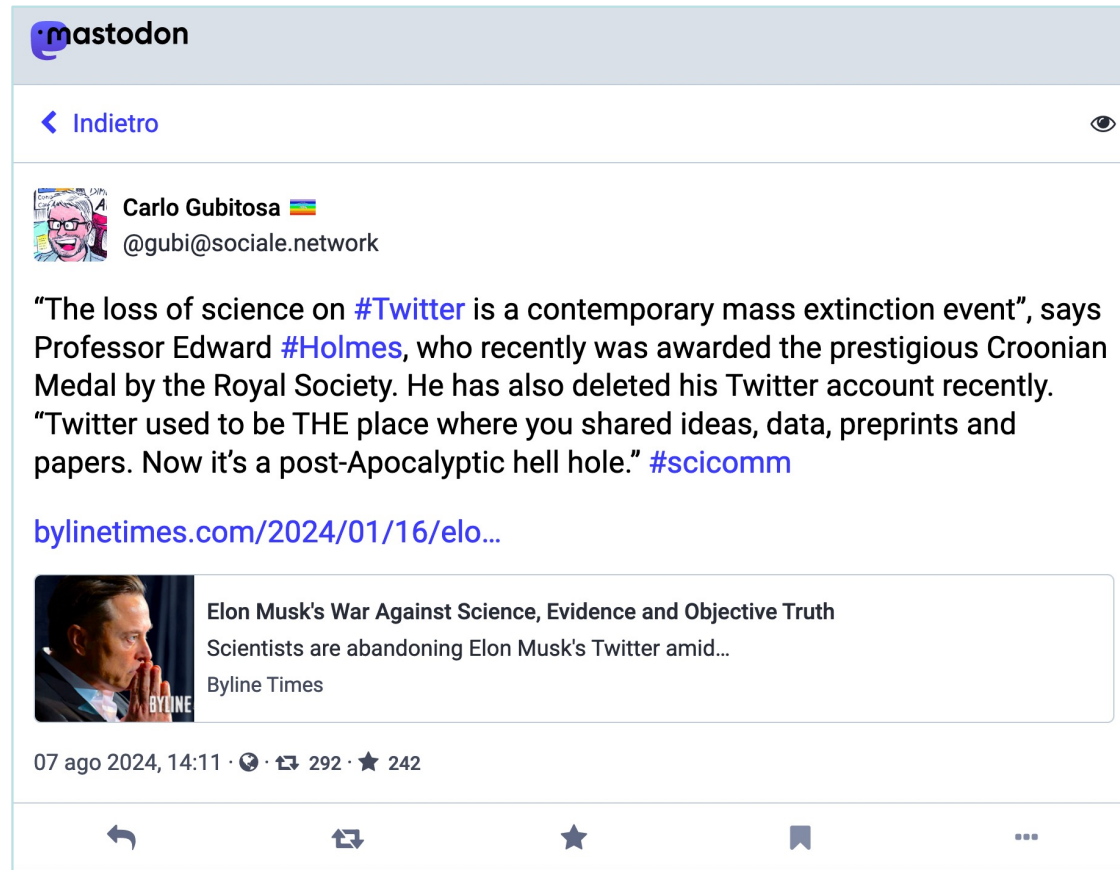




Two main and seemingly unrelated themes/challenges emerged from the meeting.

The first theme was the disappearance of informal research communication on twitter resulting from Musk's take-over. It has been characterized as a contemporary mass extinction event.

Theme 1: Xodus, Loss of Science Communication on Twitter



<https://sociale.network/@gubi/112920672314200247>


Markolin, P. (16 January 2024) Elon Musk's War Against Science, Evidence and Objective Truth
<https://bylinetimes.com/2024/01/16/elon-musks-war-against-science-evidence-and-objective-truth/>




In 2024, as part of its Innovation activities, SURF set up a Mastodon instance. It has not been successful to this date, going by the “188 active users” stat.

The question hanging around the room was whether the Fediverse with its distributed Mastodon instances could provide an alternative.

SURF Mastodon Pilot



Services ▾ Topics ▾

News Agenda About SURF 

Mastodon pilot for research and education

SURF and Universities of the Netherlands are jointly exploring Mastodon as an open source platform for education and research in the Netherlands. In which public values are paramount. We launched a pilot in February 2023. Join us and discover how students, researchers, staff and institutions can experiment with Mastodon in a low-threshold way.

Login to the Mastodon pilot environment



Mastodon pilot

About the pilot

Make available

Tutorials

Experimenting with Mastodon without effort

Mastodon is currently considered the alternative to

Straight to work


[Register and login to social.edu.nl](#)  (users)

<https://www.surf.nl/en/mastodon-pilot-for-research-and-education>

SURF Mastodon Pilot

Q Search


social.edu.nl is one of the many independent Mastodon servers you can use to participate in the fediverse.



Mastodon pilot

This is the Mastodon server run by SURF for the Dutch research and education community - more info
<https://surf.nl/mastodon-pilot>

ADMINISTERED BY:

 @SLURF Ser.
@SLURF

SERVER STATS:

188
active users

[social.edu.nl](#) · [About](#) · [Profiles directory](#) · [Privacy policy](#)


Mastodon: [About](#) · [Get the app](#) · [Keyboard shortcuts](#) · [View source code](#) · v4.3.8

Explore

Posts

Hashtags


News

 **University of Groningen Library**
@Bibliotheclaris 1h


New in our electronic #collection:

[➔ Resisting #BigTech: The Personalized is #Political](#)

bloomsburycollections.com/mono...



Resisting Big Tech
The Personalized is Political
Niels Niessen

 mastodon

Explore

Live feeds

Mastodon is the best way to keep up with what's happening.

Follow anyone across the fediverse and see it all in chronological order. No algorithms, ads, or clickbait in sight.

Login


<https://social.edu.nl/about>





The SURF Mastodon instance has a very interesting feature with this regard: Users are registered and authenticated by means of SURFConext, the single sign-on service for high education in The Netherlands that leverages institutional identity providers. Meaning posts by users of the SURF Mastodon instance can be unambiguously attributed, traced back to an actual person who works or studies at a Dutch higher education institution. Which is very attractive from a research communication perspective.

But how to get them to join? What is the carrot?

SURFconext Used for Mastodon Registration/Authentication



Services ▾Topics ▾


NewsAgendaAbout SURFNL  

[Home](#) > [Services](#) > [Identity & Access Management](#) > **SURFconext**

SURFconext

Secure access everywhere with one set of credentials


Let your users log in to all the cloud services your institution uses with one username and password. Both for services that everyone uses and services for small specialist teams. Secure, easy and privacy-friendly.





Advantages

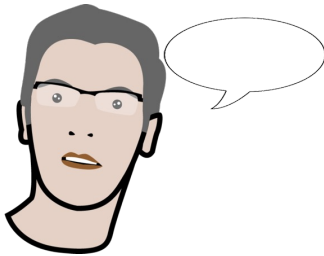
- ✓ **One interface for all cloud services**
With SURFconext, you arrange access to different services with one link instead of a separate link per service.
- ✓ **Safe and easy**
Users can securely and easily access cloud services with one account, that of the institution (single sign-on).
- ✓ **Privacy-friendly**
SURFconext transfers only the personal data that is really necessary to the service provider.
- ✓ **Also for own services**
As an institution, you can make your own services easily accessible to your own staff and students, and those of

Do you have a question about SURFconext? Get in touch.



Team SURFconext
 n.v.t.
 info@surfconext.nl

<https://www.surf.nl/en/services/identity-access-management/surfconext>



The second theme related to the efforts in the context of Open Science to assess researchers in novel ways, to take a broad range of activities/contributions into account for assessments.

- There is global activity in this realm, for example, under the umbrella of COARA, the Coalition for Advancing Research Assessment.
- And recently I came across a UNESCO document “Principles of Open Science Monitoring” that is in request for feedback status. It actually lists Roberto Di Cosmo as one on the contributing experts.

Theme 2 - Holistic Assessment of Researcher Contributions

Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.

Call for inputs: Global Consultation on the Draft Principles of Open Science Monitoring

As part of the ongoing efforts to advance open science monitoring in alignment with the 2021 UNESCO Recommendation on Open Science, UNESCO is calling for inputs and comments from all the regions and interested stakeholders on the Draft Principles for Open Science Monitoring.

Deadline for inputs: 30 November 2024



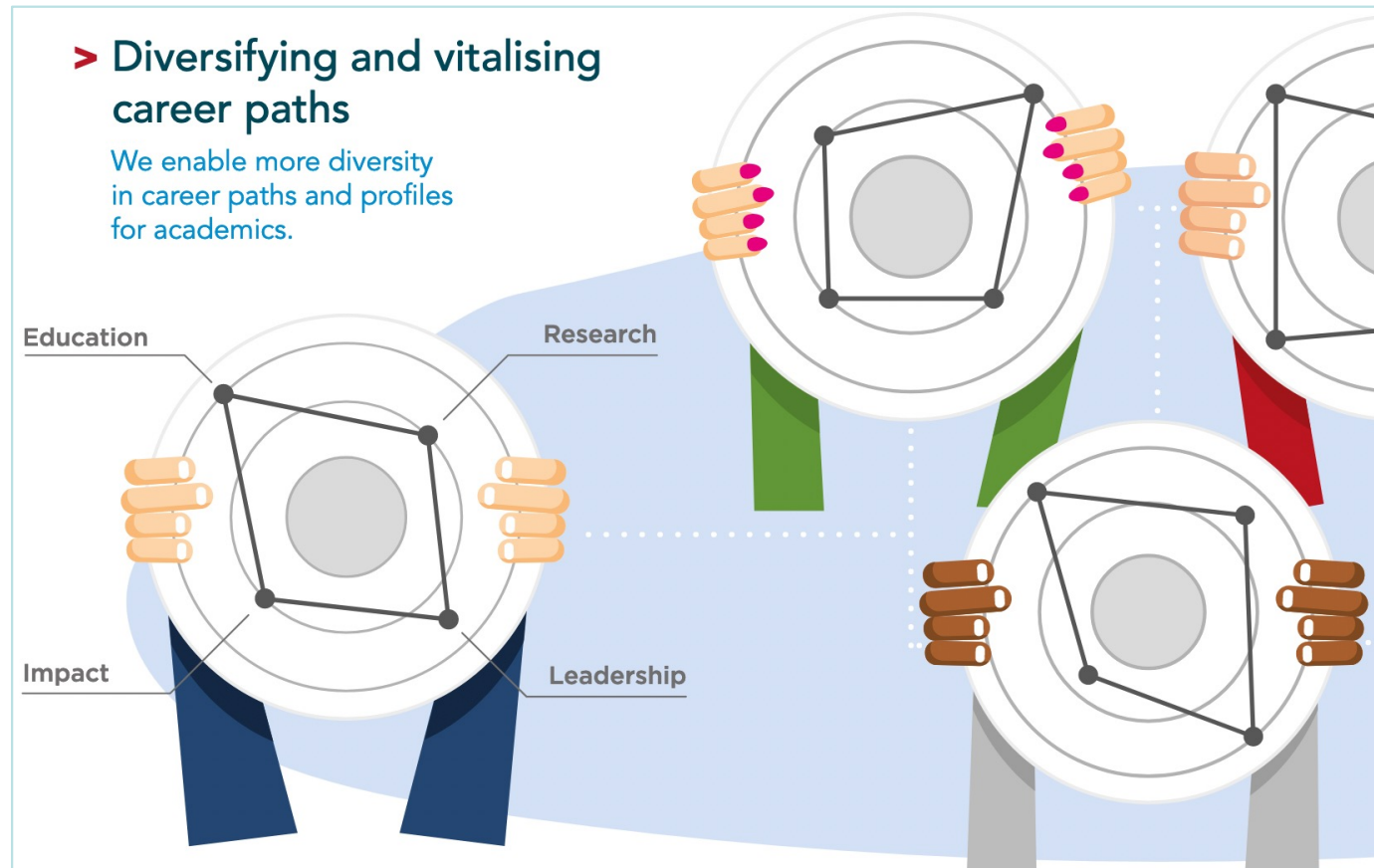
<https://coara.eu/>

<https://www.unesco.org/en/articles/call-inputs-global-consultation-draft-principles-open-science-monitoring>



This desire to assess researchers in novel ways also very much exists in the Netherlands. And there is clear interest in areas that go beyond the creation of scholarly artifacts, including, under the "impact" umbrella penetration of research results into society, the media, i.e. a newspaper article about a researcher's work, an interview on the radio, an appearance on a TV program, etc.

Theme 2 - Holistic Assessment of Researcher Contributions



graspos – Open Research Assessment Dataspace

28

graspos Researcher Profile

Narrative that goes beyond research outputs

- Supported by evidence
- Using a variety of open data sources

How

- Different templates for different uses
- Novel ideas for presentation

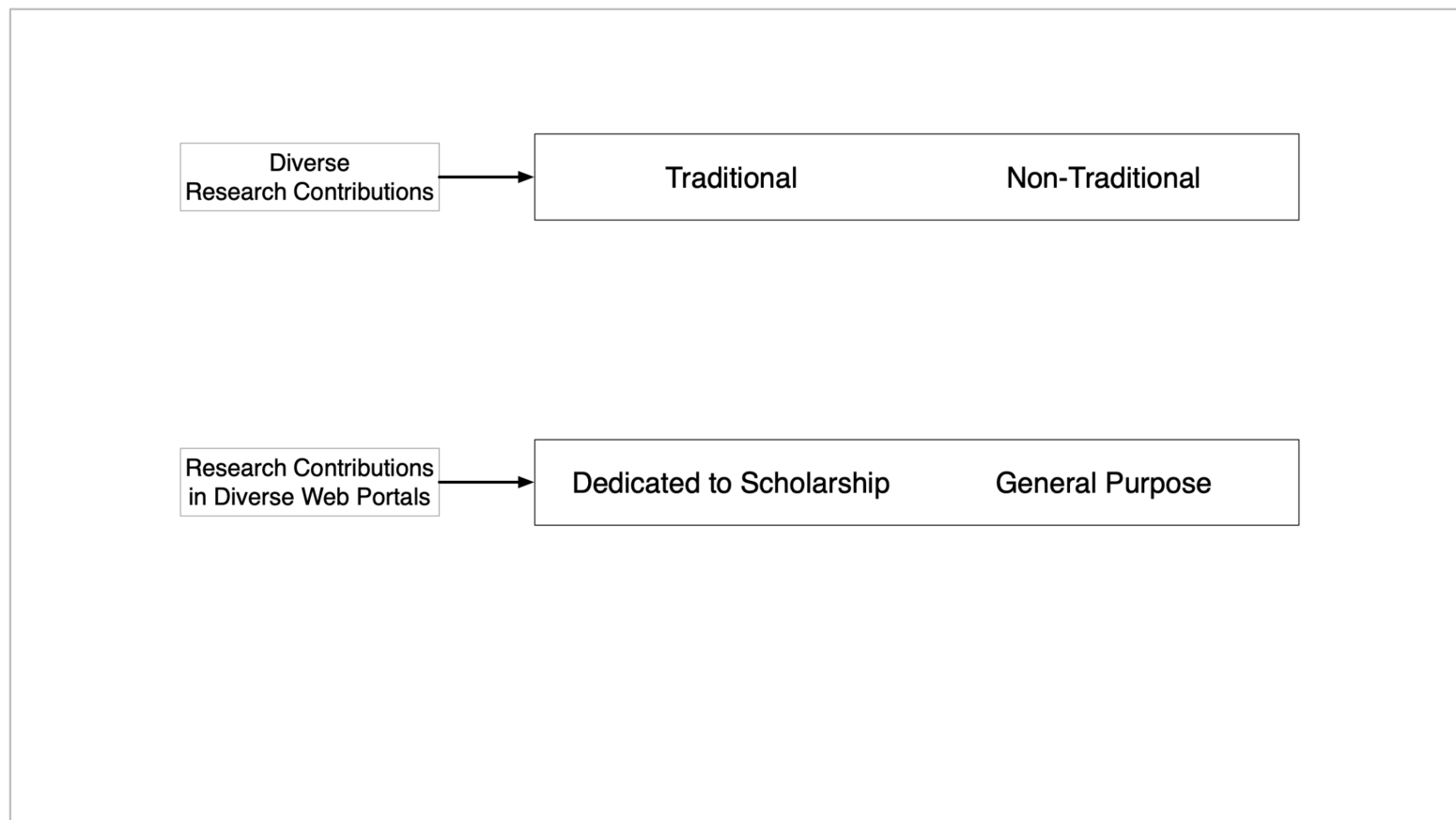
Challenges

- How to capture information beyond research outputs
- ORCID extensions a valuable tool, but far from becoming extensively used

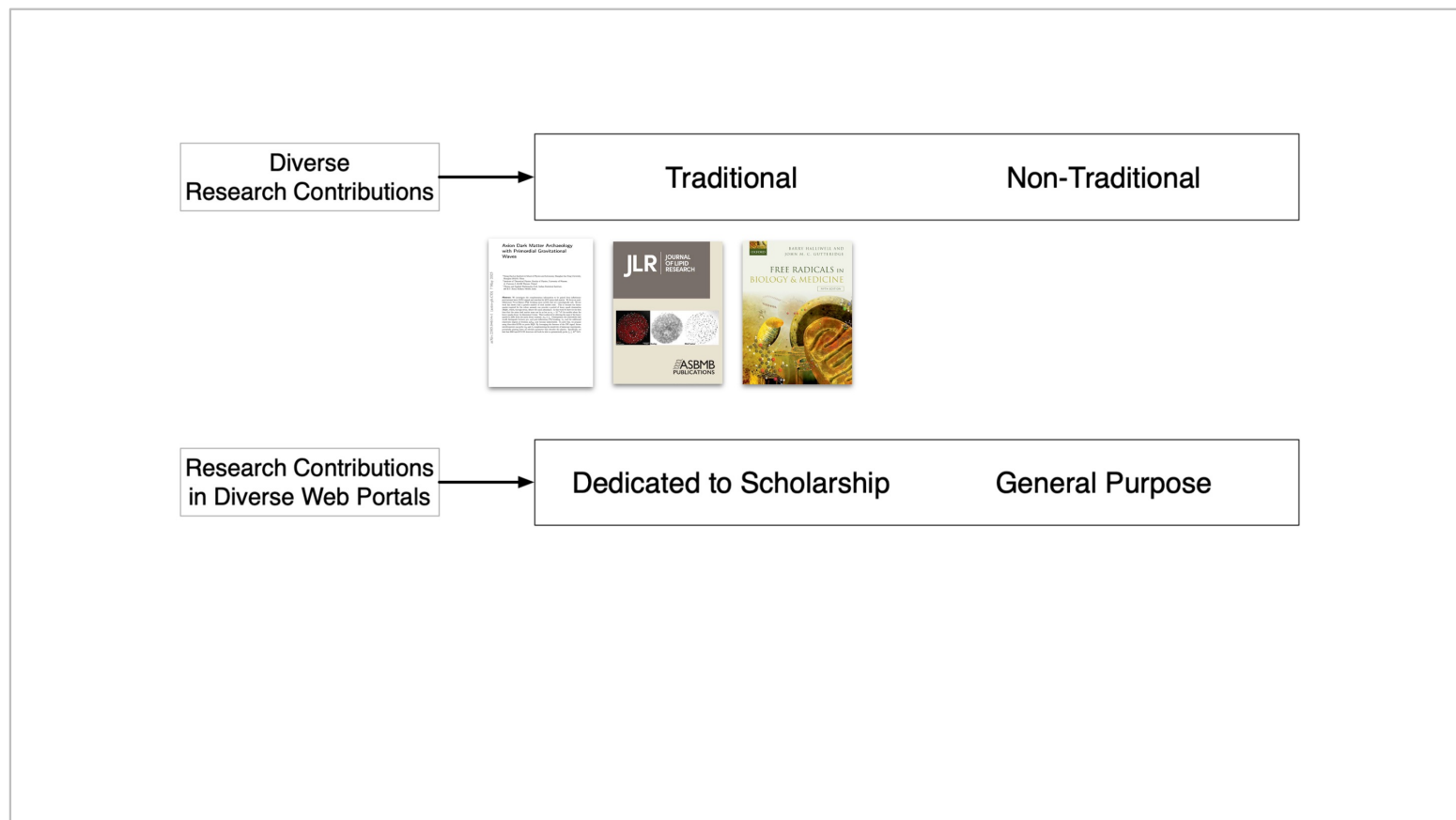


EuroCRIS 2024 | Vienna | May 15-17, 2024

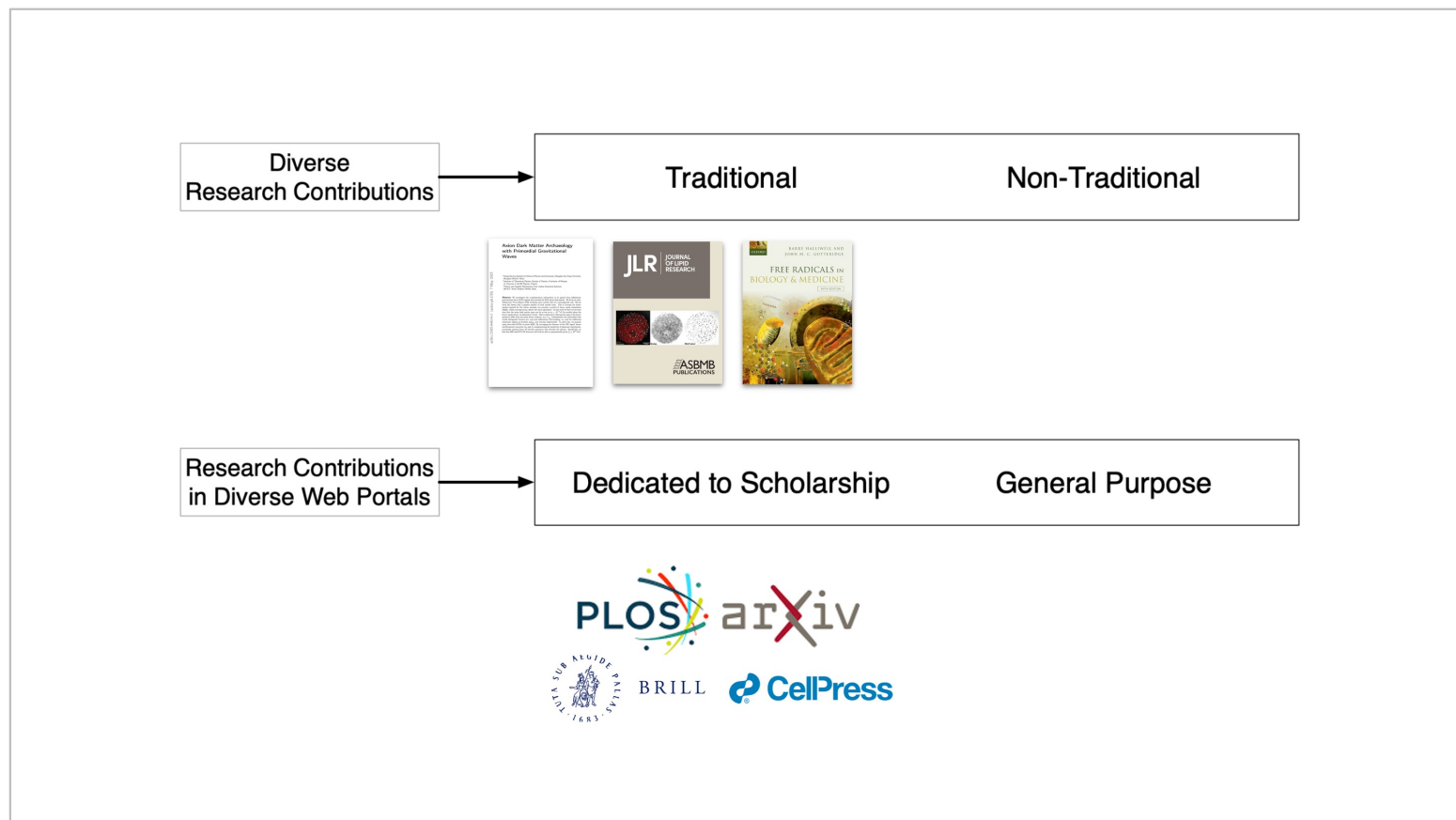
How to Find Non-Traditional Research Contributions?



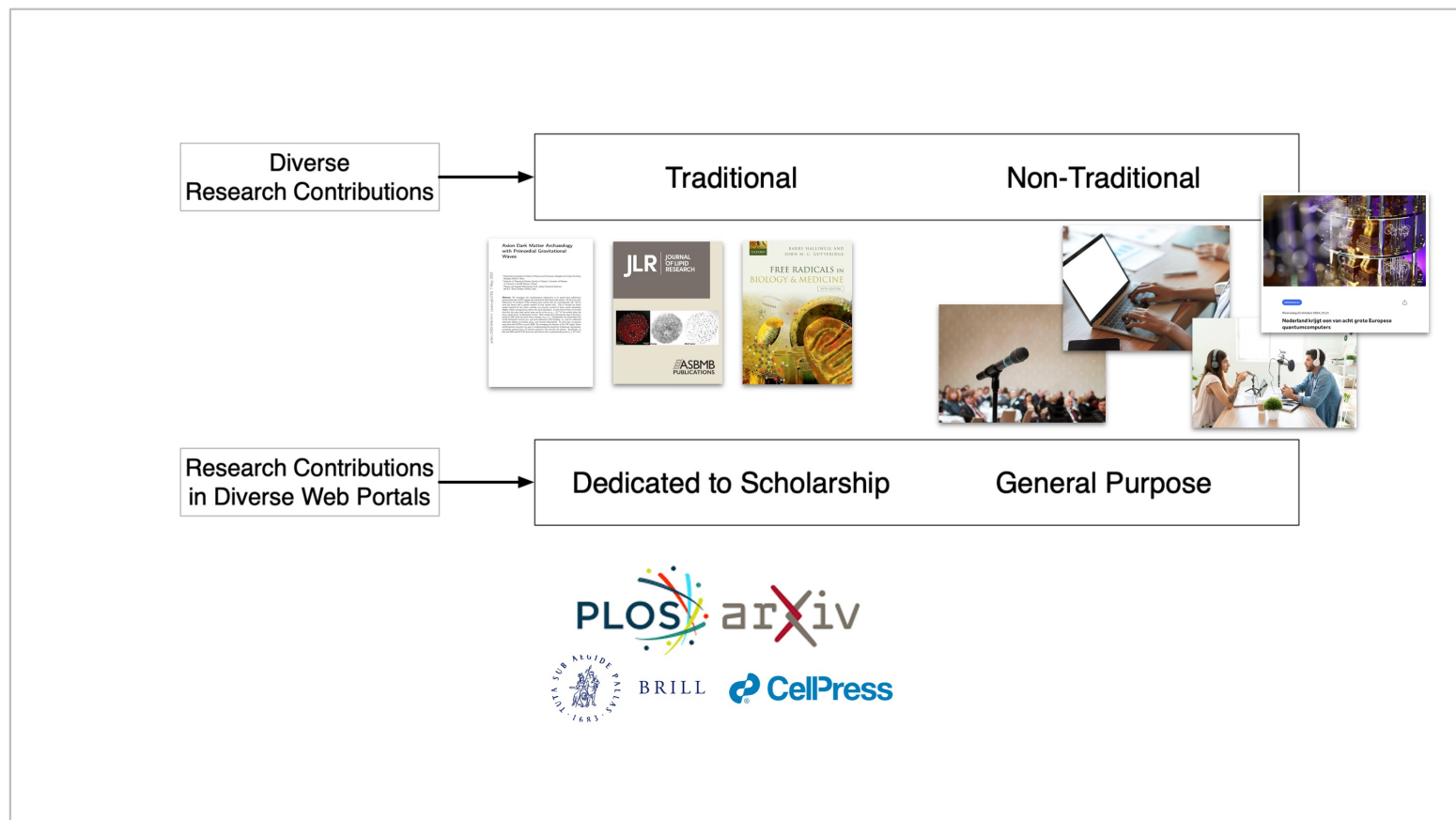
How to Find Non-Traditional Research Contributions?



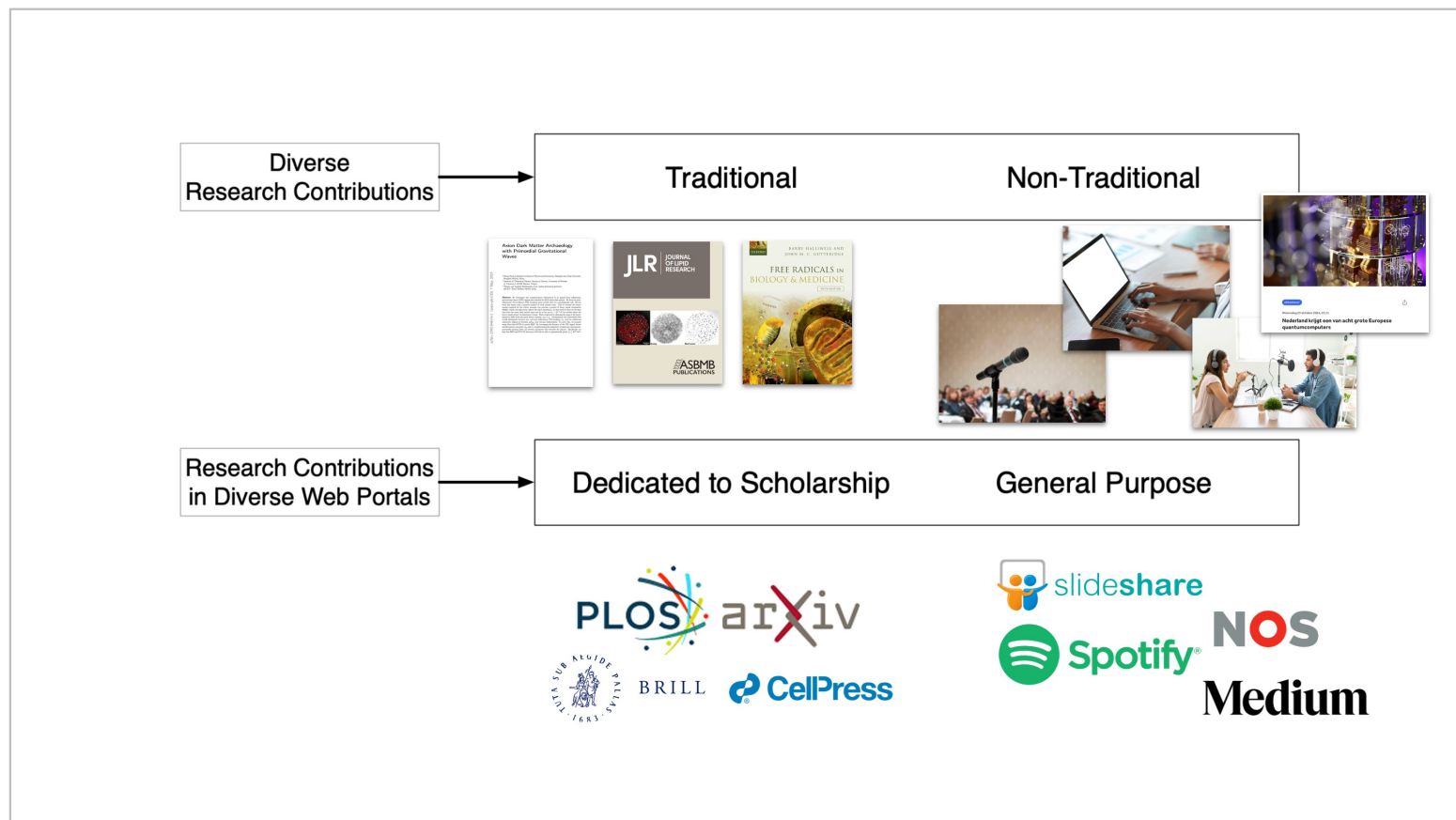
How to Find Non-Traditional Research Contributions?



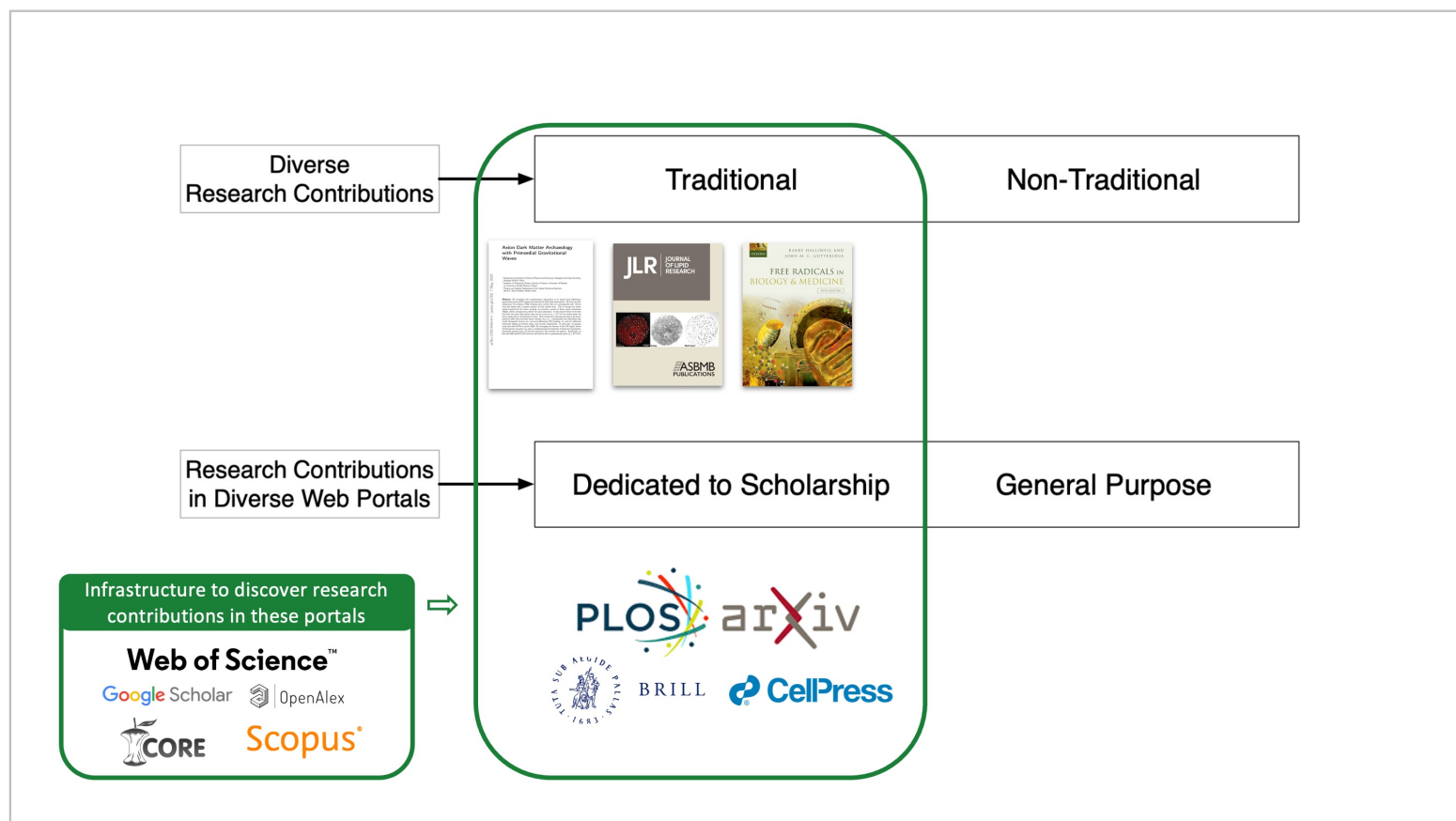
How to Find Non-Traditional Research Contributions?



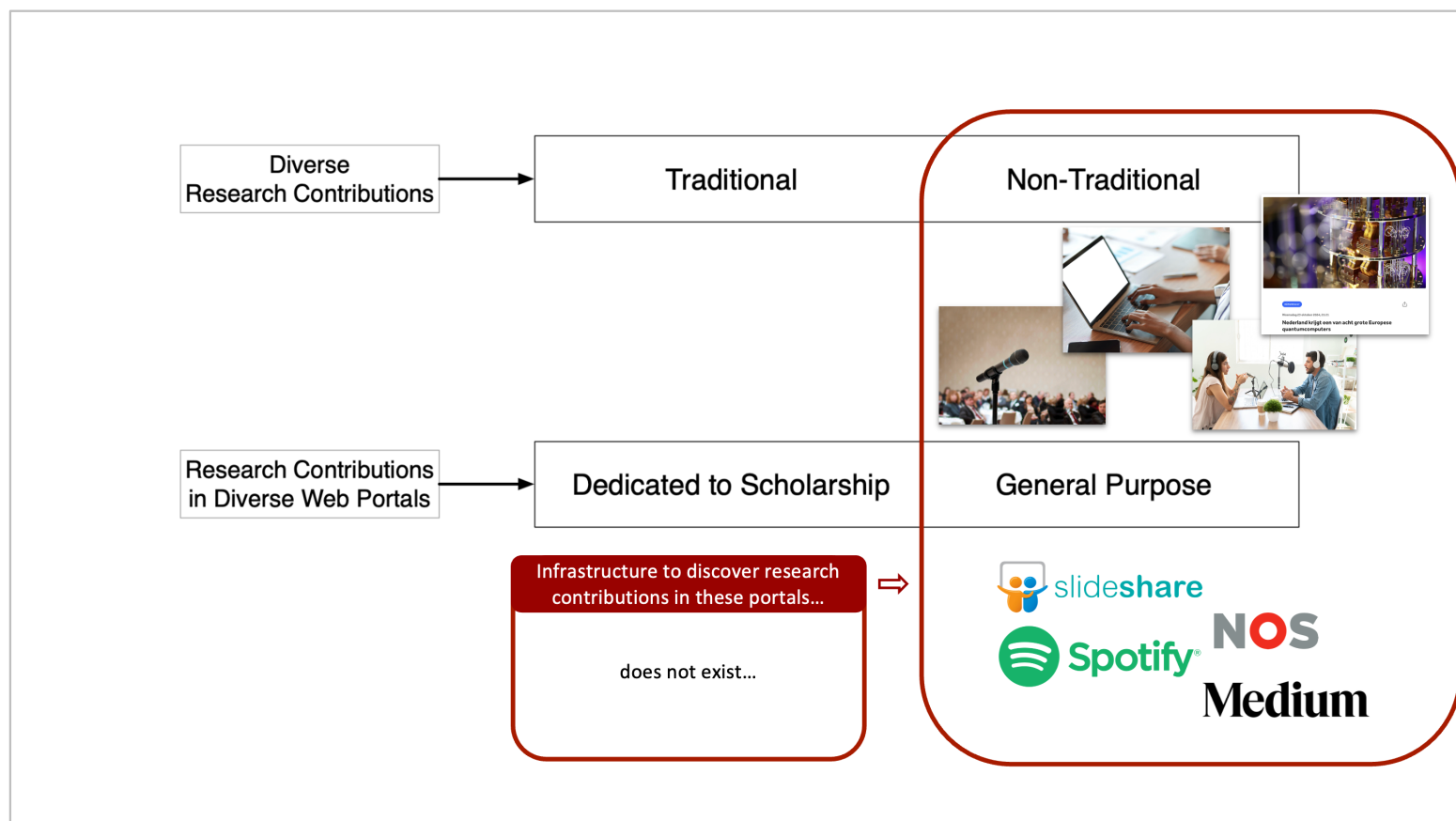
How to Find Non-Traditional Research Contributions?

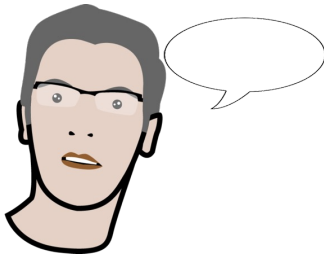


How to Find Non-Traditional Research Contributions?



How to Find Non-Traditional Research Contributions?





A challenge with this regard, though, is how to get an overview of all pertinent activities/contributions in a manner that doesn't put all the burden on the researchers.

The EC-funded GRASPOS project operates in this realm and aims at generating holistic researcher profiles, for example, even including things like conference attendance. I heard OpenAIRE's Nathalia Manola talk about the project at the recent CRIS2024 conference in Vienna. She confirmed the challenge.

As we learned in the myresearch.institute perspective, one can't monitor it all. The monitoring approach can work when a select number of environments are deemed relevant but it breaks down when the environment that is to be monitored is the entire web.

Claim It Project Context – SURF Meeting 02/2024

- Build a prototype for a solution that:
 - **Addresses the problem of finding research contributions needles in the web haystack**
 - Can be a catalyst to (re)start information scholarly communication on social media
- Inspiration from prior work:
 - myresearch.institute experiment (2018-2020)
 - bots on Twitter





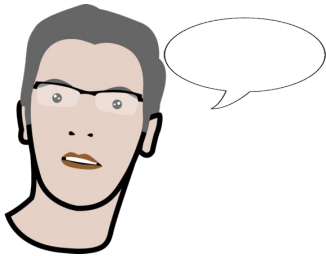
Overall the considerations at the basis of the project were that:

- Researchers use a variety of web platforms for collaboration and communication
- They use platforms that are dedicated to scholarship as well as general purpose
- The researchers' institutions are not aware of contributions made in general purpose portals and many times also not of those in scholarly portals that deal with less traditional contributions.
- Although contributions are not systematically archived, institutions have an incentive to obtain broad coverage of the contributions of their researchers. After all, institutions have set up institutional repositories and/or CRIS systems to cover contributions by their researchers to the research literature. Institutions have started setting up repositories to cover research data. It seems like a logical next step for institutions to start looking into systems that can cover a broader range of their researchers' outputs.

myresearch.institute Experiment - Considerations

- Researchers use a variety of web platforms for collaboration and communication.
- They use platforms that are dedicated to scholarship as well as general purpose.
- The researchers' institutions are not aware of contributions made in general purpose portals.
- Institutions have an incentive to obtain broad coverage of the contributions of their researchers.





We decided to explore the potential of the Personal Web Observatory approach in an institutional setting, i.e. institutions operate infrastructure to automatically track contributions of researchers in web portals.

We created a fictitious research institution and populated it with 16 researchers who had given their consent to track them across a range of portals.

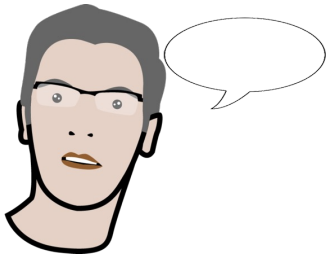
In our institutional observatory, the researchers were uniquely identified by means of their ORCID and in each portal they had a web identity that our observatory knew about. That is essential to be able to track them.

myresearch.institute Experiment - Researchers

- 16 researchers
- Uniquely identified by ORCIDiDs
- Web identities in multiple portals
- Make a variety of contributions



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025



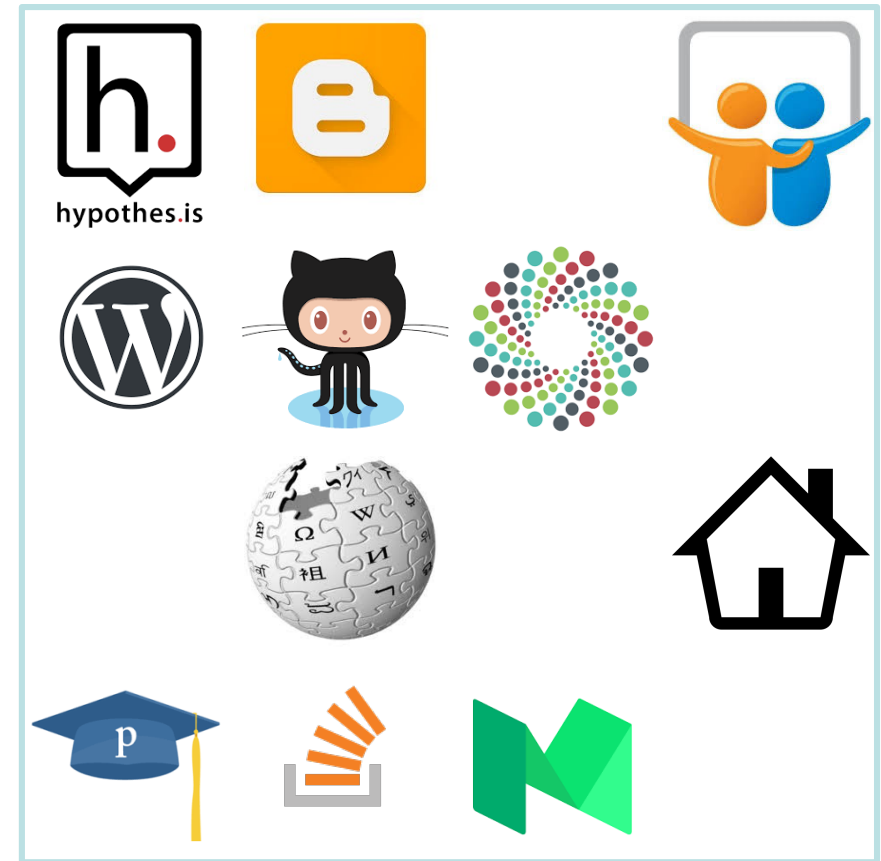
We selected 11 portals in which to track our researchers. Some are general purpose portals and some are focused on research: hypothes.is, blogger, slideshare, wordpress, github, figshare, wikipedia, home page, publons, stack overflow, ...

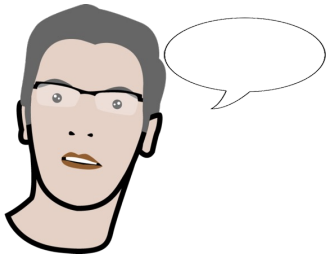
10.187 unique contributions were tracked between August 2018 and March 2020.



myresearch.institute Experiment - Portals

- 11 portals
- Tracking started August 2018
- Tracking ended March 2020
- 10.187 unique contributions tracked

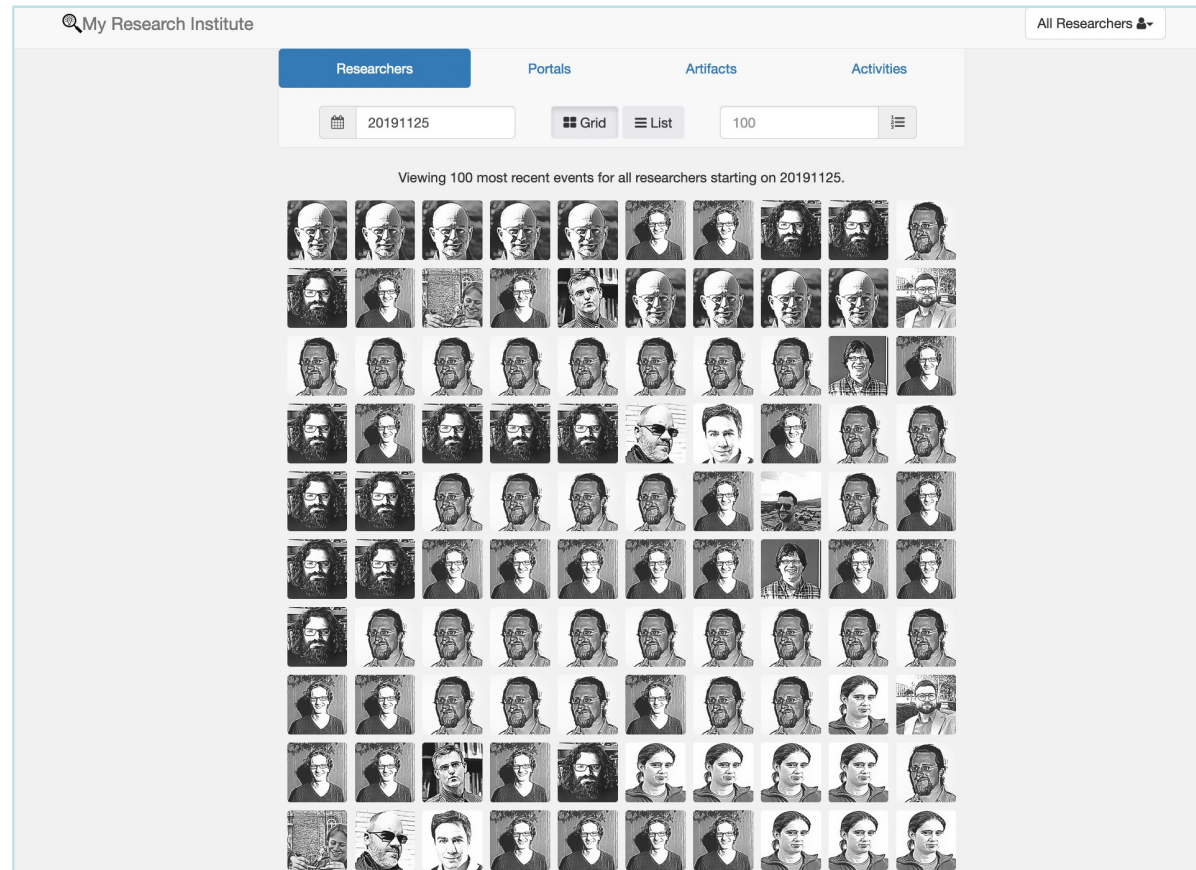




The event store was gradually populated over time and a user interface, rather inspired by Amy Guy's personal web observatory, allowed getting an overview of our researcher's contributions.

It provided ways to get a chronological overview of contributions by all researchers, organized by the researcher who made the contribution, the portal, the contribution type, and the activity type. It also allowed to zoom in on a single researcher.

myresearch.institute – Contributions - Researcher View



<https://myresearch.institute/>



And there are also a few insights gained from this effort that informed the overall approach of the mycontributions.info experiment that I will talk about in a bit:

- It's not just institutions that do not know about researcher contributions in those web portals. The researchers know they make them but lose track of it all. As such, they much appreciated the overview provided in the institutional web observatory. They even asked whether our experimental setup had an API so they could get their stuff out.
- One can track only so many portals and obviously researchers will make contributions to portals that were not tracked
- Not all tracked contributions seem to be scholarly; there's definitely a mix of private and scholarly on general purpose platforms

myresearch.institute – Insights Gained

- Researchers don't remember all the contributions they make and as such much appreciated the overview provided in the institutional web observatory
- Researchers make contributions to portals that were not tracked; can't track them all
- Not all tracked contributions seem to be scholarly; there's definitely a mix of private and scholarly on general purpose platforms





So, one needs to turn the whole thing around. And that was a major insight of the brainstorm at SURF: instead of detecting researchers' traces by monitoring portal APIs, have researchers themselves announce them, in the case of the SURF experiment by posting about them on Mastodon.

The specific approach chosen draws inspiration from tools that had been created on Twitter, such as #ICanHazPDF to obtain a copy of a research article that sits behind a paywall and @icanhazmemento to request the creation of a Memento for a certain URL. In this case, researchers will post about resources that they consider their contributions.

The incentive for the researcher to post is to obtain holistic data that can be used for assessments but, honestly, also to just keep track of what they've been up to. And the hope is that the carrot of the contribution list will be sufficient to attract Dutch researchers to Mastodon and start informal research communication there. We will learn whether that hope is justified once SURF decides on a community in which to run a pilot.



#ICanHazPDF

WIKIPEDIA

The Free Encyclopedia

Q

∞

Create account

Log in

...

ICanHazPDF

8 languages

Article

Talk

Read

Edit

View history

Tools

From Wikipedia, the free encyclopedia

The correct title of this article is **#ICanHazPDF**. The omission of the **#** is due to *technical restrictions*.

#ICanHazPDF is a [hashtag](#) used on [Twitter](#) to request access to [academic journal](#) articles which are behind [paywalls](#).^[1] It began in 2011^[2] by scientist [Andrea Kuszewski](#).^{[3][4]} The name is derived from the meme [I Can Has Cheezburger?](#)^[4]


Process

edit

Users request articles by tweeting an article's title, [DOI](#) or other linked information like a publisher's link,^[5] their email address, and the hashtag "#ICanHazPDF". Someone who has access to the article might then email it to them. The user then deletes the original tweet.^[6] Alternatively, users who do not wish to post their email address in the clear can use direct messaging to exchange contact information with a volunteer who has offered to share the article of interest.

Part of a series on

File sharing



Technologies

[show]

Video on demand sites

[show]

BitTorrent sites

[show]

Academic/scholarly

[show]

File sharing networks

[show]

P2P clients

[show]


Streaming programs

[show]

Anonymous file sharing

[show]

@icanhazmemento



I Can Haz Memento
@icanhazmemento

Following

Use [#icanhazmemento](#) to archive URLs in a tweet. Or reply a tweet with [#icanhazmemento](#) to archive the URLs in the tweet. Via [@WebSciDL](#).

Joined July 2015


1 Following 53 Followers

Followed by [archive.today](#), [Anna Perricci](#), [Michael L. Nelson](#), and 7 others

Posts


Replies

Media



I Can Haz Memento @icanhazmemento · 8/27/20
Replying to @shawnmjones
.@shawnmjones, here's your newly archived page:
[timetravel.mementoweb.org/memento/202008...](#) ().

See other versions: [timetravel.mementoweb.org/list/202008271...](#)
See machine readable versions: [memgator.cs.odu.edu/timemap/json/h...](#)

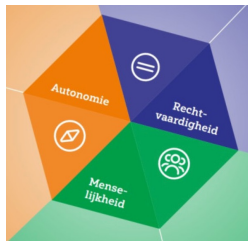


I Can Haz Memento @icanhazmemento · 8/27/20
Replying to @shawnmjones
.@shawnmjones, here's your newly archived page:
[timetravel.mementoweb.org/memento/202008...](#) ().

See other versions: [timetravel.mementoweb.org/list/202008271...](#)
See machine readable versions: [memgator.cs.odu.edu/timemap/json/h...](#)

<https://en.wikipedia.org/wiki/ICanHazPDF>
<https://x.com/icanhazmemento>

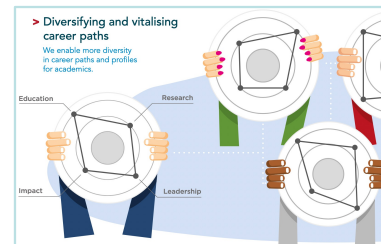
Claim It Prototype - Overview



publieke waarden



mastodon



research contributions



Carol

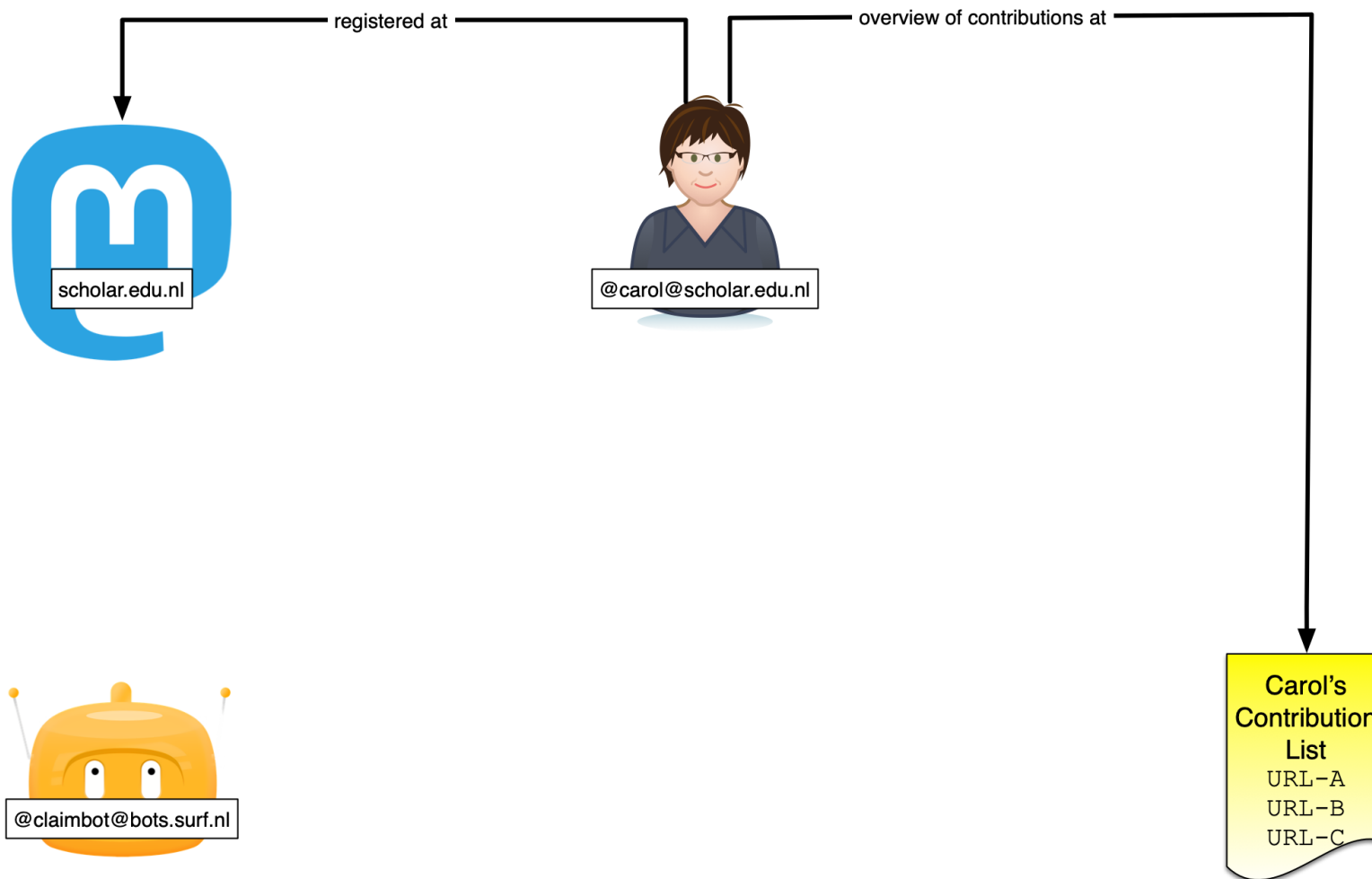


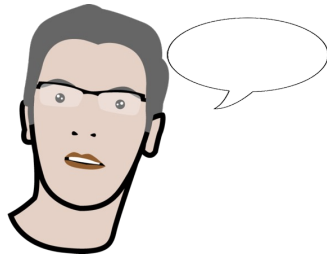
Here's a high-level overview of the the design of the experimental setup that Patrick is currently working on.

I am registered with the SURF Mastodon instance.

I have a Contribution List.

There's also a bot in the Fediverse – the claimbot - that accepts requests from users registered at the SURF Mastodon instance.

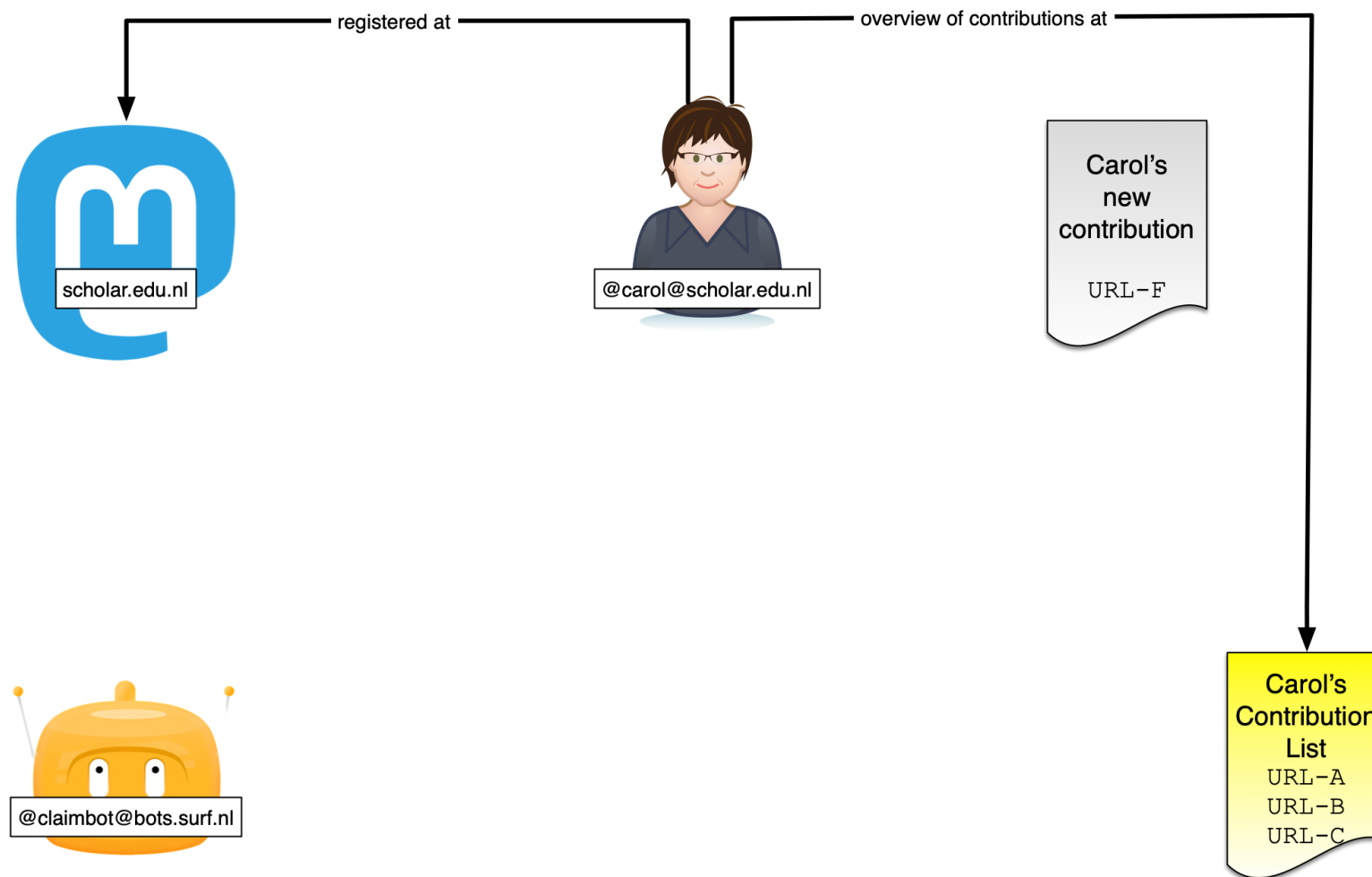




Let's now introduce a new contribution I made, which I want to claim.

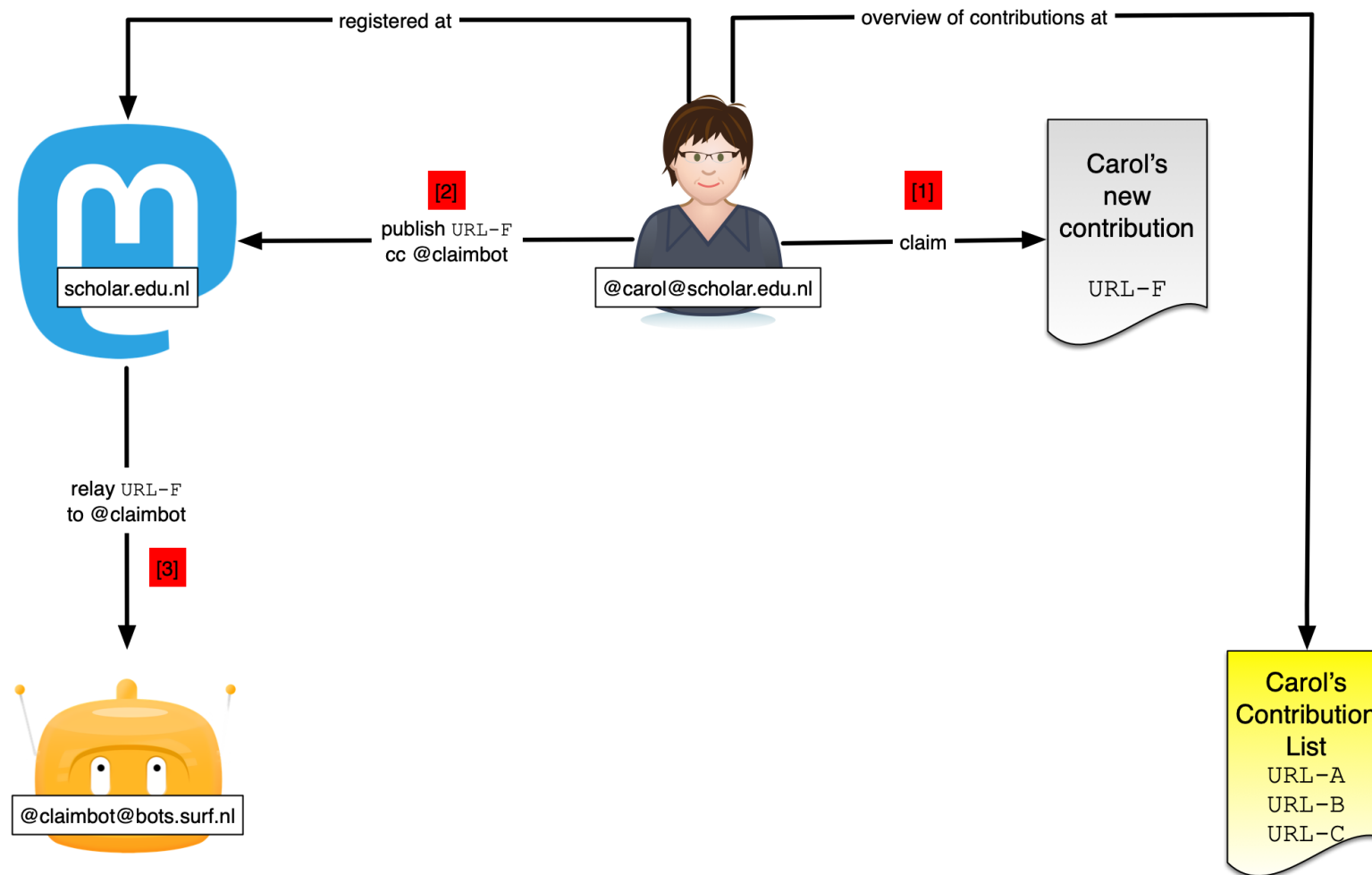


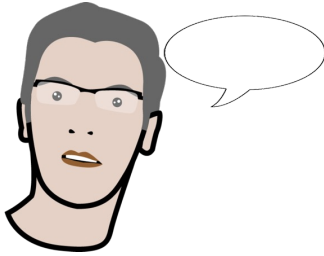
@hvdsonp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025





So, I am going to create Mastodon posts with the URL of these pages and I will target them at the claimbot.





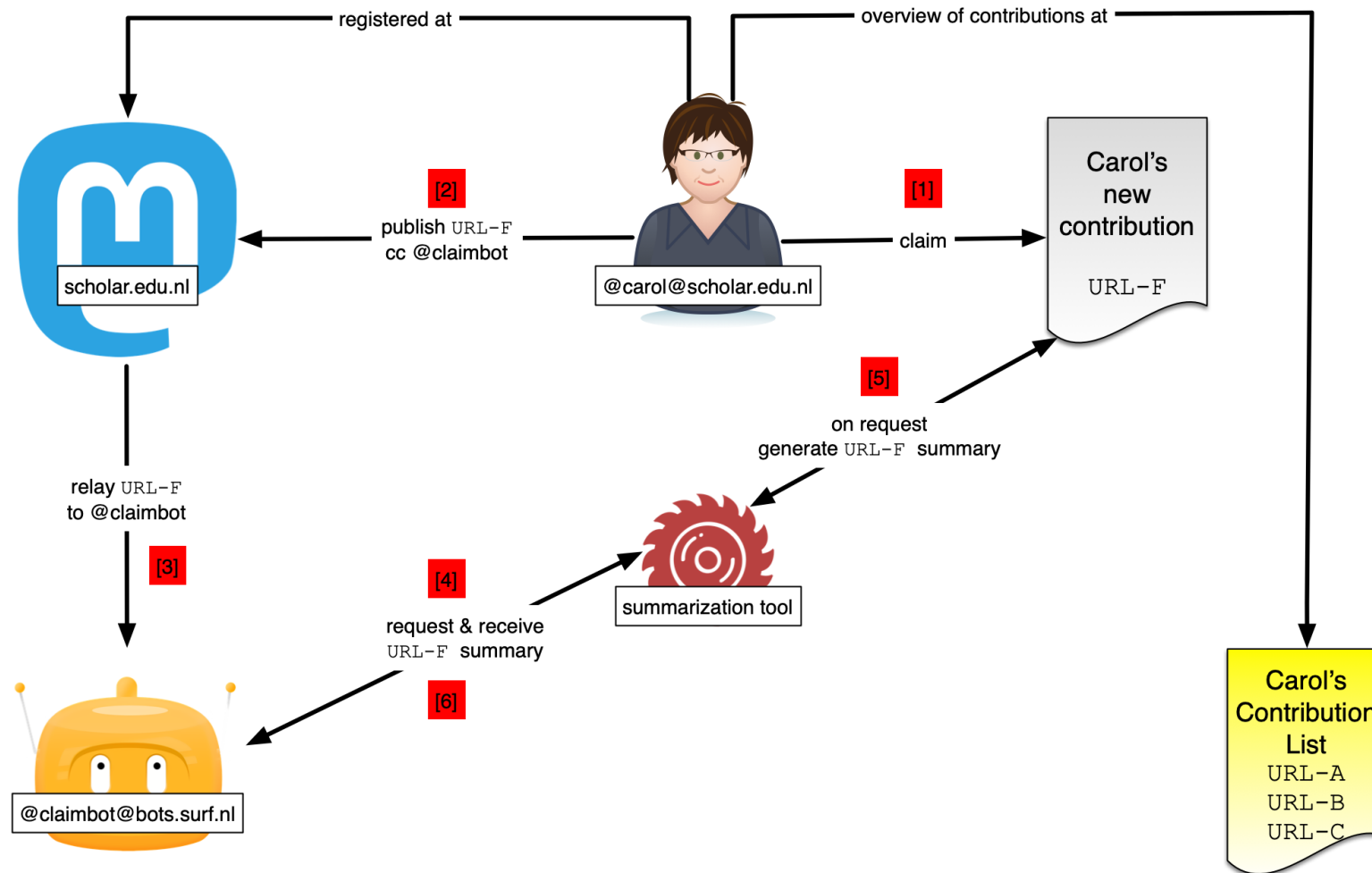
- Prior to the Conference, I express my excitement to be speaking
- After the talk, I share my slides

Note that I have sent these as private messages so as to not spam the SURF Mastodon instance. But in reality I would send these publicly so that any one can see my new contributions.



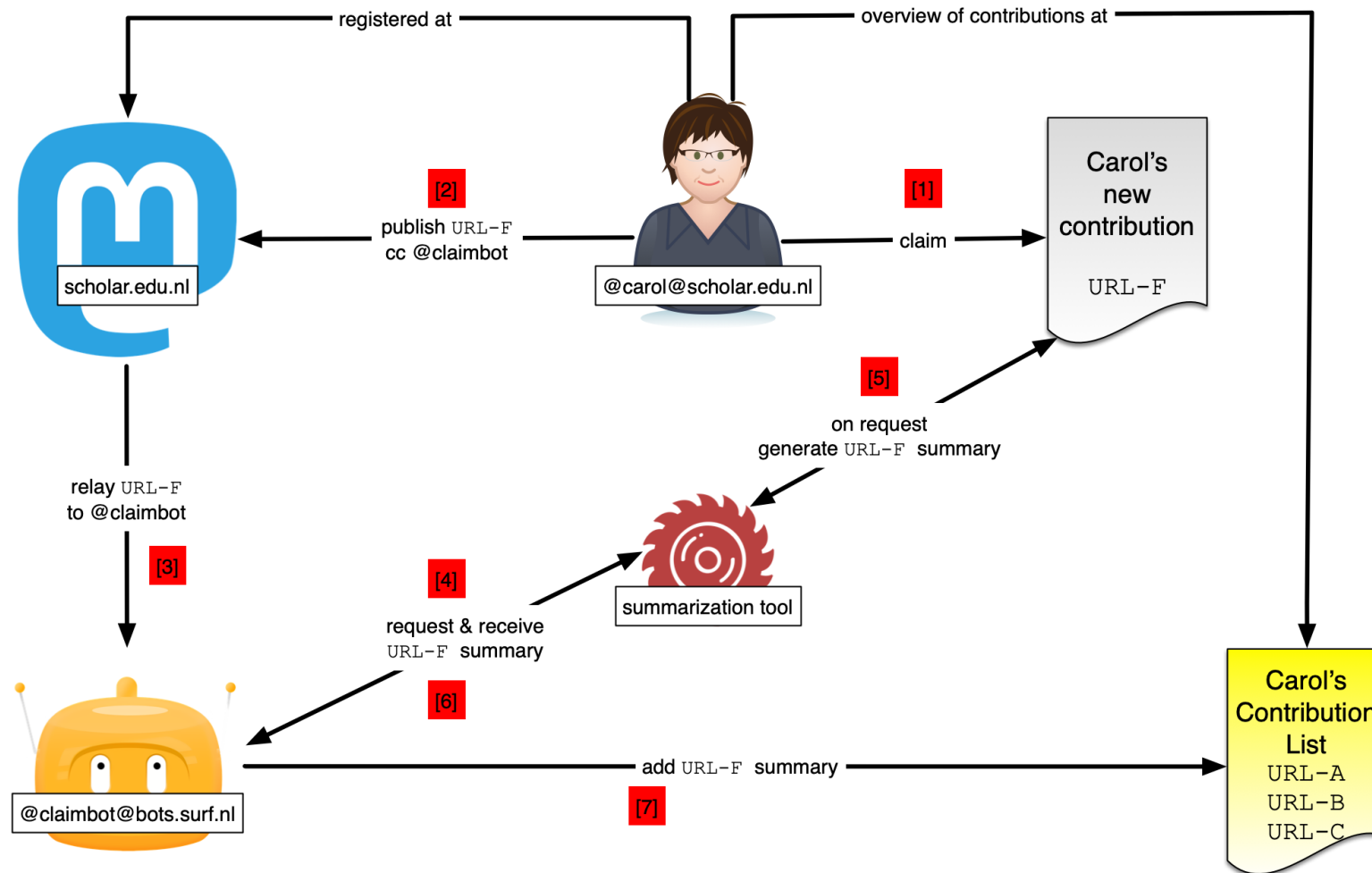
The claimbot receives the post, checks whether it originates from someone registered with the SURF Mastodon instance, and if so, calls a service that automatically extracts pertinent metadata about the contribution.

That information extraction is not conducted by calling APIs as was the case in myresearch.institute but by obtaining information via the contribution page.





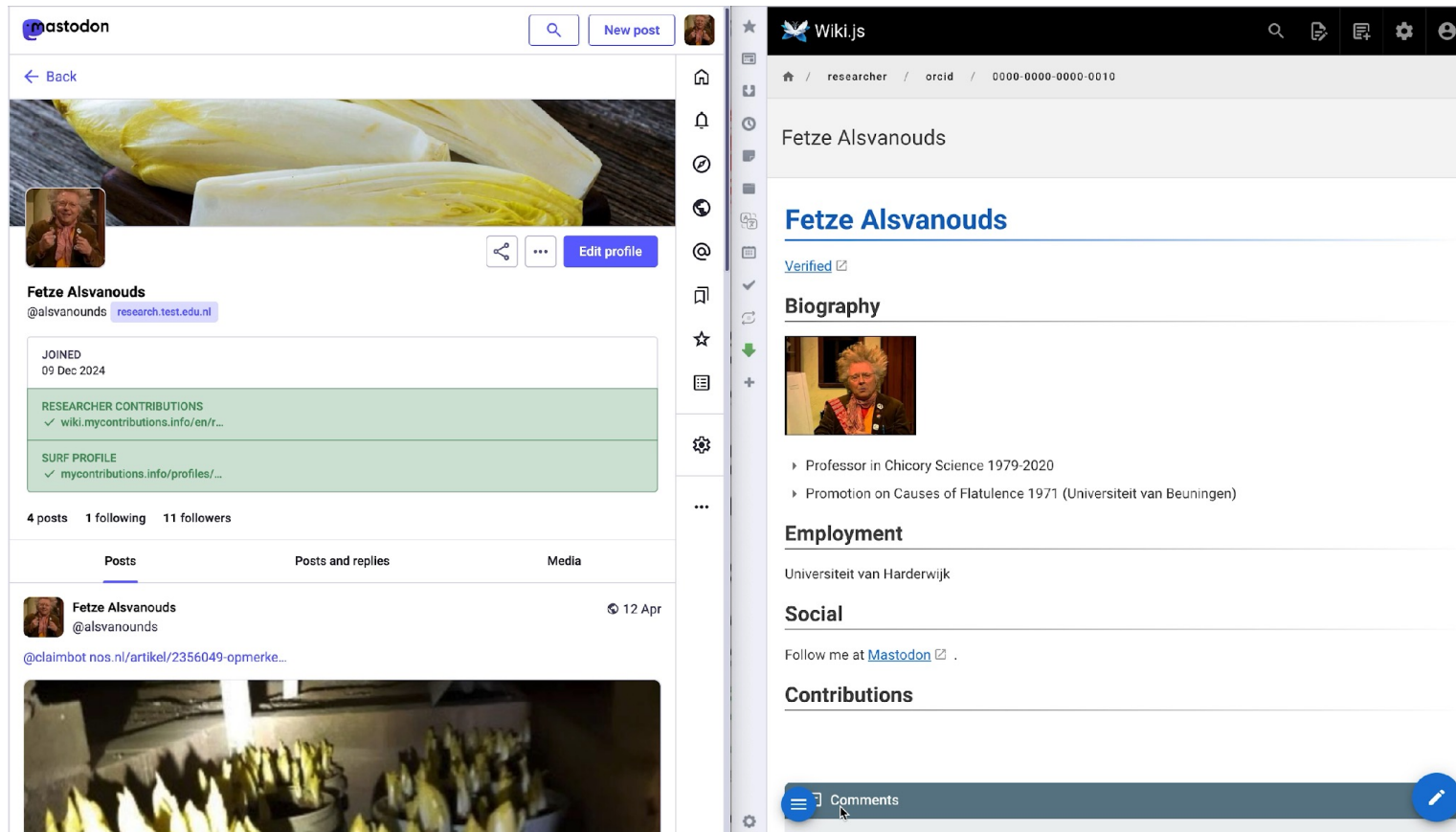
The claimbot adds an entry for these new contributions to my Contribution List.





As you can see the two new contributions were added to my Contribution List. The metadata extracted is pretty minimal in this case but there's at least an entry for the contributions.

Claim It Prototype - Screencam Demonstration



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

Claim It – Next Steps

- Pilot with Hogescholen
- **Addressing challenges**
- **Devising generic architecture that allows for diverse yet compatible implementations**
- Advocacy for implementations in other countries/communities



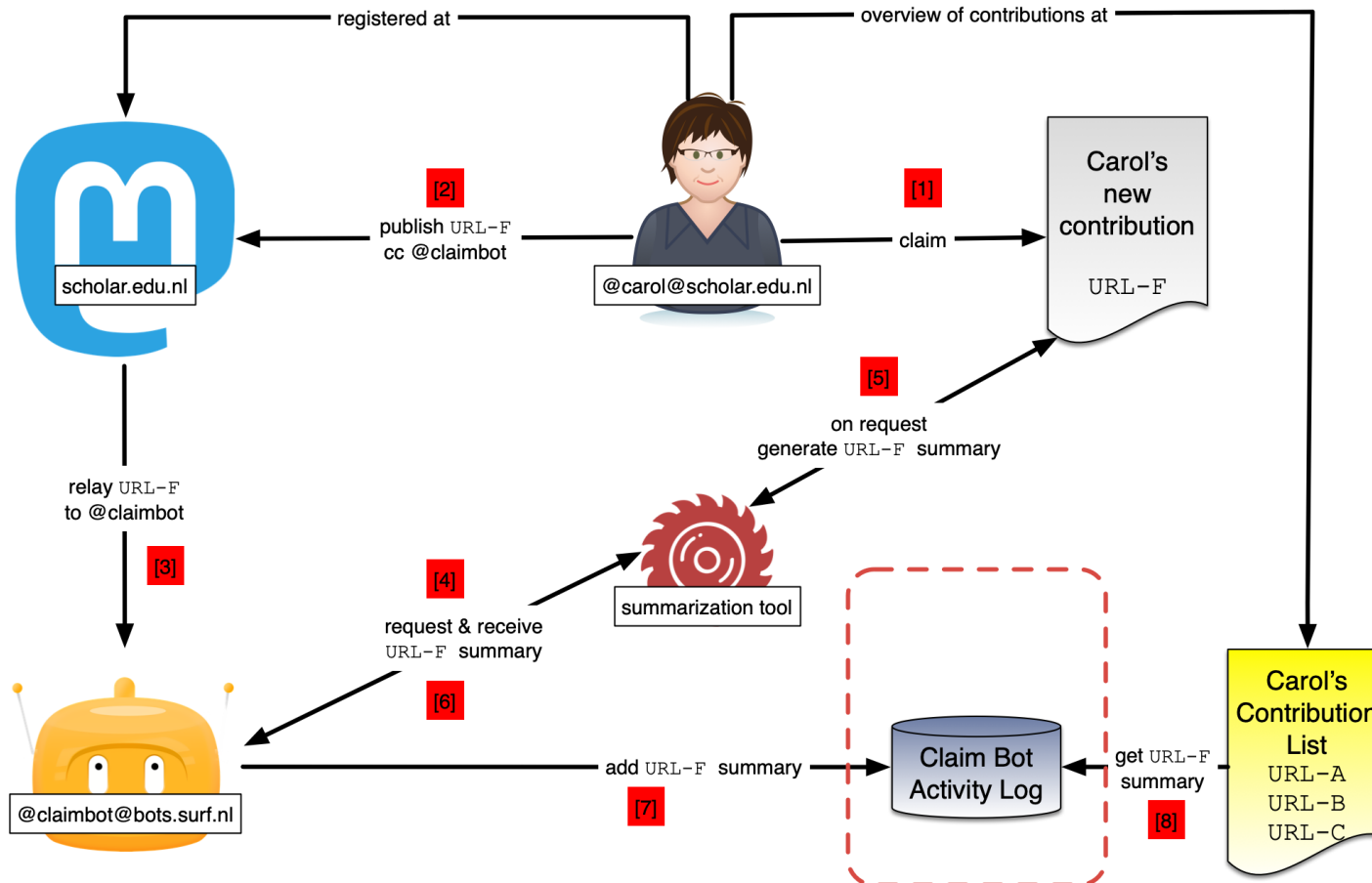
@hvdsonp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

Claim It – Next Steps

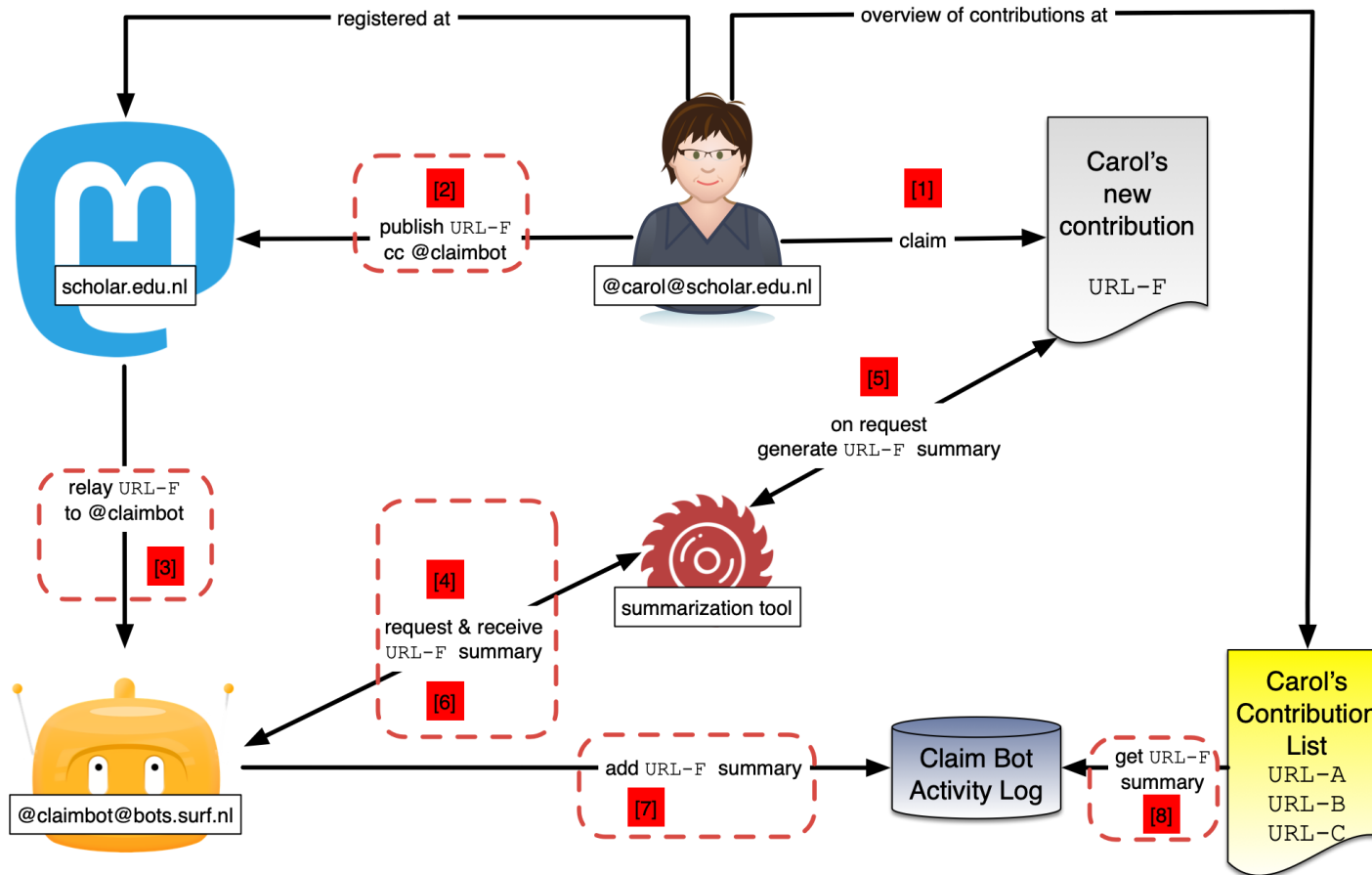
- Integration with existing repositories
- Use of open interoperability protocols
- Identity verification
- Trusted archiving of (traces of) research contributions
- Summarization challenges



Repository Integration



Use of Open Interoperability Protocols





The communication between the claimbot and the extraction tool uses the notification approach described in the Event Notifications in Value-Adding Networks specification authored by Patrick Hochstenbach and colleagues.

The approach is itself based on Linked Data Notifications and ActivityStreams2 vocabulary and is, among others, also used in the COAR Notify project for communication between repositories and overlay peer-review services. But. the notification approach is increasingly attracting attention for other use cases such as requesting archiving of resources, requesting/announcing augmentation of metadata, registration of datasets in registries, etc.

Mastodon

ActivityPub

W3C Recommendation 23 January 2018



This version:

<https://www.w3.org/TR/2018/REC-activitypub-20180123/>

Latest published version:

<https://www.w3.org/TR/activitypub/>

Latest editor's draft:

<https://w3c.github.io/activitypub/>

Test suite:

<https://test.activitypub.rocks/>

Implementation report:

<https://activitypub.rocks/implementation-report>

Previous version:

<https://www.w3.org/TR/2017/PR-activitypub-20171205/>

Editors:

[Christine Lemmer-Webber](#)

[Jessica Tallon](#)

Authors:

[Christine Lemmer-Webber](#)

[Jessica Tallon](#)

[Erin Shepherd](#)

[Amy Guy](#)

[Evan Prodromou](#)

Activity Streams 2.0

W3C Recommendation 23 May 2017



This version:

<https://www.w3.org/TR/2017/REC-activitystreams-core-20170523/>

Latest published version:

<https://www.w3.org/TR/activitystreams-core/>

Latest editor's draft:

<http://w3c.github.io/activitystreams/core/>

Test suite:

<https://github.com/w3c/activitystreams/tree/master/test>

Implementation report:

<https://github.com/w3c/activitystreams/tree/master/implementation-reports>

Previous version:

<https://www.w3.org/TR/2017/PR-activitystreams-core-20170413/>

Editors:

[James M Snell](#), IBM

[Evan Prodromou](#), Fuzzy.ai

<https://www.w3.org/TR/activitypub/>
<https://www.w3.org/TR/activitystreams-core/>

Event Notifications in Value-Adding Networks

Event Notifications in Value-Adding Networks

Living Document, 6 September 2023

This version:

<https://www.eventnotifications.net>

Latest published version:

<https://www.eventnotifications.net>

Previous Versions:

<https://www.eventnotifications.net/0.1/>

Feedback:

[Inline In Spec](#)

Editors:

[Patrick Hochstenbach](#) (Ghent University Library)

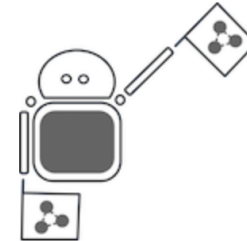
[Miel Vander Sande](#) ([meemoo](#) - Flemish Institute for Archives)

[Ruben Dedecker](#) (IDLab - Ghent University)


[Paul Walk](#) ([Antleaf](#))

[Martin Klein](#) (Los Alamos National Laboratory)

[Herbert Van de Sompel](#) (IDLab - Ghent University)




authorIDy: Listing Contributions by Contributor Identifier

 Home Conventions Patterns ▾ FAIR FAIRiCat Signmap authorIDy Adopters About Feedback

authorIDy: Listing Contributions by Contributor Identifier

Warning: Discussion document - Not a specification (yet)

This version, created 20240224: <https://signposting.org/authorIDy/> 

Check out [the experimental demonstration of the authorIDy interface](#) as described in this document

Please provide feedback in the [GitHub Signposting repository](#), using label authorIDy

Increasingly, repositories allow, recommend, or require providing unique contributor identifiers, such as ORCID or ISNI, when depositing contributions. Despite this evolution, few repositories provide a machine interface that allows listing the contributions made by a researcher using their contributor identifier as a key. An interface with this capability, supported across repositories, would facilitate a range of use cases and could be an important next step towards realizing [COAR's Next Generation Repositories](#) vision. This document describes core characteristics of authorIDy, a conceptual proposal for such a machine interface. Hopefully, this description will evolve into a specification.

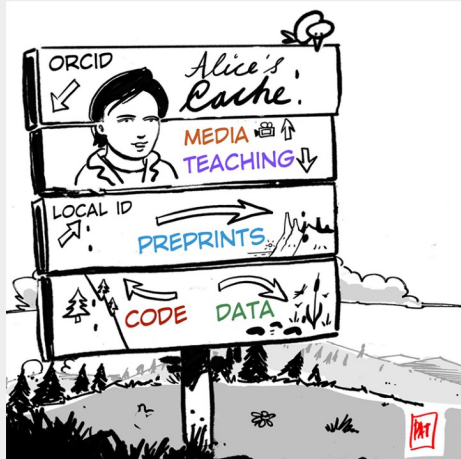
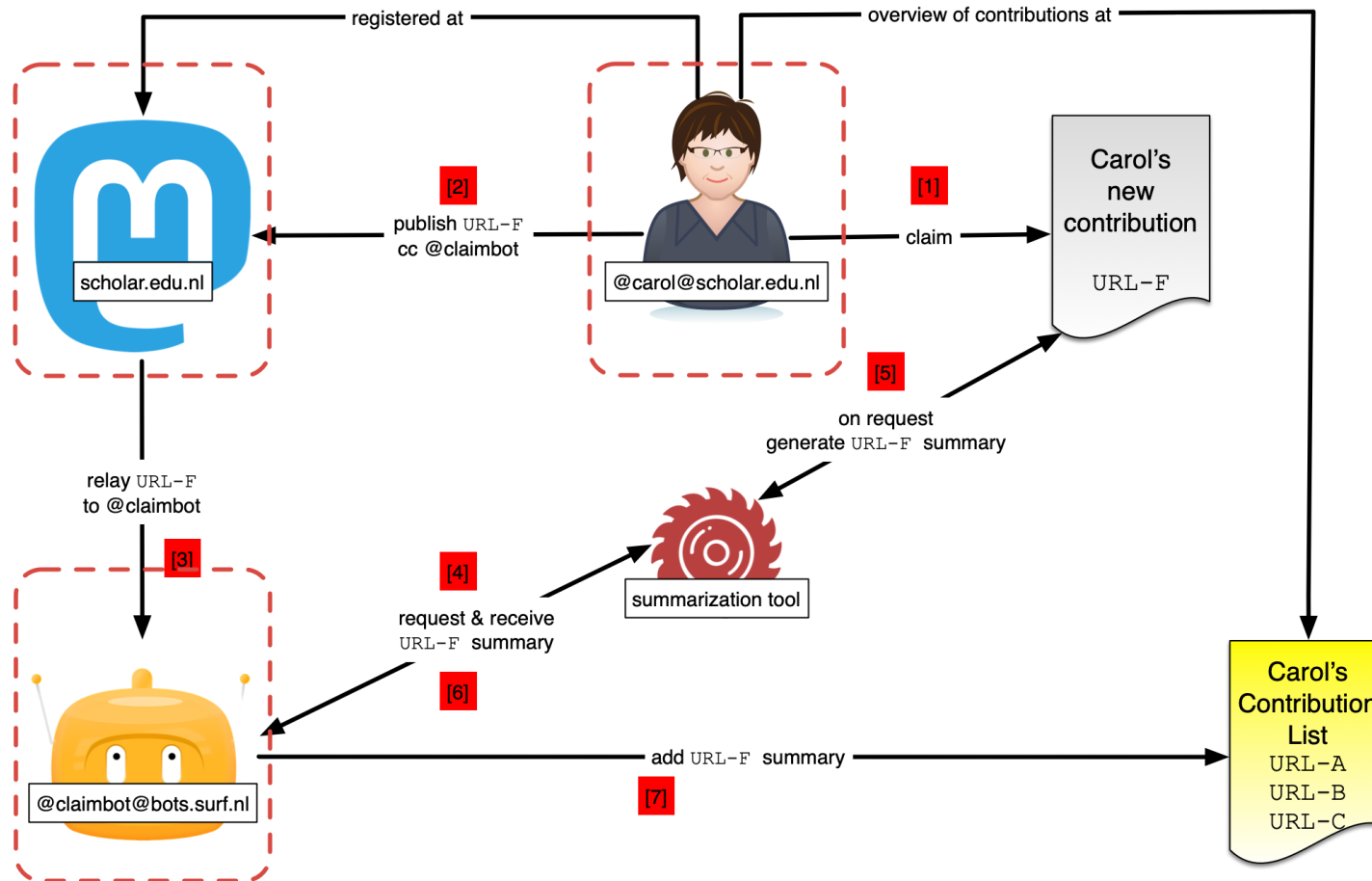


Image courtesy of [Patrick Hochstenbach](#).

authorIDy: Listing Contributions by Contributor Identifier (2024)
<https://signposting.org/authorIDy/>

Identity Verification





Identity Verification

- Mastodon natively supports rel="me" identity verification
 - demonstrates that the user with a given Mastodon account has editing rights to a given webpage
- Additional *academic* verification is required:
 - to allow extending the approach to other Mastodon instances (that don't have something like SURFconext identification)
 - to allow the claim bot to determine whether it will accept to process a claim
 - to be able to attribute informal scholarly communication on Mastodon to a person with verified academic affiliation



InAcademia Experiment

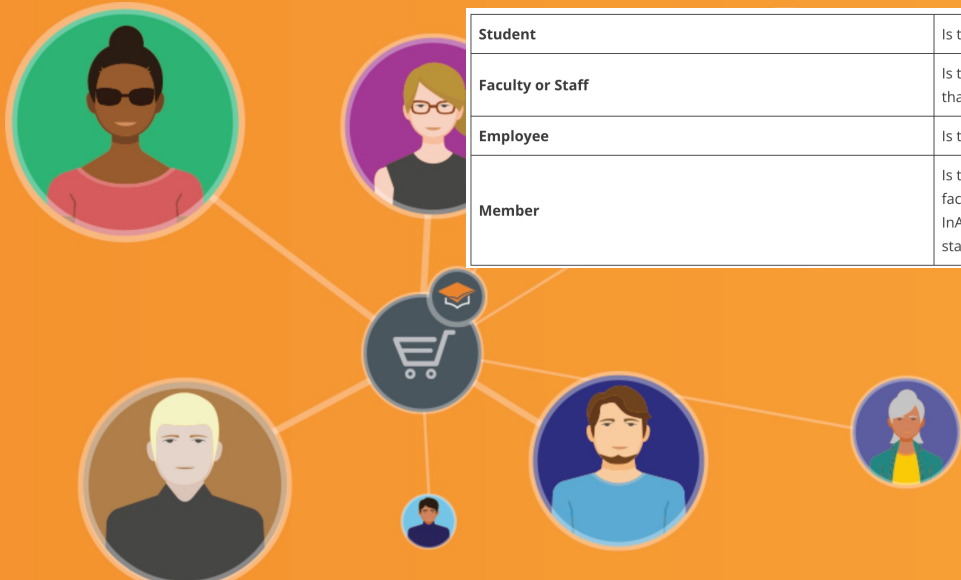


ABOUT INACADEMIA ▾ TECHNICAL OVERVIEW ▾ SUPPORT ▾ FAQs STUDENT DISCOUNT FOR WOOCOMMERCE ▾ CONTACT US BLOG Q

in

Welcome to InAcademia

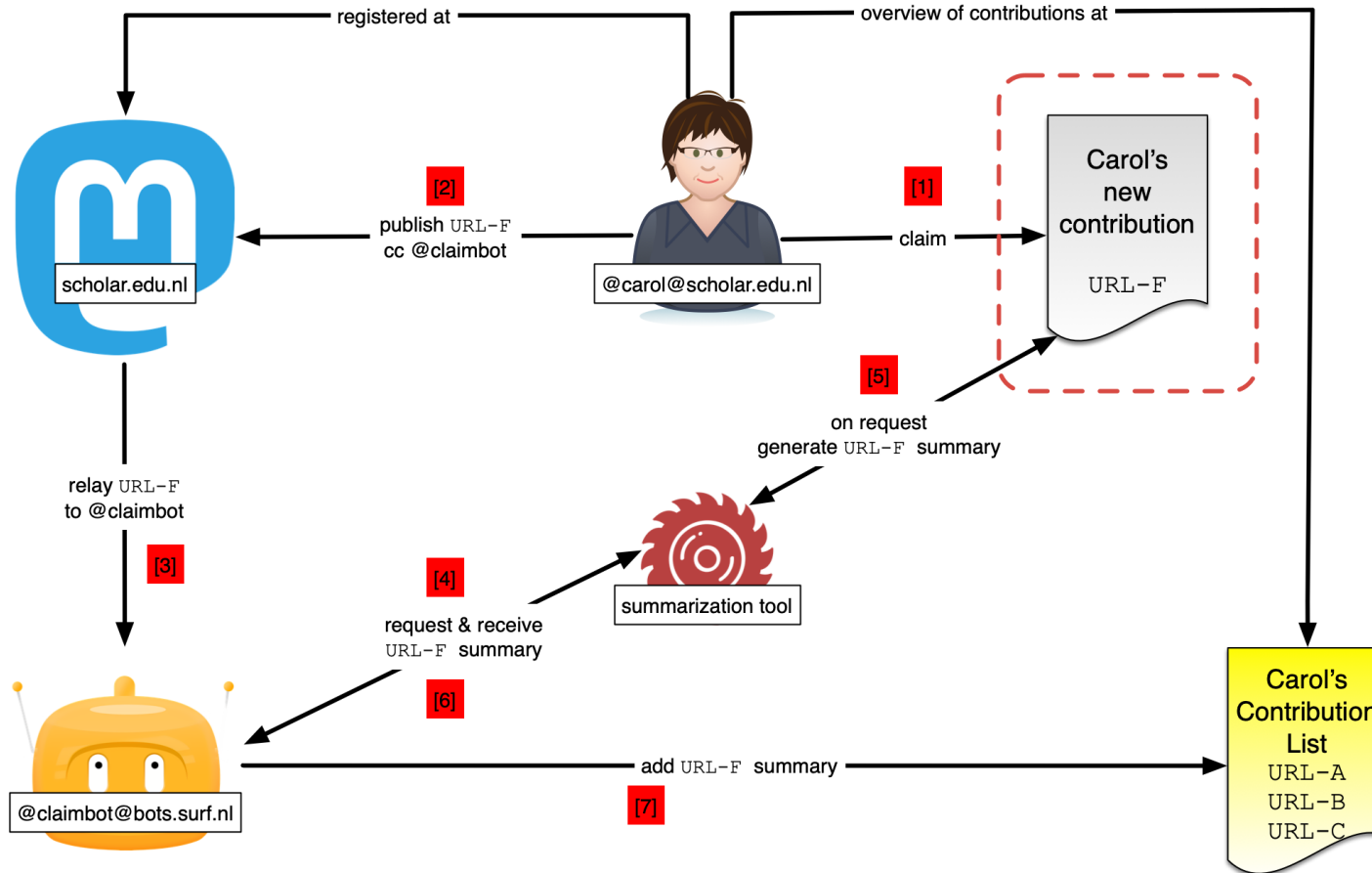
InAcademia provides real-time, secure validation of student affiliation, so retailers can validate eligibility for special offers cost-effectively while preserving user privacy.



Student	Is the end-user a student at the institution?
Faculty or Staff	Is the end-user a teacher/researcher (faculty) or a worker (other than teacher/researcher, staff) at the institution?
Employee	Is the user an employee at the institution?
Member	Is the end-user affiliated to the institution (i.e. either a student, faculty member, staff or employee)? Please note that in this case, InAcademia will not distinguish whether the user is a student, staff, faculty member or employee.

<https://inacademia.org>

Trusted Archiving



Trusted Archiving

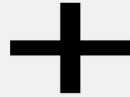



- As a means to compensate for
 - link rot and content drift
 - metadata extraction challenges
- Not necessarily archiving the contribution itself, but (at least) a trace of the contribution
 - The URL that was tooted
 - Think: difference between landing page and the actual contribution linked from the landing page



myresearch.institute – Landing Page for Contribution

My Research Institute

Herbert Van de Sompel




Artifact

Artifact URLs:

[Embed](#) <https://github.com/dret/I-D/issues/106#issuecomment-439398379>

Published	2018-11-16T13:50:01Z
Researcher	Herbert Van de Sompel
ORCID	https://orcid.org/0000-0002-0715-6126
Portal	Github
User	hvdsonp



Add a before-sunset link relation type · Issue #106 · dret/I-D

Preserved by [SCHOLARLYORPHANS.ORG](#)

While the sunset link relation type allows us to convey what would be the upcoming URI of the current resource (or the resource explicitly provided as the context using the anchor attribute), there

[github.com](#) @ 2018-11-19T12:50:31Z

[Other Versions](#) || [Current version](#)

Resources Captured

Mementos:

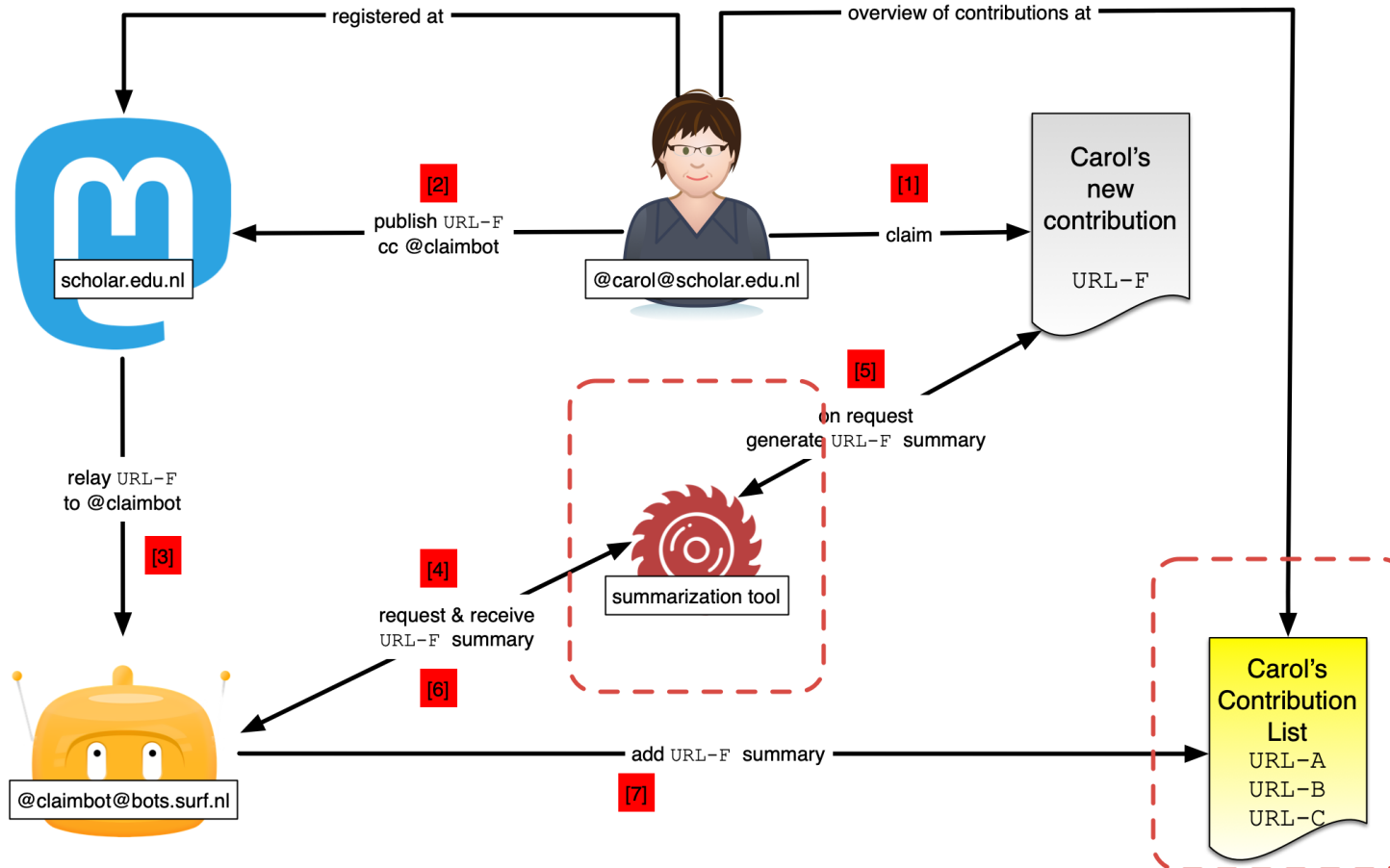
- <https://scholarlyorphans.org/memento/20181119125031/https://github.com/dret/I-D/issues/106#issuecomment-439398379>

WARCs:

- <https://myresearch.institute/capture/warc/WARCPROX-20181119125031432-00000-evw6kc9z.warc>

<https://myresearch.institute/event/7644e5d859e84360895d34a31573ee4f/>

Summarization Challenges



Summarization Challenges - Metadata Extraction

- Experiment using URLs from toots by 100 researcher as research contribution claims
- Total URLs 555; for 54 no metadata could be extracted
- Success rate for metadata extraction depends on the type of web resource for which metadata is scraped

	title	abstract	language	author	date published
webpage	335 = 99%	152 = 45%	144 = 42%	44 = 13%	118 = 35%
post weblog	40 = 100%	34 = 85%	35 = 88%	25 = 63%	33 = 83%
motion picture	11 = 100%	9 = 82%	0	11 = 100%	11 = 100%
article- newspaper	23 = 100%	23 = 100%	23 = 100%	19 = 83%	23 = 100%
article- journal	56 = 100%	48 = 86%	47 = 84%	56 = 100%	56 = 100%

Summarization Challenges – Closed Repositories

A Word of Caution

- Imagine Notify, Event Notifications, Signposting, etc. becoming hugely successful
- Imagine zillions of actions by bots involved in notifications, automated interactions, ...
- Such a success will undoubtedly attract unethical bots
- Add that to the alarming problems that repositories already face because of voracious and unethical LLM bots
- ... we are at risk that open repositories close their doors to machine interactions



COAR Annual Meeting

Göteborg, June 7 2024



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025



Summarization Challenges – Closed Repositories



herbert

@hvdsomp@w3c.social

Remember how the origins of #FAIR are all about machine access to scholarly objects? And how the A stands for Accessible? Well, despite all the megaphoning about FAIR, scholarly objects are rapidly becoming less accessible to machines. Because repositories are increasingly blocking machine access in order to protect themselves from aggressive AI bots. To achieve FAIR, an urgent focus on addressing this challenge is needed. And, no, PIDs are not the solution however much you'd love them to be.

Mar 13, 2025, 07:00 PM

🌐 · Web

29 boosts · 15 favorites



@hvdsomp@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

Aggressive AI Bots



go to hellman

- Ignore robots.txt
- Exhaust server connection space
- Use large blocks of IP addresses
- Use randomly generated, spoofed, user agent strings

AI bots are destroying Open Access

Aggressive AI Bots – Not Just Scholarly Repositories



IMPROVEMENT

MAY 8, 2025 • 1 MINUTE READ

Updated rate limits for unauthenticated requests



Tommaso Gagliardini
@tomgag@infosec.exchange

In a move that surprises absolutely noone, GitHub now requires users to login in order to browse public repositories (including open source projects). After a few (~10) requests, you get blocked (I can confirm). In order to fight AI scrapers, I guess.

So, GitHub decided to blanket-limit access to open source projects as a defense against the very scourge that they(r parent company) unleashed on the world.

I won't be hypocrite: it's a bit embarrassing, but undeniably satisfying to say "told you so". I moved away from GitHub long ago and I moved all my stuff to Codeberg instead. And so happy I did!

Next step: radicle.xyz maybe?

[github.com/orgs/community/disc...](https://github.com/orgs/community/discussions)

Whoa there!

Hide

You have exceeded a secondary rate limit.

Please wait a few minutes before you try again;

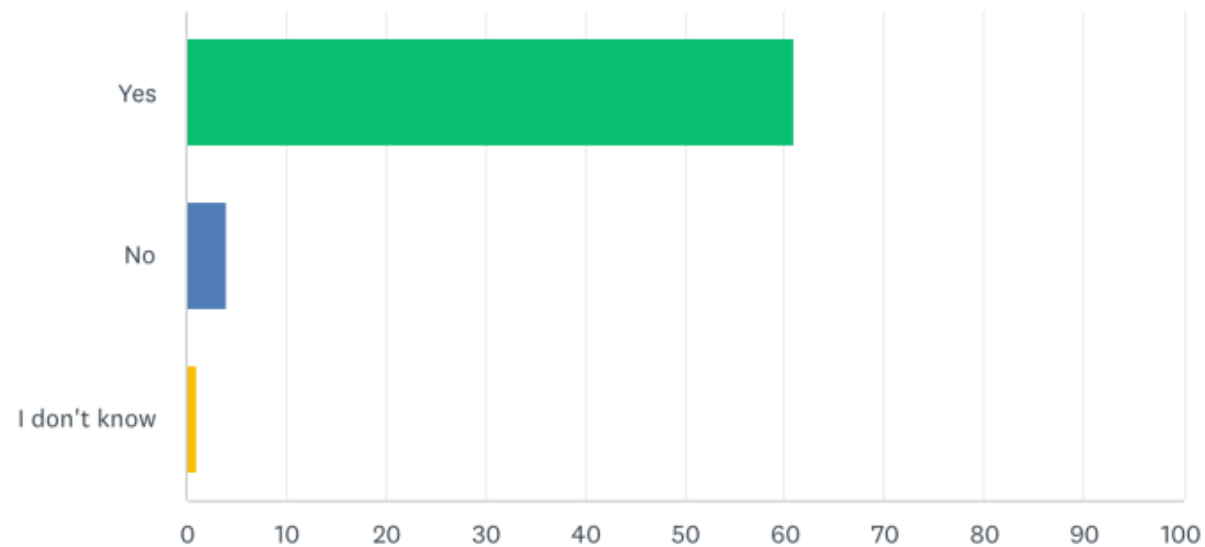
<https://infosec.exchange/@tomgag/114504958452945090>

<https://github.blog/changelog/2025-05-08-updated-rate-limits-for-unauthenticated-requests/>

Aggressive AI Bots – COAR Survey

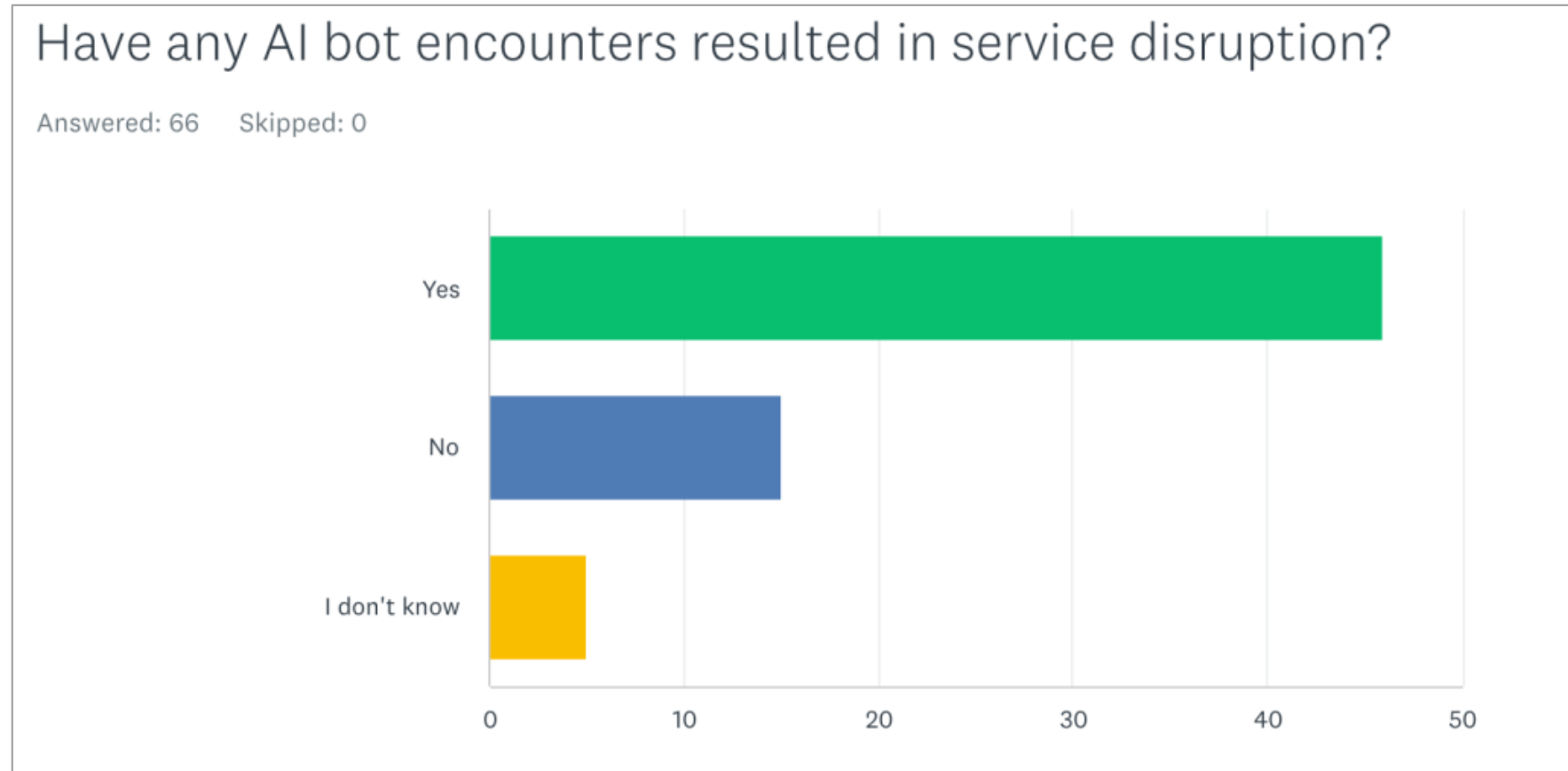
Is your repository encountering AI bots that are scraping or extracting your content?

Answered: 66 Skipped: 0



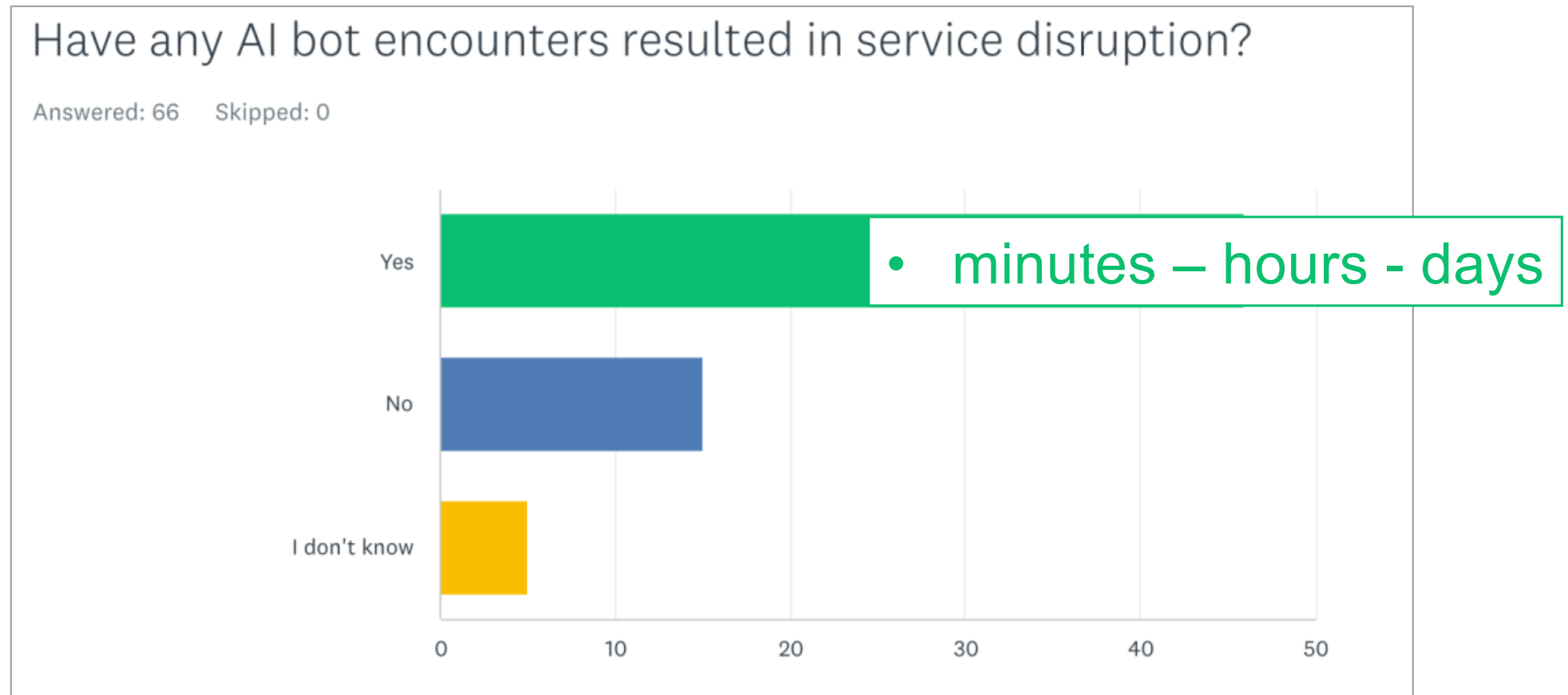
<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

Aggressive AI Bots – COAR Survey



<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

Aggressive AI Bots – COAR Survey

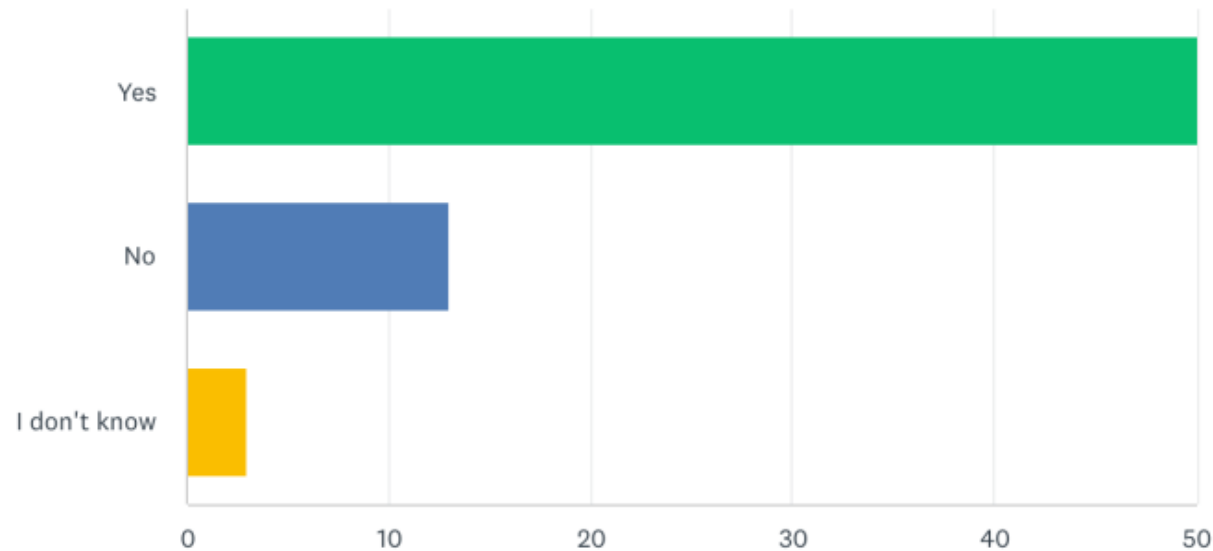


<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

Aggressive AI Bots – COAR Survey

Have you applied any measures to minimize or stop access to the repository by AI bots?

Answered: 66 Skipped: 0

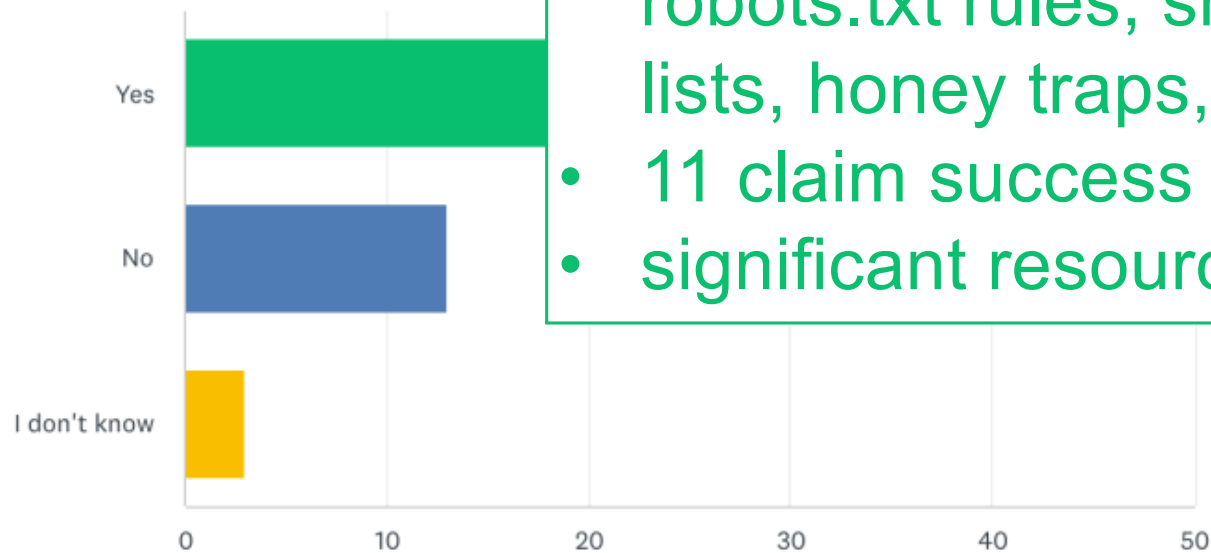


<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

Aggressive AI Bots – COAR Survey

Have you applied any measures to minimize or stop access to the repository by AI bots?

Answered: 66 Skipped: 0

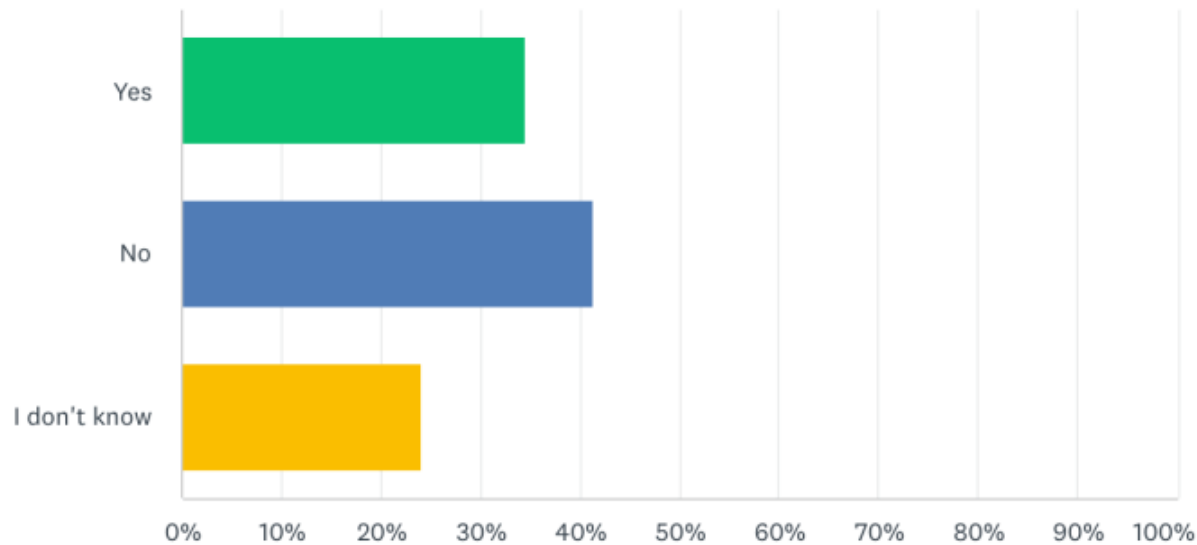


- rate-limiting, firewall rules, robots.txt rules, shared allow-lists, honey traps, cloudflare
- 11 claim success
- significant resource investment

Aggressive AI Bots – COAR Survey

If yes, do these measures also inhibit access to your repository by “friendly” services?

Answered: 58 Skipped: 8

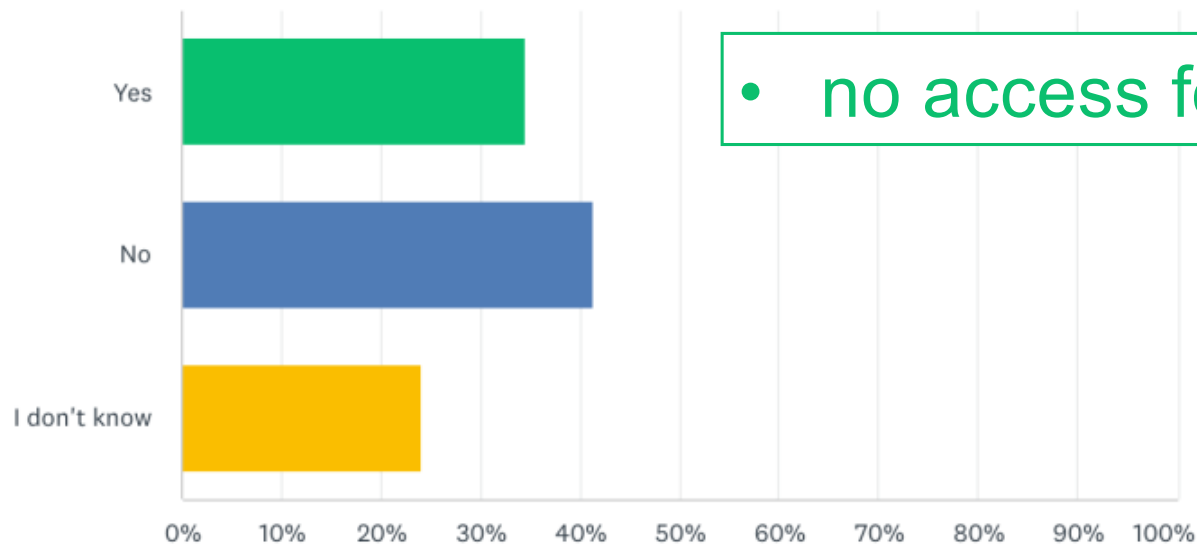


<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>

Aggressive AI Bots – COAR Survey




If yes, do these measures also inhibit access to your repository by “friendly” services?

Answered: 58 Skipped: 8





- no access for friendly bots

Summarization Challenges – Bot Impersonates a Human

 README  MIT license 

curl-impersonate

 [Build and test](#)  [Docker images](#)

A special build of [curl](#) that can impersonate the four major browsers: Chrome, Edge, Safari & Firefox. curl-impersonate is able to perform TLS and HTTP handshakes that are identical to that of a real browser.

curl-impersonate can be used either as a command line tool, similar to the regular curl, or as a library that can be integrated instead of the regular libcurl. See [Usage](#) below.

<https://github.com/lwthiker/curl-impersonate>

Aggressive AI Bots – Ongoing/Planned Activities

- COAR Task Force on Managing AI Bots in Repositories
 - gain understanding of current approaches to handle AI bots
 - review potential solutions aimed at keeping repositories open
 - develop recommendations for the repository community
- IETF AI Preferences
 - expressive extension of robots.txt
 - legally binding?
- IETF web bot authentication
 - more reliable bot identification than e.g. IP address, user agent
 - using certificate infrastructure
 - one-off bot access?

New Open Science Contribution?

Toot it to Claim it!



claimbot



mastodon



summarizer



Carol

Herbert Van de Sompel

[@hvdSompel@w3c.social](https://hvdSompel@w3c.social) - <https://hvdSompel.info>



@hvdSompel@w3c.social - Toot It to Claim It
DANS, The Hague – May 19 2025

