University of New England

Using data mining techniques to enhance repository collections

CAUL Research Repositories Community Event, Sydney. 28–29 October, 2019.

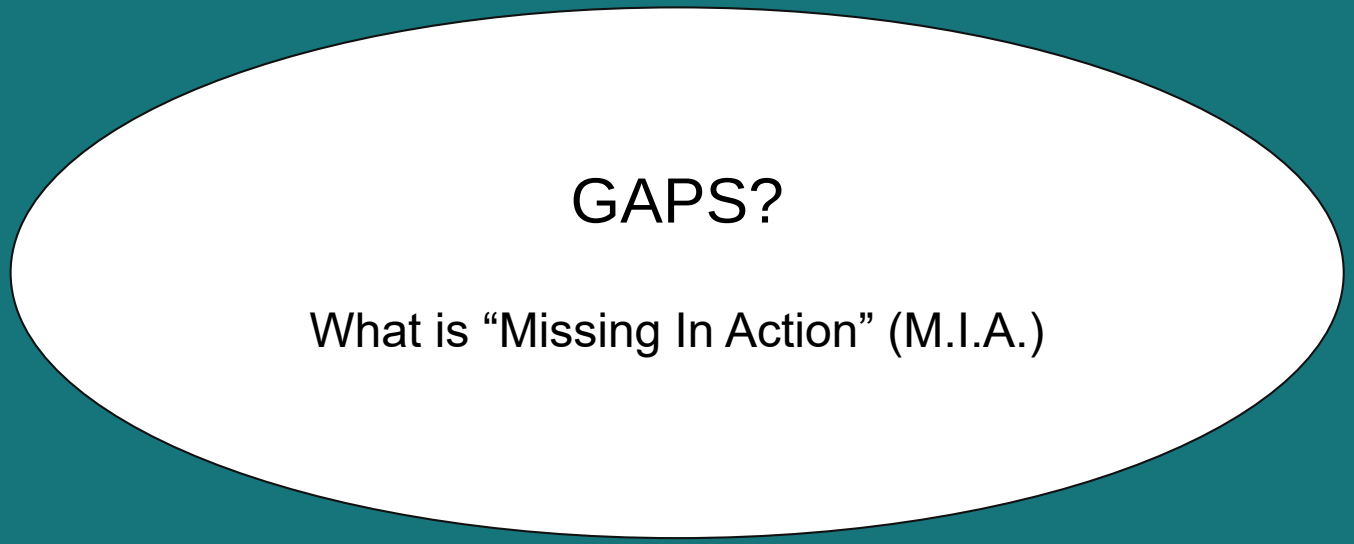
Allison Brown & Matt Watts, University of New England.

Creative Commons Attribution (CC BY)

Gap analysis background

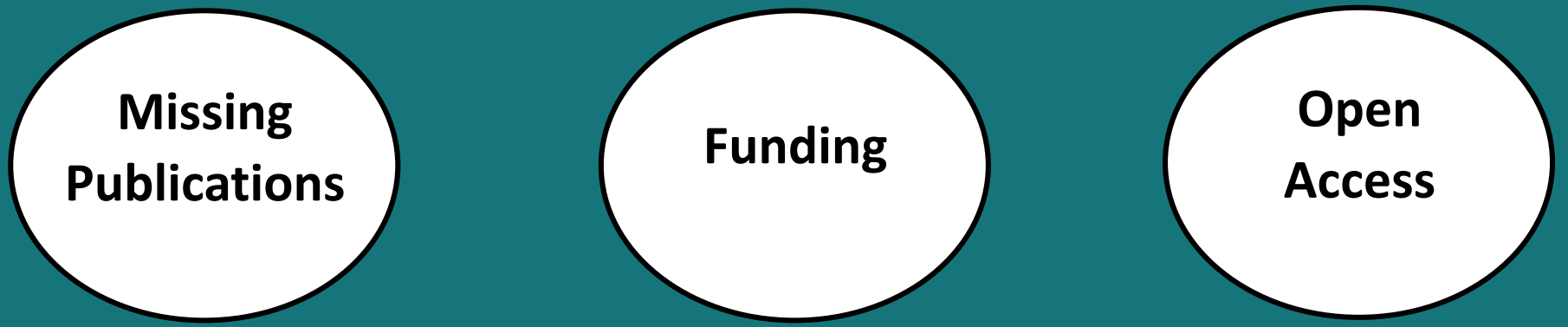
- Context of RUNE (repository) and UNE
- What is a gap analysis?
- Why we did a gap analysis?

Gap(s)



une
University of
New England

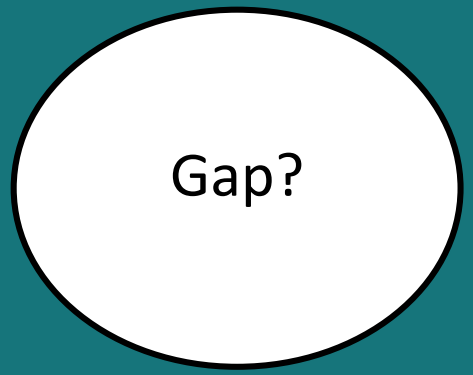
Gaps UNE Identified



Activity 1

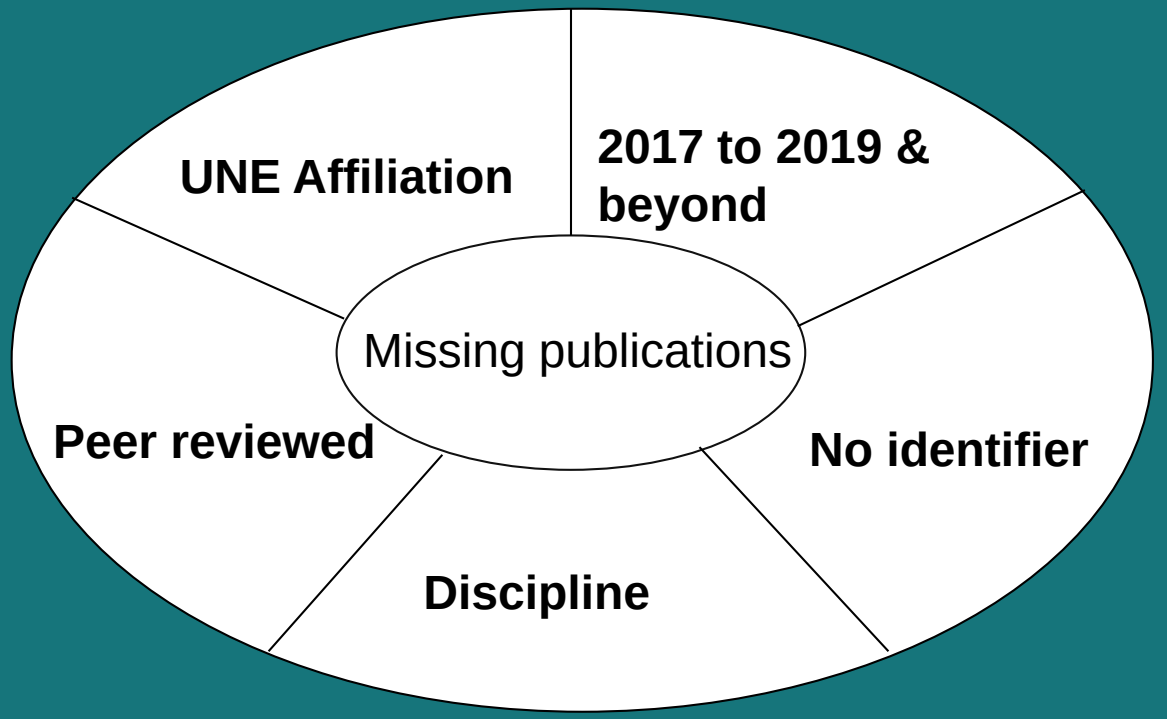
Working in groups or by yourself, identify a *possible gap* in your repository.

Write this gap in the middle of the mind-map.



une
University of New England

Factors


UNE Affiliation
2017 to 2019 & beyond
Missing publications
Peer reviewed
No identifier
Discipline

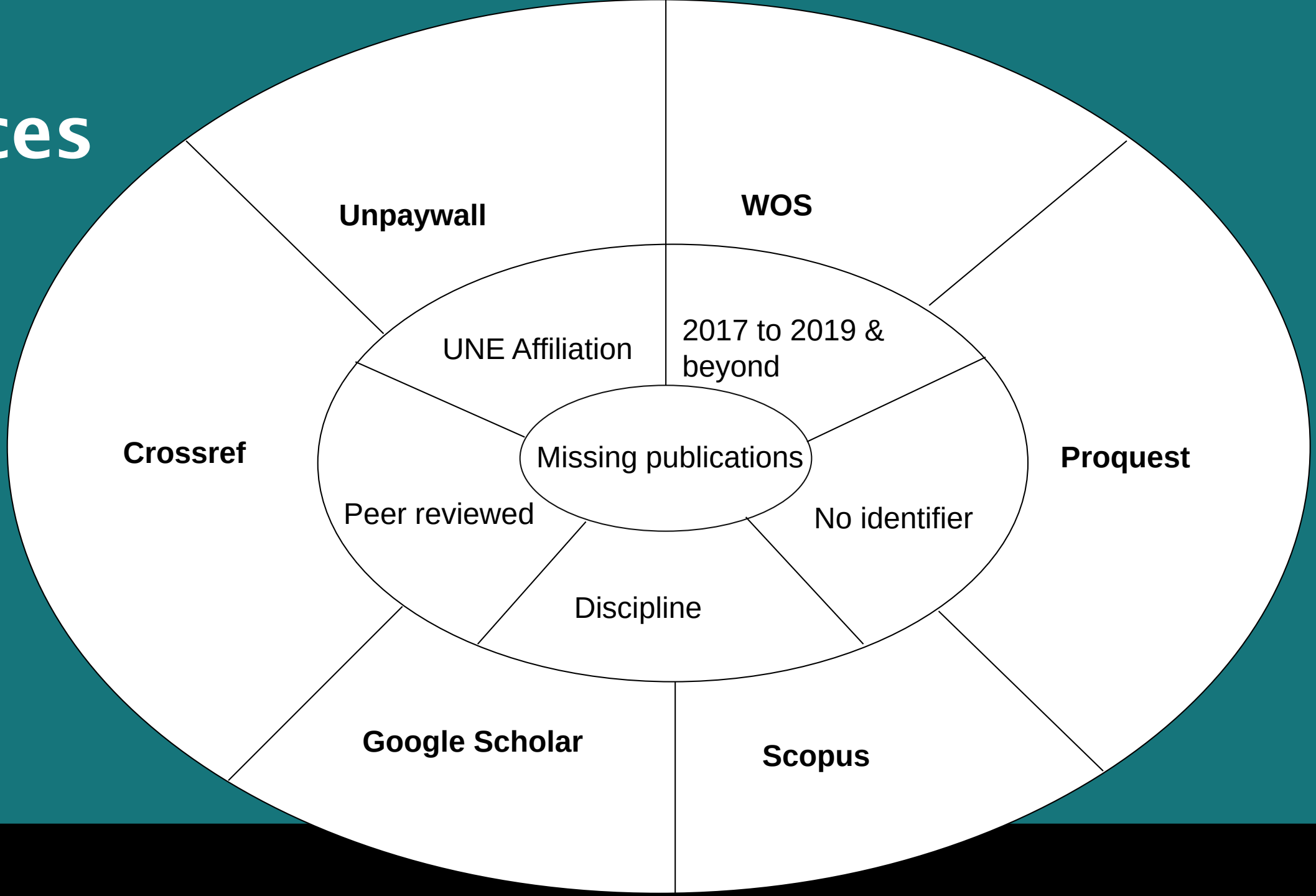
Activity 2

Working in the same groups, brainstorm different *factors* of the identified gap.

Add potential factors around the second ring.

University of New England

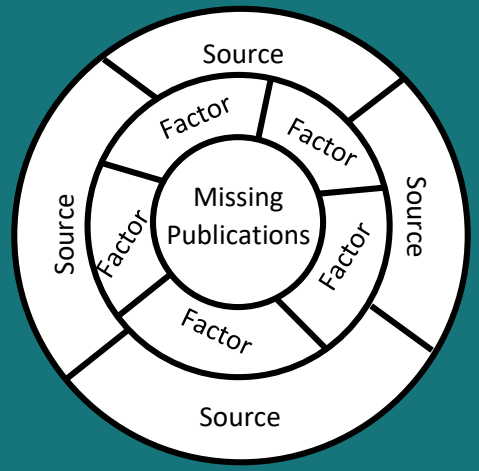
Sources


Unpaywall
WOS
Crossref
Proquest
Google Scholar
Scopus
UNE Affiliation
2017 to 2019 & beyond
Peer reviewed
No identifier
Discipline
Missing publications

Activity 3

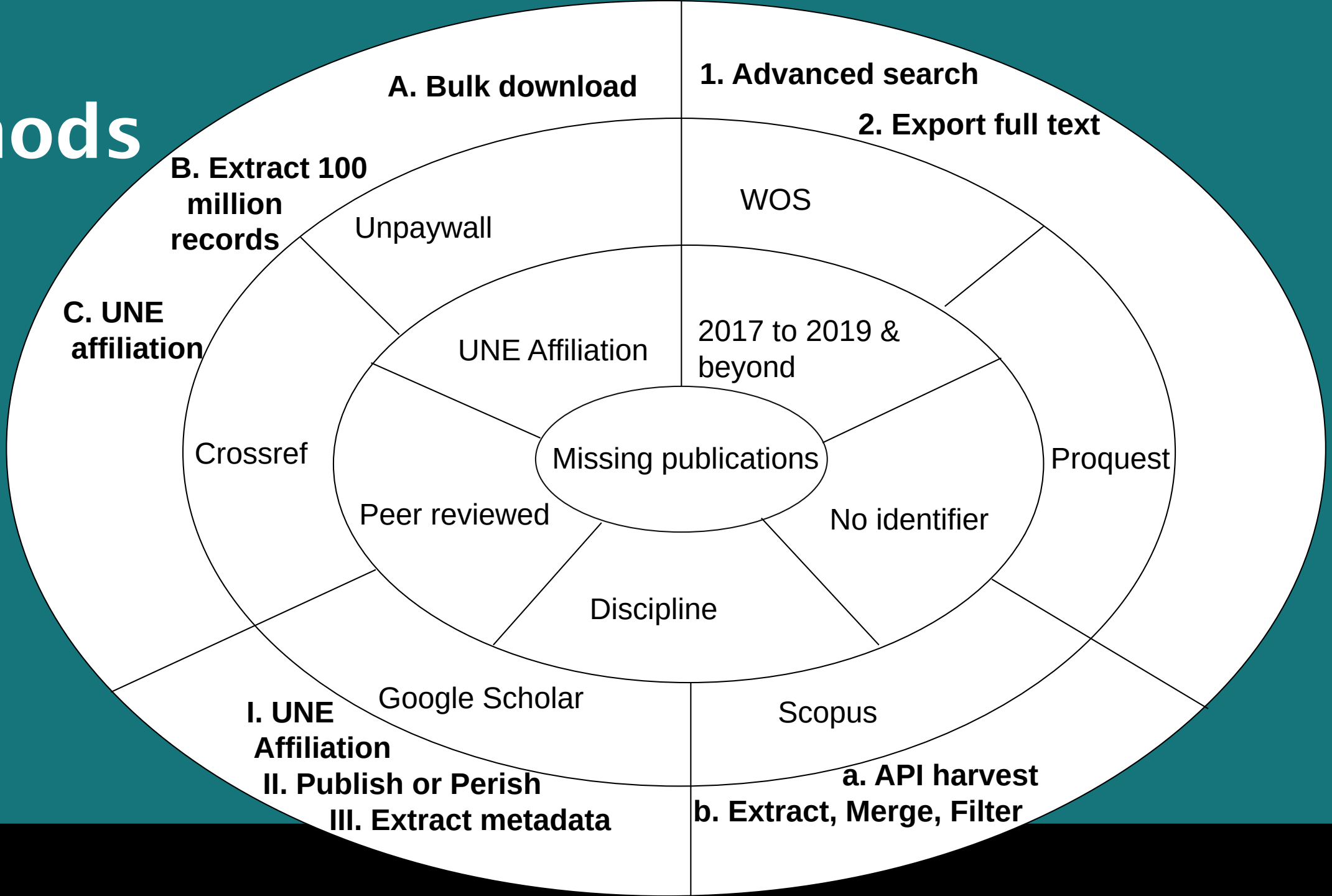
Working in the same groups, brainstorm ideas for different sources to find data/metadata on the identified factors.

Add potential sources around the third ring.



University of New England

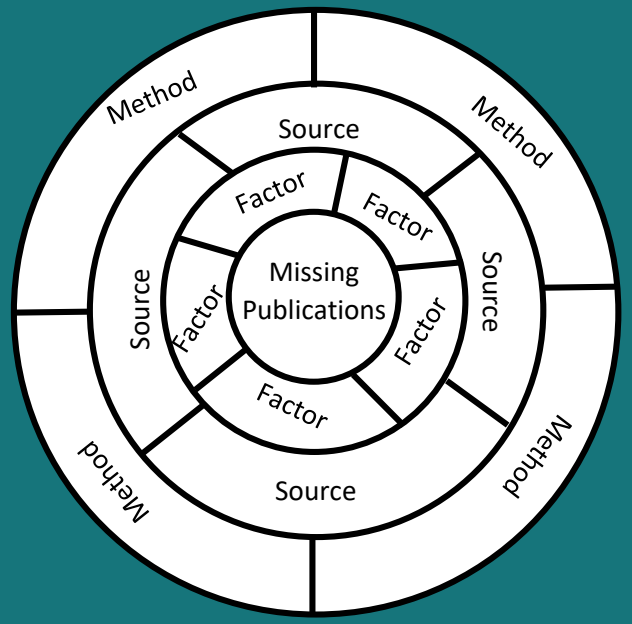
Methods


A. Bulk download
1. Advanced search
2. Export full text
B. Extract 100 million records
WOS
Unpaywall
C. UNE affiliation
UNE Affiliation
2017 to 2019 & beyond
Crossref
Missing publications
Proquest
Peer reviewed
No identifier
Discipline
Google Scholar
Scopus
I. UNE Affiliation
II. Publish or Perish
III. Extract metadata
a. API harvest
b. Extract, Merge, Filter

Activity 4

Brainstorm possible methods to extract data/metadata from the different identified sources.

Add potential methods around the outer ring.



University of New England

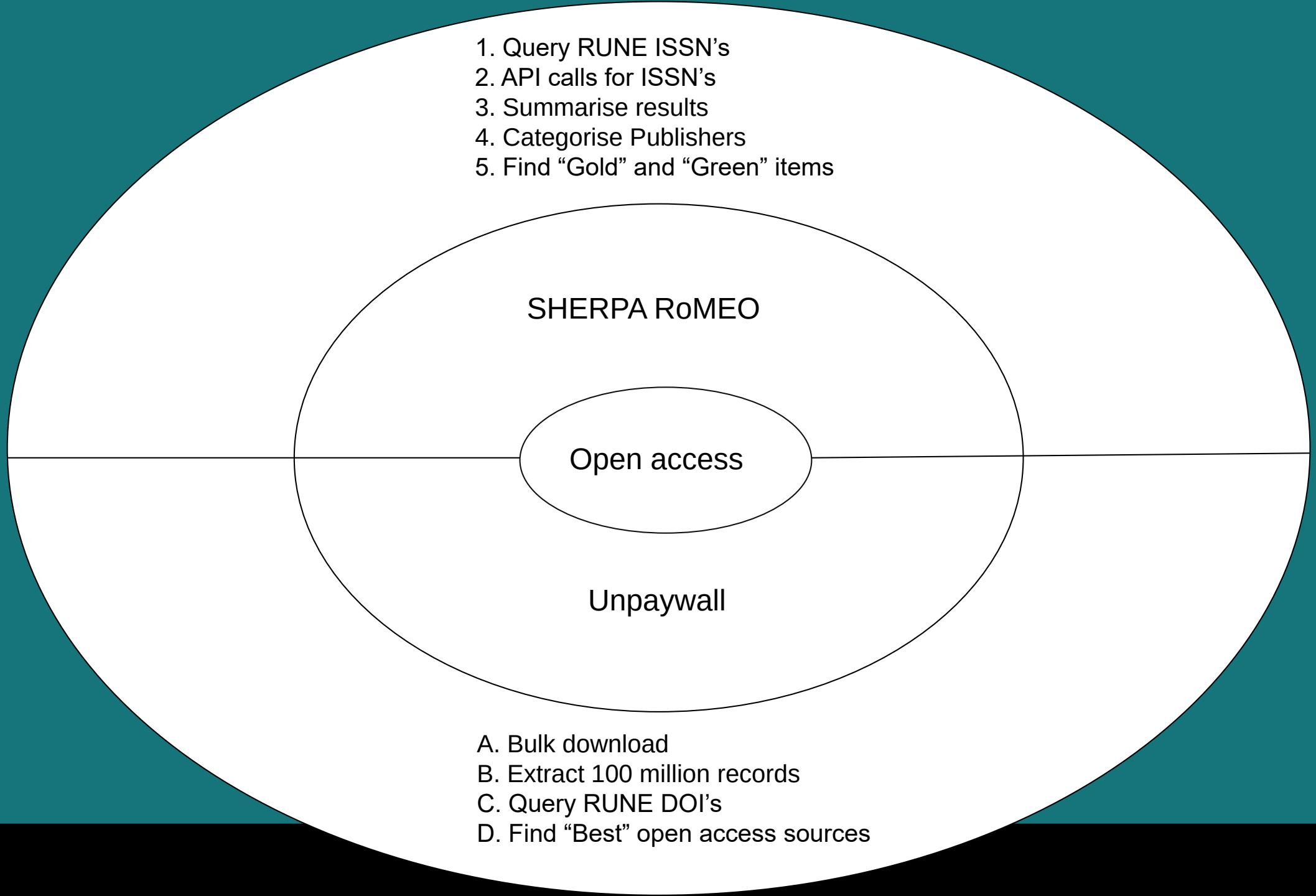
Outcomes: Missing Publications

Year	Scopus	WOS	Proquest	Crossref	Unpaywall	Total
2014	212	55		7		274
2015	204	69	4	4		281
2016	235	89	1	3		328
2017	169	109		8	1	287
2018	256	142	1	10		409
2019	139	110				249

Outcomes: Lessons

- Remaining gaps
- Capturing transient UNE researchers
- Information Explosion
- Open Access

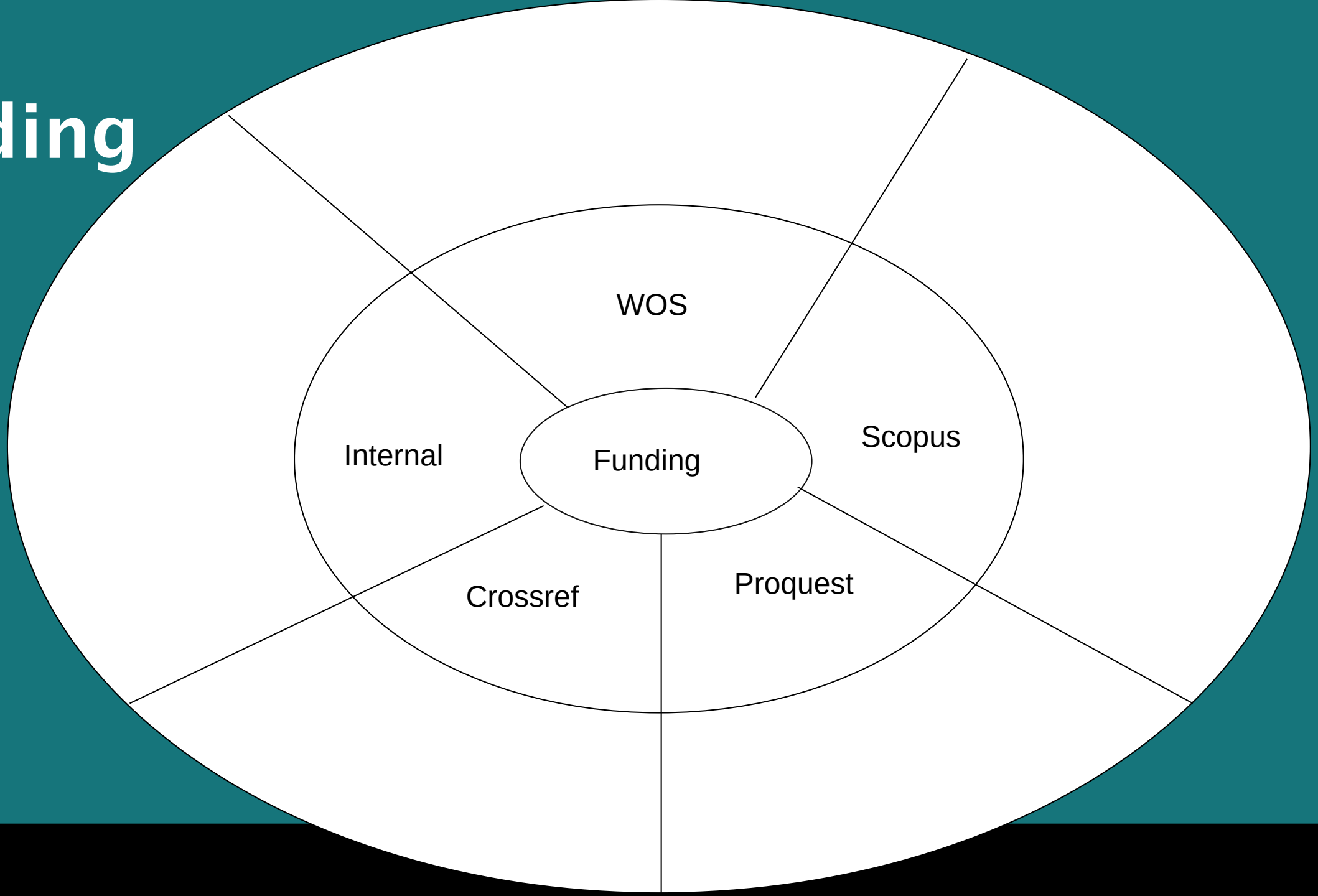
OA


1. Query RUNE ISSN’s
2. API calls for ISSN’s
3. Summarise results
4. Categorise Publishers
5. Find “Gold” and “Green” items
SHERPA RoMEO
Open access
Unpaywall
A. Bulk download
B. Extract 100 million records
C. Query RUNE DOI’s
D. Find “Best” open access sources

Future Directions

- Repeating gap analysis
- Capturing processes
- Additional sources
- Remaining gaps - Funding

Funding



Applause

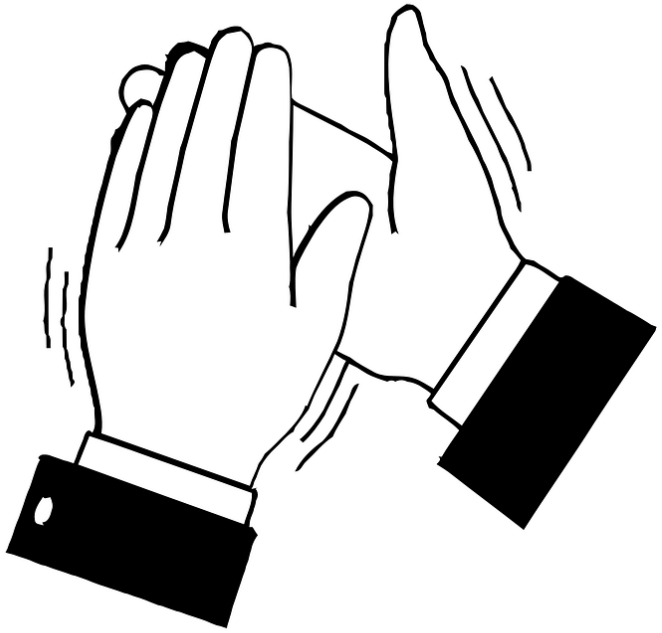


Image Source: https://pixabay.com/vectors/applause-clapping-hands-black-297115/

Questions

Image Source: kmicican https://www.needpix.com/photo/836507/question-questions-the-question-mark-question-mark-ask-asking-multidirectional-wonder-choice

Thank you to…

- Project team: Allison Brown, Gilbert Meyns, Kristal Spreadborough and Matt Watts
- Library: RUNE team, Leadership team and other staff for their support
- Research Services and UNE researchers

Image Source: https://www.maxpixel.net/Together-Teamwork-Cooperation-Team-Team-Spirit-2309035

Takeaways

- UNE R Group – https://github.com/rgroupune/
- Sharing code: 20190712_crossref_metadata_analysis

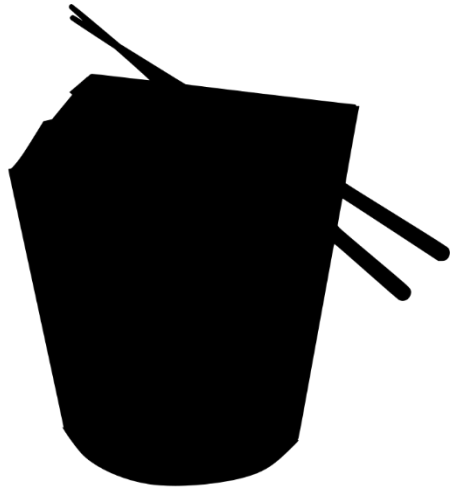
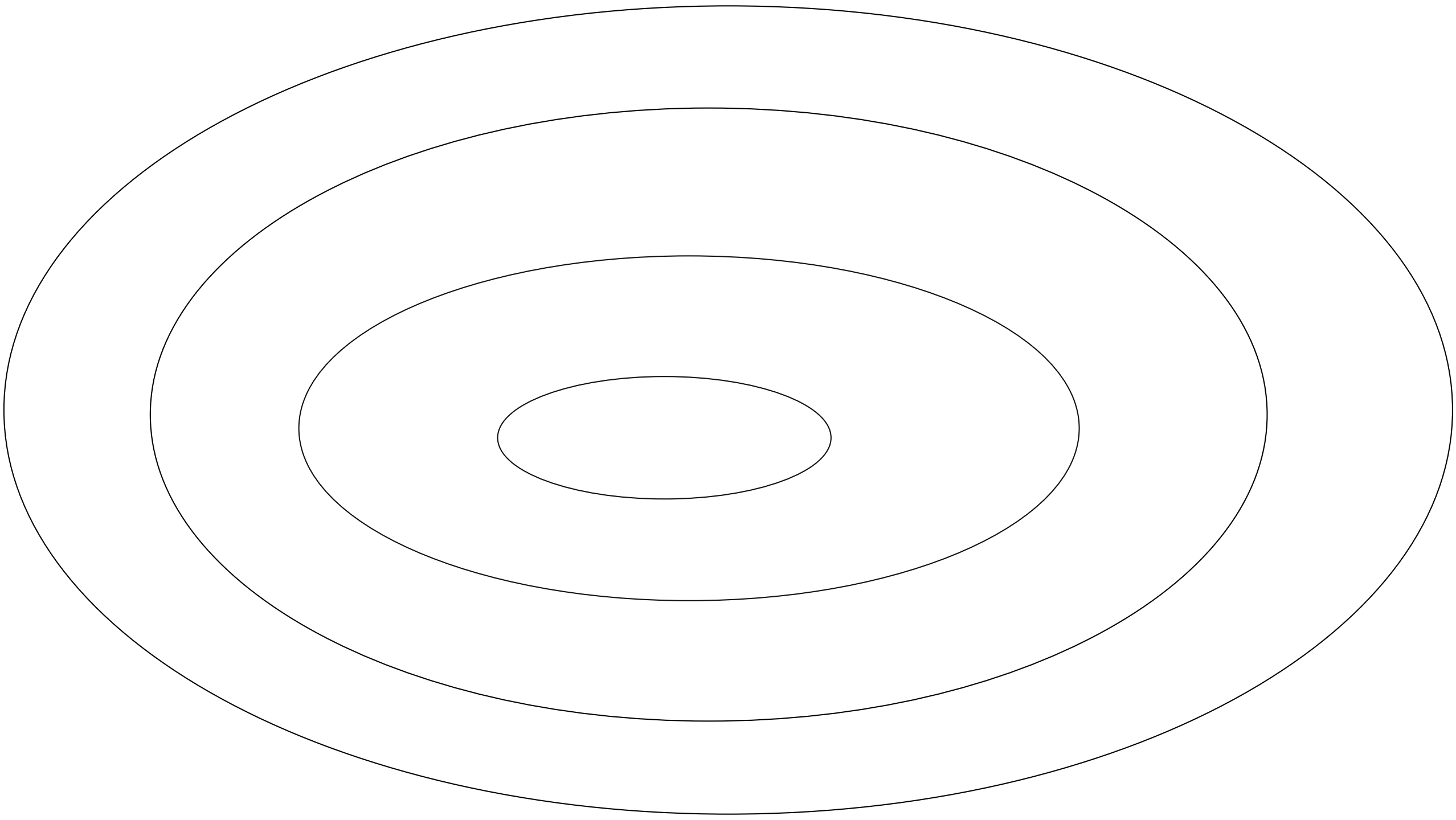



Image Source: https://svgsilh.com/image/154591.html