



QUEEN'S
UNIVERSITY
BELFAST

CSIT

CENTRE
FOR SECURE
INFORMATION
TECHNOLOGIES

UK
RI



Delivered by
Innovate UK
and EPSRC



UK Government
Northern Ireland



Invest
Northern
Ireland

Exploring global academic repositories for software

Dr. Domhnall Carlin
EPSRC Research Software Engineering Fellow

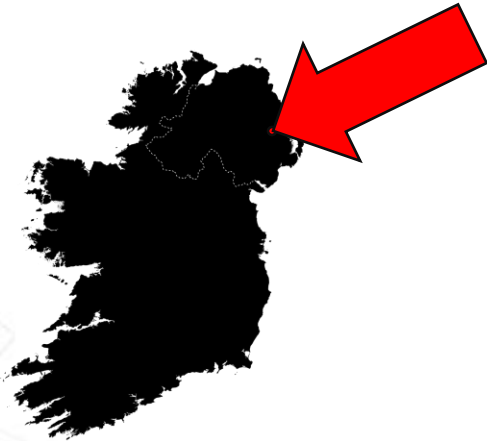
Queen's University Belfast
N.Ireland

The logo for the Open Repositories 2025 conference, featuring a white silhouette of a city skyline (Chicago) against a black background.

OPEN REPOSITORIES 2025
THE 20TH INTERNATIONAL CONFERENCE * * * * CHICAGO, IL, USA

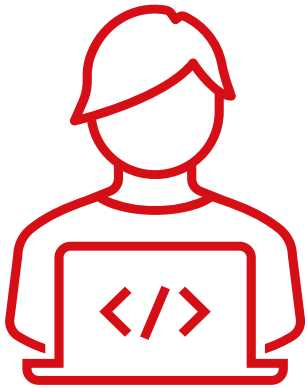
QUB

- Est. 1845
- 25,000 students
- Russell Group member
- Had no RSE function



Fellowship aims

- To raise awareness of the value of research software in an open science workflow
- To help researchers gain recognition for code they write



As a researcher, I should:

- know I can publish my research software
- within my institute's guidelines
- with my institute's research repository.

Background

- REF: The UK's Research Excellence Framework
- Assesses the quality and impact of higher education research in the UK
- 6-7 years
- Is kind of a big deal
- Software returns to REF plummeted to 13 in the last round (2021) from >115 (2008)

Background

- Recent work examined 182 academic institutional repositories (IRs) from 157 universities in the UK ^[1]
- Found:
 - ~60% **did not** recognise software as a discrete type of output
 - ~10% **did**, but had no records
 - ~30% **had** records of software as distinct outputs of research process

[1]Carlin, D, Rainer, A & Wilson, D 2023, 'Where is all the research software? An analysis of software in UK academic repositories', PeerJ Computer Science, vol. 9, e1546. <https://doi.org/10.7717/peerj-cs.1546>

Background

- The most popular OS repository platform could enable 'software' as a type with a single word in a config file
- It's not a **technical** problem
- Some IRs appeared to have had defaults overridden

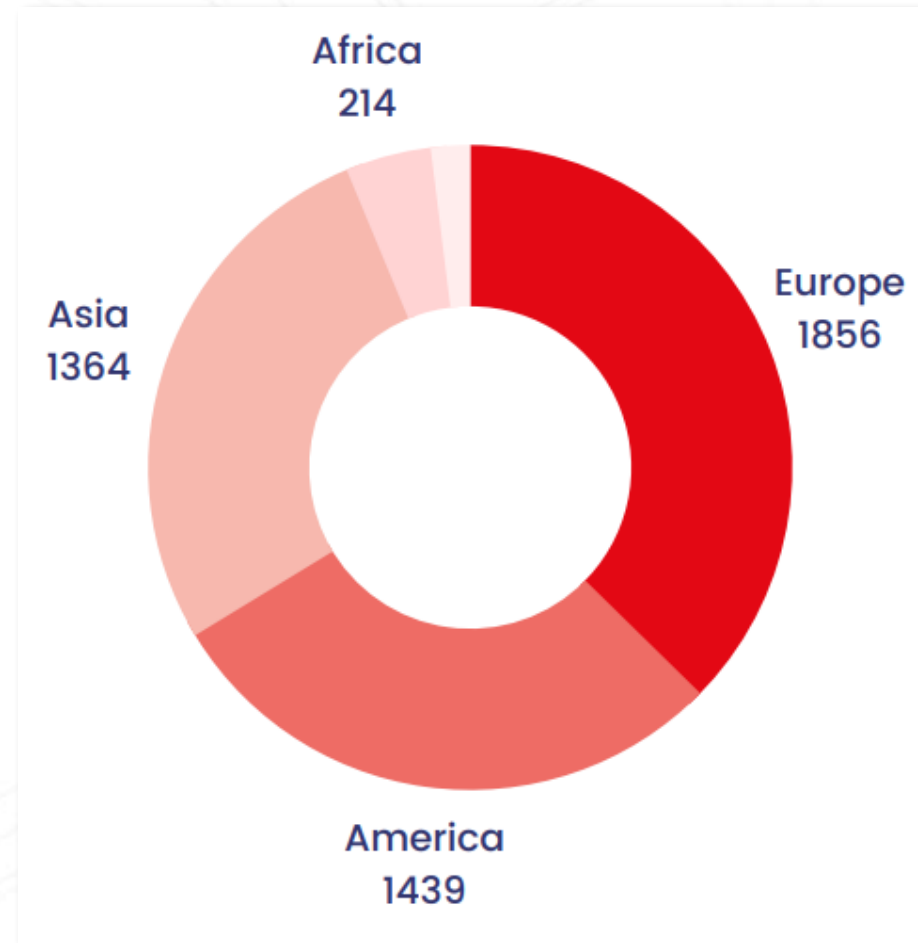
Research Questions

- 1) What is the prevalence of research software in IRs in a global sample?
- 2) Does this replicate the work by [1] in a larger sample?
- 3) Does a larger global sample provide greater insight into the recognition of research software?
- 4) Do any factors in the underlying RIS platforms provide insight into usage?

Method

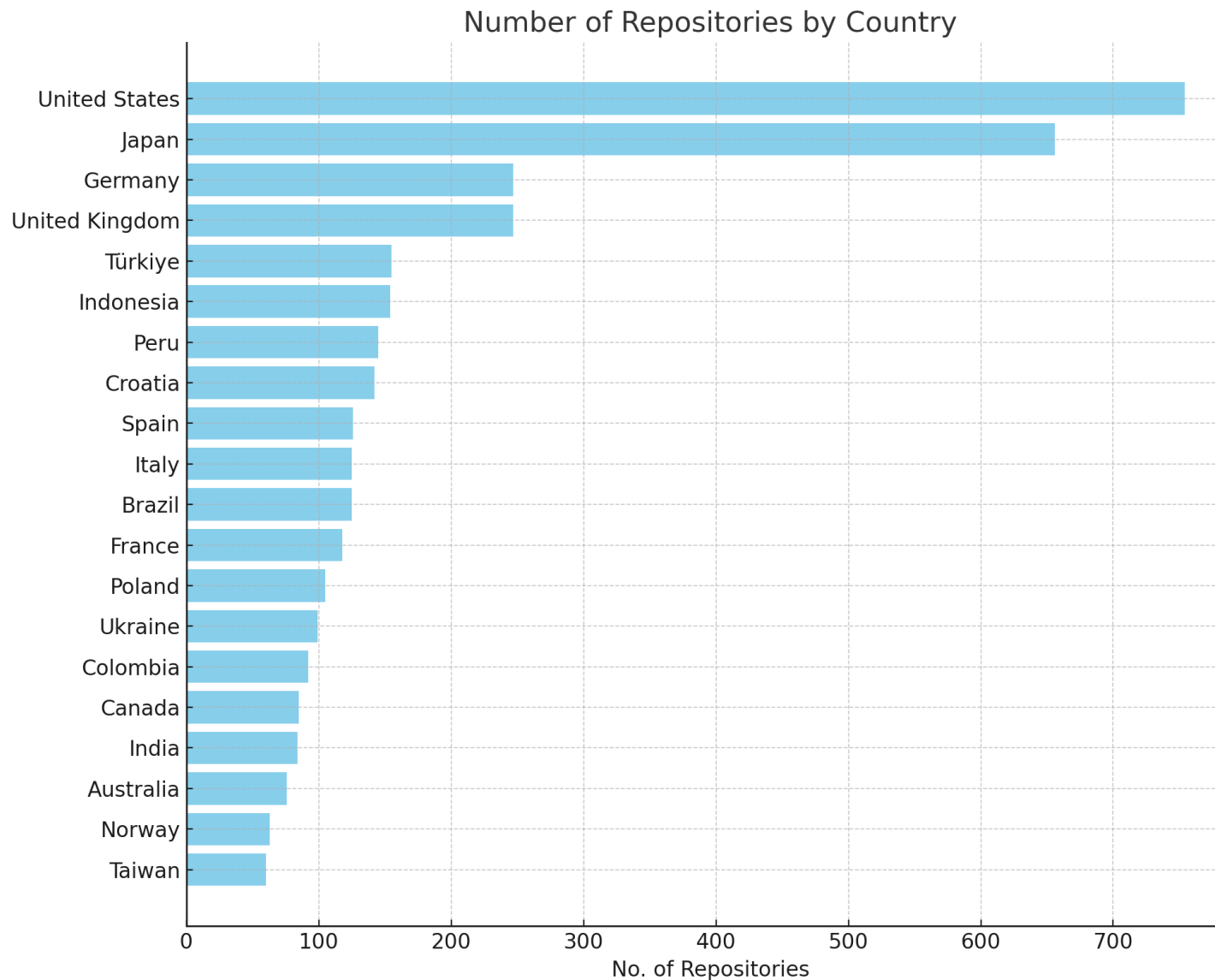
- Querying the IR of every university in the world is a mammoth task (25k universities?)
- Employed an existing list of IRs
- OpenDOAR via Sherpa API from JISC
 - Quality-assured, global Directory of Open Access Repositories.
 - Filtered using 'institutional' entries
- This yielded a list of URLs with associated metadata

Locations



Locations

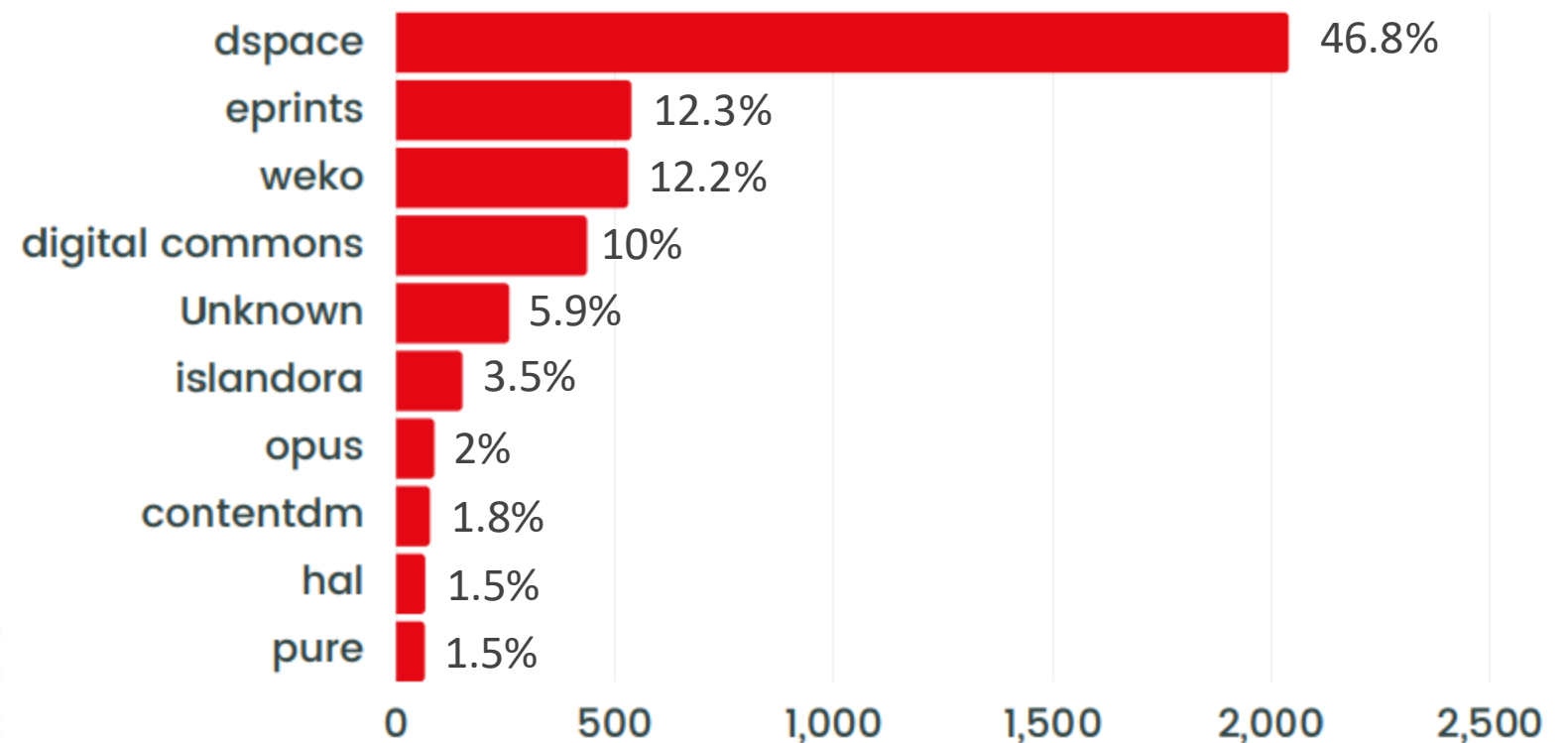
- IRs from 125 different countries
- Top 20:



RIS platforms

- ~170 RIS platforms
- Messy data

Top 10 RIS platforms



Method

- OpenDOAR provided a list of URLs with metadata
- not counts of software records*
- Manually visited each of ~4970 URLs

Categories

Does not contain software

- Software exists as searchable item
- Has no records
- Note *records*

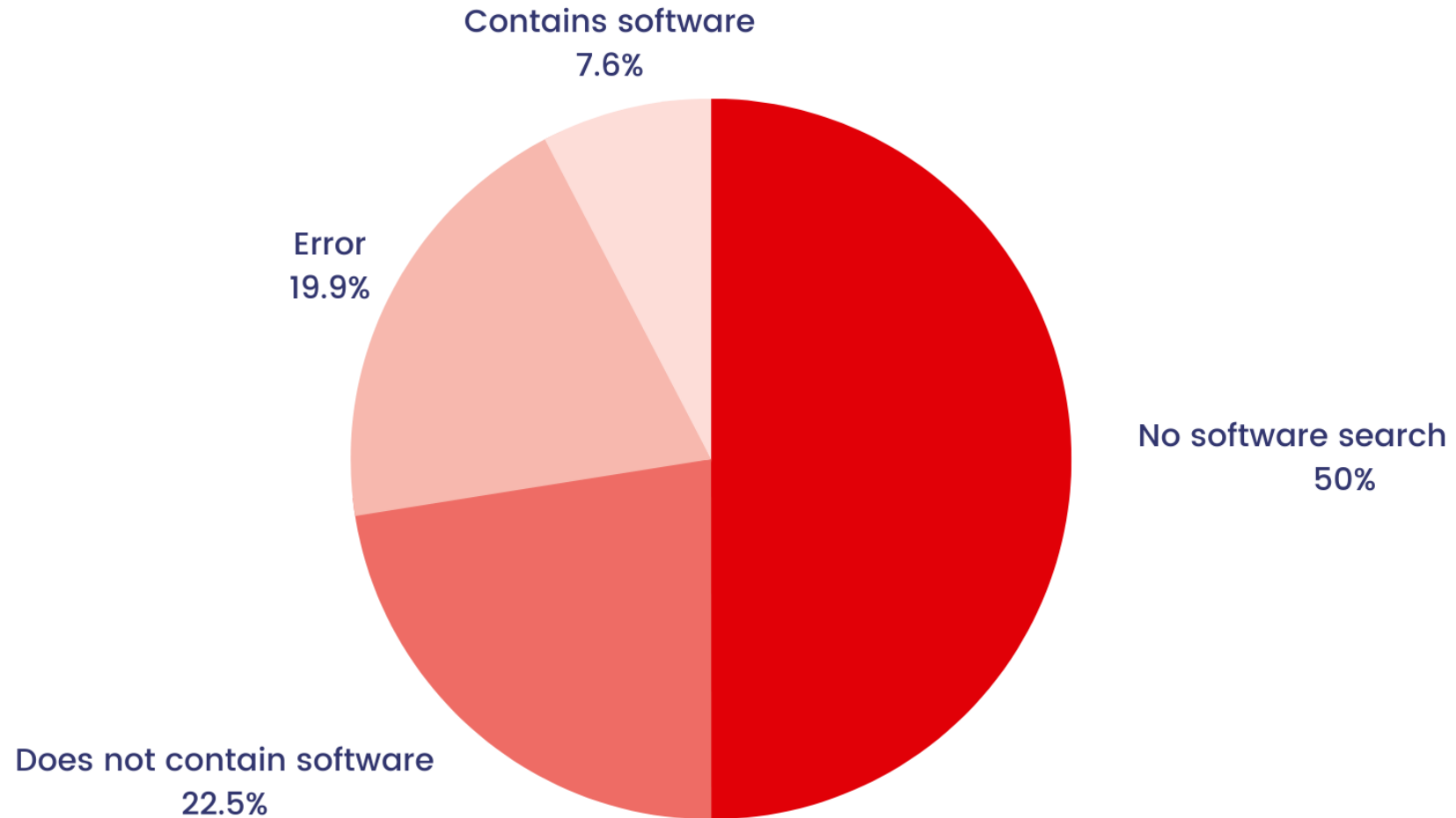
Contains software

- Software exists as searchable item
- Has records

No software search capability

- Software doesn't exist as searchable item
- Does not mean there is no software (e.g. defined as other)

Results: research software



Changes over time

Feb'24

13047

May'25

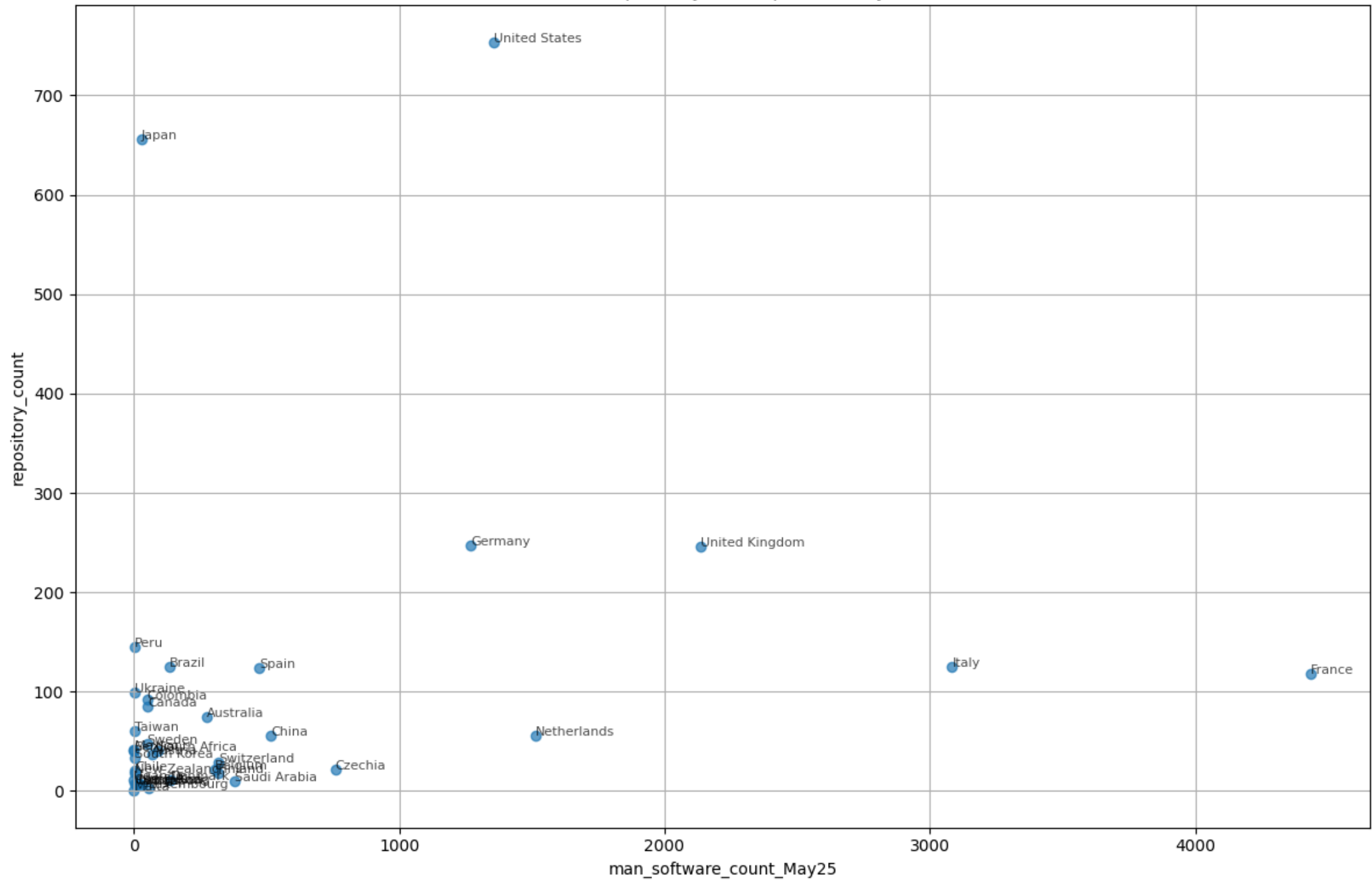
17919

+37.34%

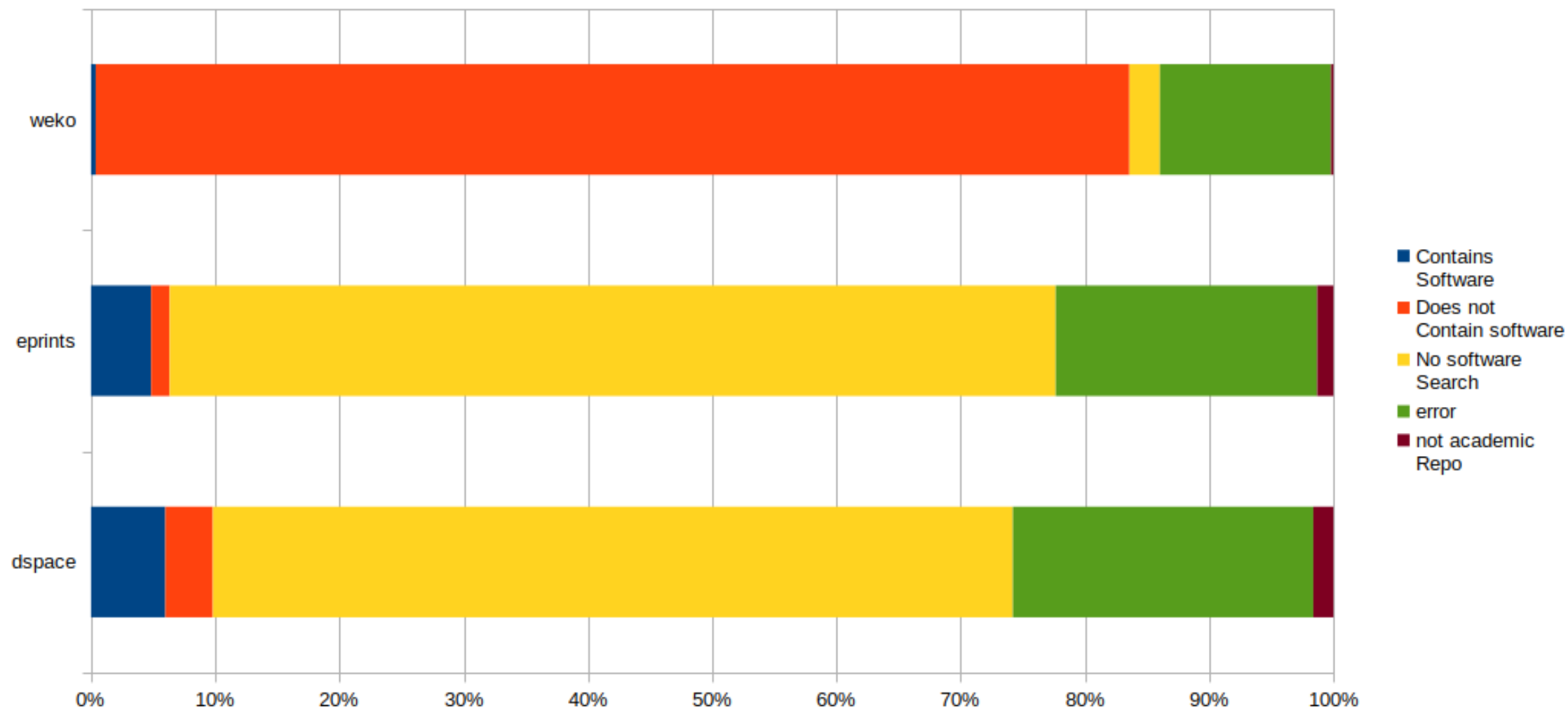
- Range: **-24** - **+889**
- 1828 - 1

	Software May25	Software Feb24	% change
France	4437	2260	96.327434
Italy	3084	2848	8.286517
United Kingdom	2138	1627	31.407498
Netherlands	1516	970	56.288660
United States	1356	1298	4.468413
Germany	1270	768	65.364583
Czechia	759	713	6.451613
China	517	136	280.147059
Spain	473	346	36.705202
Saudi Arabia	379	364	4.120879
Switzerland	320	211	51.658768
Finland	319	311	2.572347
Belgium	305	254	2078740
Australia	274	270	1.481481

Software Count vs Repository Count per Country (>0 software)



Results: research software in top 3 RIS



Discussion

- Of IRs sampled:
 - ~50% did not have software records as a discrete type of output
 - ~30% could contain software records
 - ~7.6% did contain records of software
- DSpace accounts for ~40% of IRs, yet:
 - >64% had no search item for software
- EPrints: 6% had software term, but isn't default

Discussion

- Using automation to replicate/maintain this will be difficult
- It was difficult to do manually:
 - Some IRs are really well hidden
 - Links in OpenDOAR can often be to library
 - Issues with translations – e.g. 'program' being translated as software, when it meant 'course of study'
 - Use of institutional data repositories for software, rather than main IR
 - Use of free-text search, rather than items
- Having software as an item does not mean it is used (e.g. Weko)

Open questions

- Does this speak to how software is considered within the research process?
- Does it matter?
- IRs are not development-friendly
- Is operating as a registry for software the best hope?
 - Linkable to development repositories
 - Ensures software is 'seen' by the university
- What would this workflow look like?

Acknowledgments

- With thanks to the organising & program committees of OR25
- Work supported by UKRI EPSRC Research Software Engineering Fellowship (EP/V052284/1).



**QUEEN'S
UNIVERSITY
BELFAST**

CSIT

**CENTRE
FOR SECURE
INFORMATION
TECHNOLOGIES**

github.com/DSpace/DSpace/blob/main/dspace/config/submission-forms.xml#L1461

DSpace / dspace / config / **submission-forms.xml**

Code

Blame

1991 lines (1943 loc) · 85.5 KB

```
1451         </pair>
1452         <pair>
1453             <displayed-value>Recording, musical</displayed-value>
1454             <stored-value>Recording, musical</stored-value>
1455         </pair>
1456         <pair>
1457             <displayed-value>Recording, oral</displayed-value>
1458             <stored-value>Recording, oral</stored-value>
1459         </pair>
1460         <pair>
1461             <displayed-value>Software</displayed-value>
1462             <stored-value>Software</stored-value>
1463         </pair>
```