

# Data Metrics at Repositories : A Survey

RDA Philadelphia April 2019

RDA Data Usage Metrics WG

# Background

- Idea of a survey conceived during a Working Group virtual call
- Questions drafted by various members of group and agreed upon by leads
- Advertised through twitter, RDA WG channels, email newsletters
- Total of 85 responses
- Initial analysis undertaken by Thomas Jouneau, Université de Lorraine

# Questions

- Demographics
  - Do you represent a repository manager/curator?
  - Which country is your repository primarily based in?
  - Repository Type?
- Current Practices
  - Do you track metrics for metadata, data downloads, or both?
  - Which usage metrics do you track?
  - Which usage metrics do you expose?
  - If you expose metrics, how do you display them?
  - If you track but do not expose data metrics, why?
  - What kind of tracking software do you use?
  - Do you combine usage metrics for datasets with usage metrics for eg. articles?
  - Do you follow any privacy policies for collecting usage metrics?
  - Do you receive requests from depositors to provide metrics about use of their datasets?

# Questions

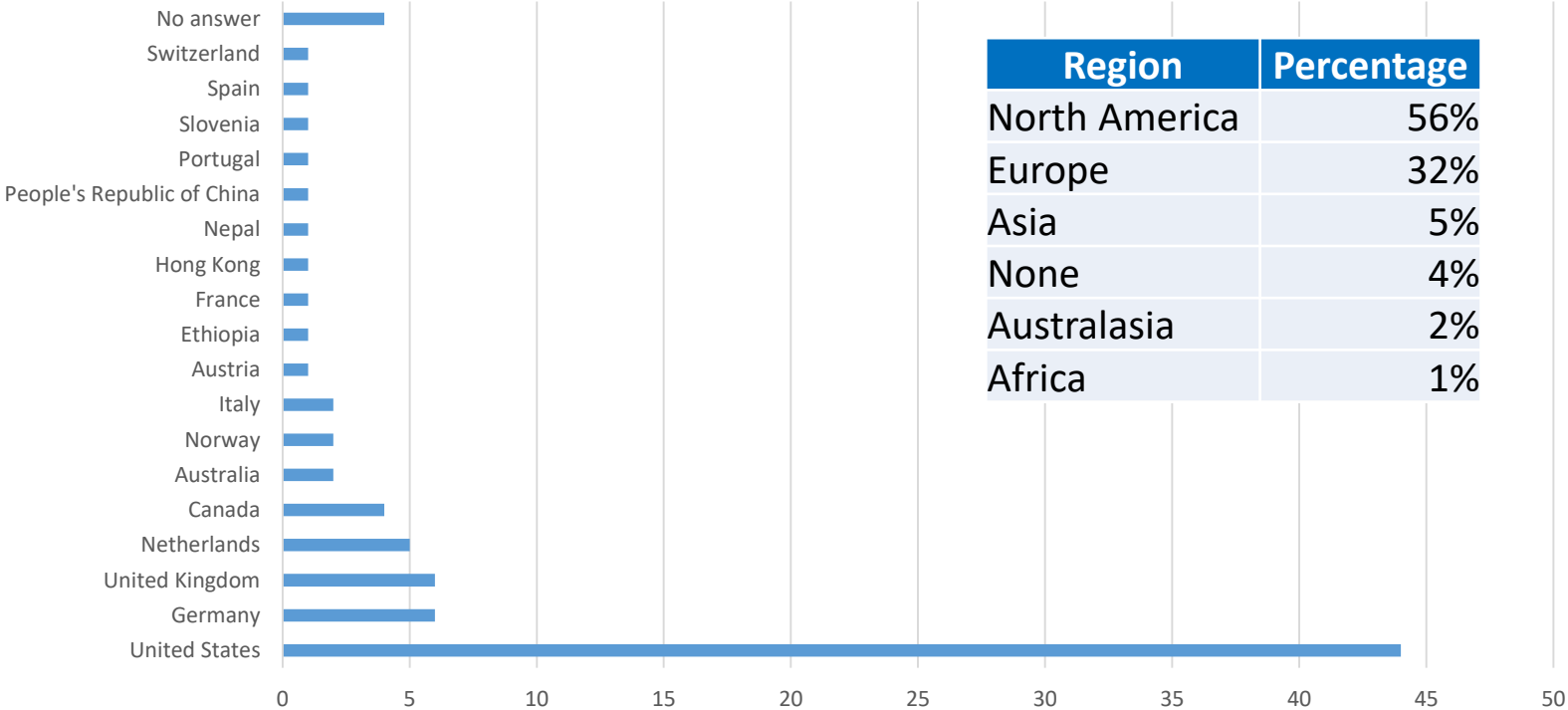
- COUNTER Code of Practice
  - Are you aware of the COUNTER Code of Practice for Research Data?
  - Do you plan to standardize your usage metrics against this Code of Practice?
  - If you do not plan to standardize your usage metrics against this Code of Practice - why not?
  - What do you see as barriers that stand between your repository and implementation of comparable usage metrics?

# Survey Demographics



N=85

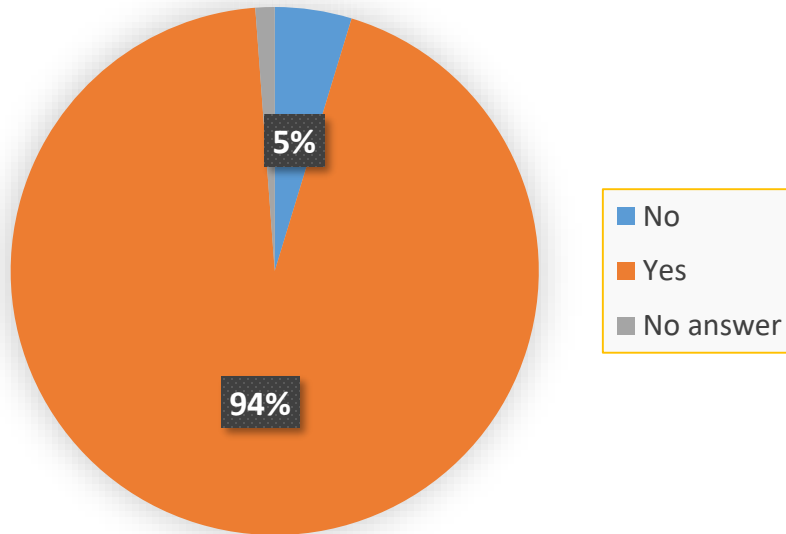
# Survey Demographics



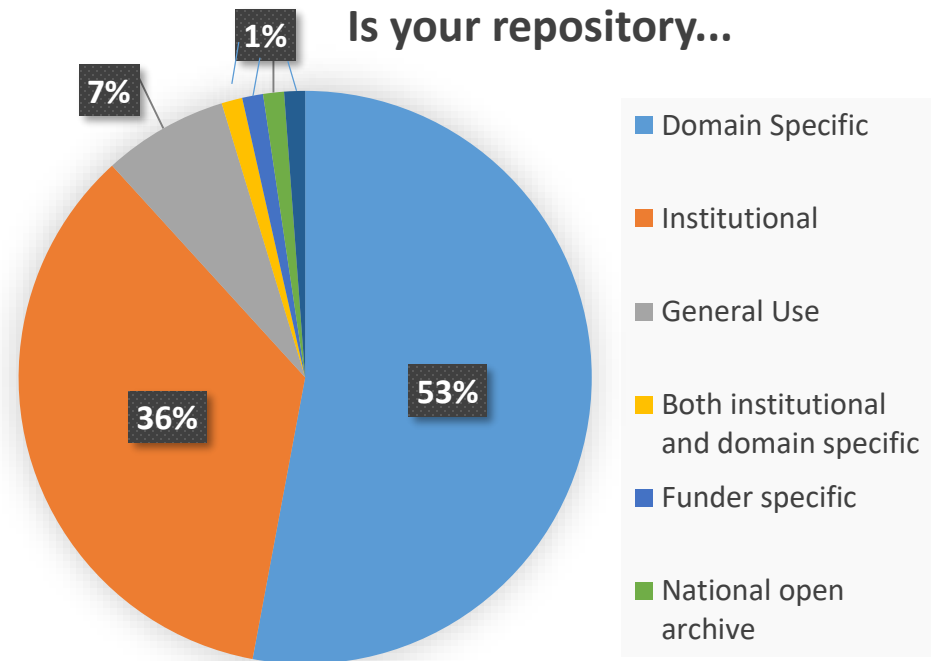
N=85

# Repository Information

Do you represent a repository manager/curator?



Is your repository...

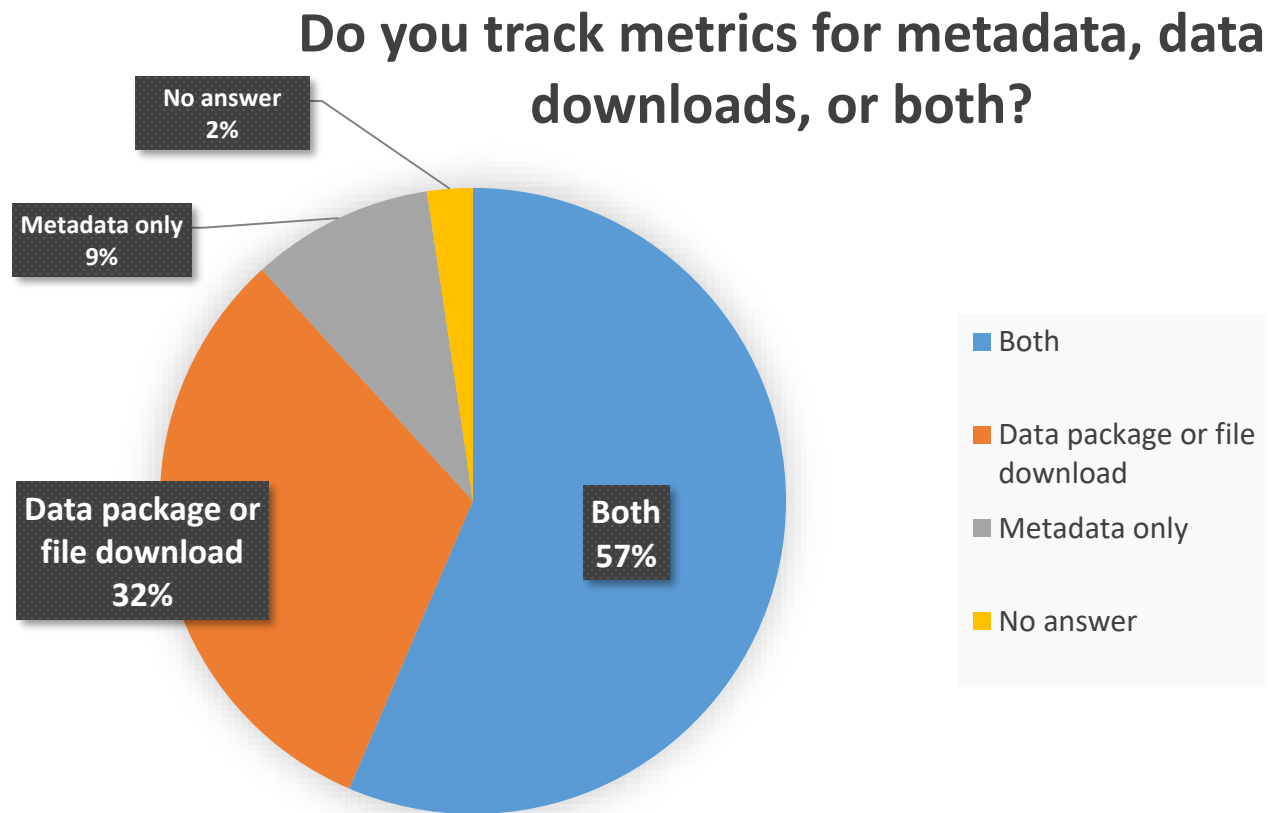


# Summary

- Demographics
  - 88% from North America and Europe
  - 94% from Repository Managers
  - 53% from Domain Specific, 36% Institutional

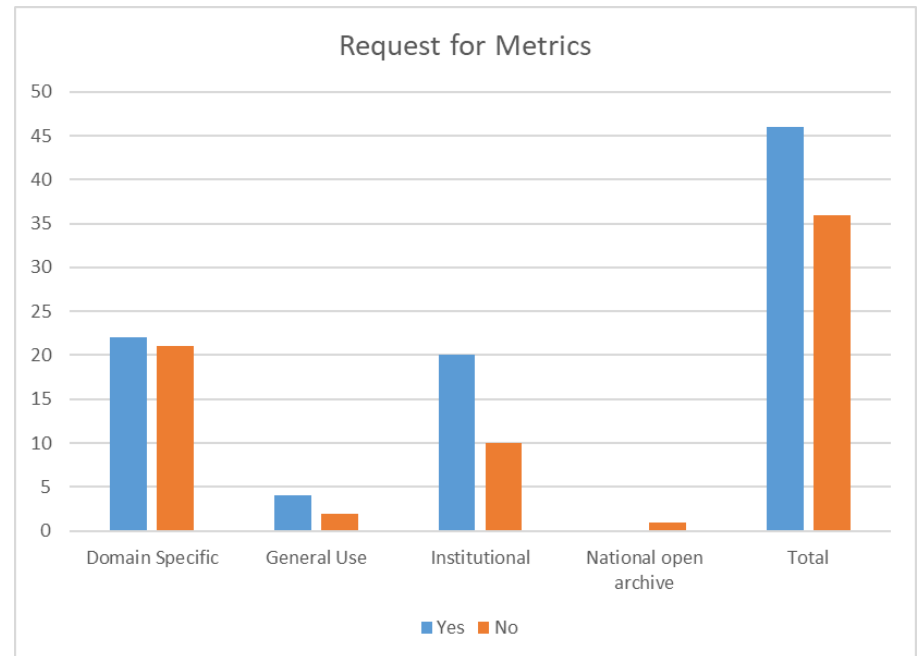
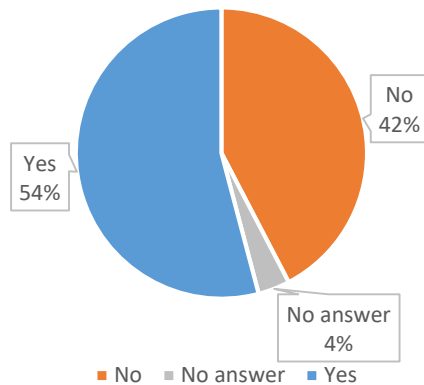


# Usage Metrics Tracking



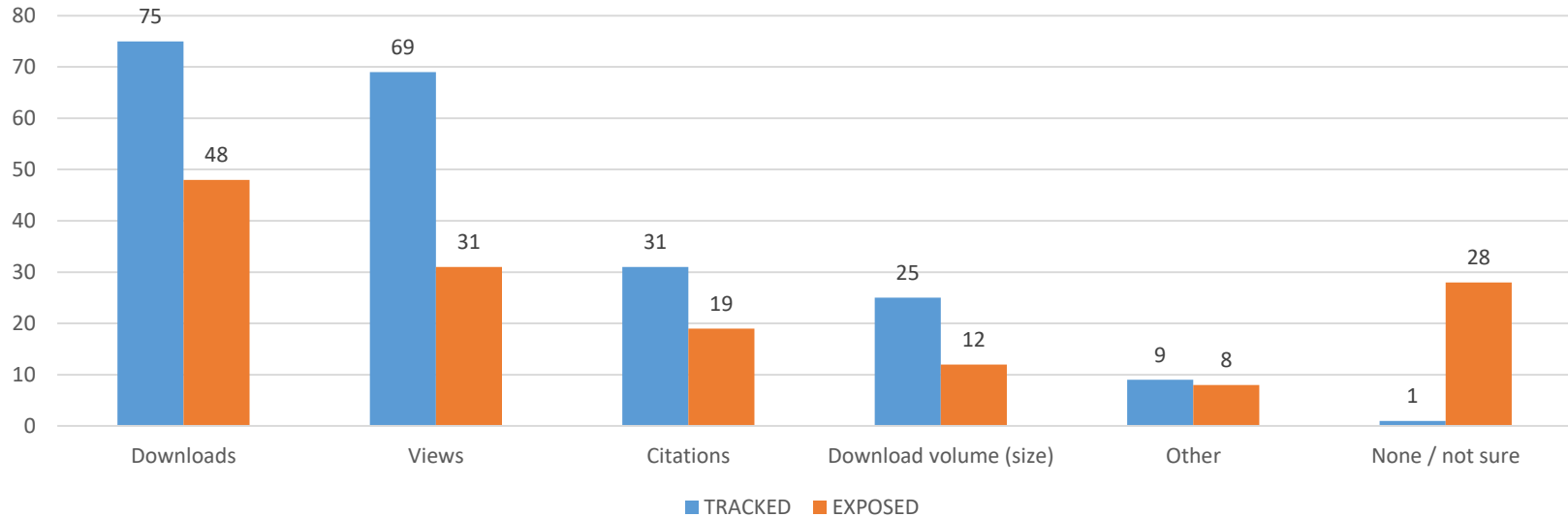
# Requests for Metrics

Do you receive requests from depositors to provide metrics about use of their datasets?



# Usage Metrics Tracking & Exposure

Which metrics do you track? Expose?



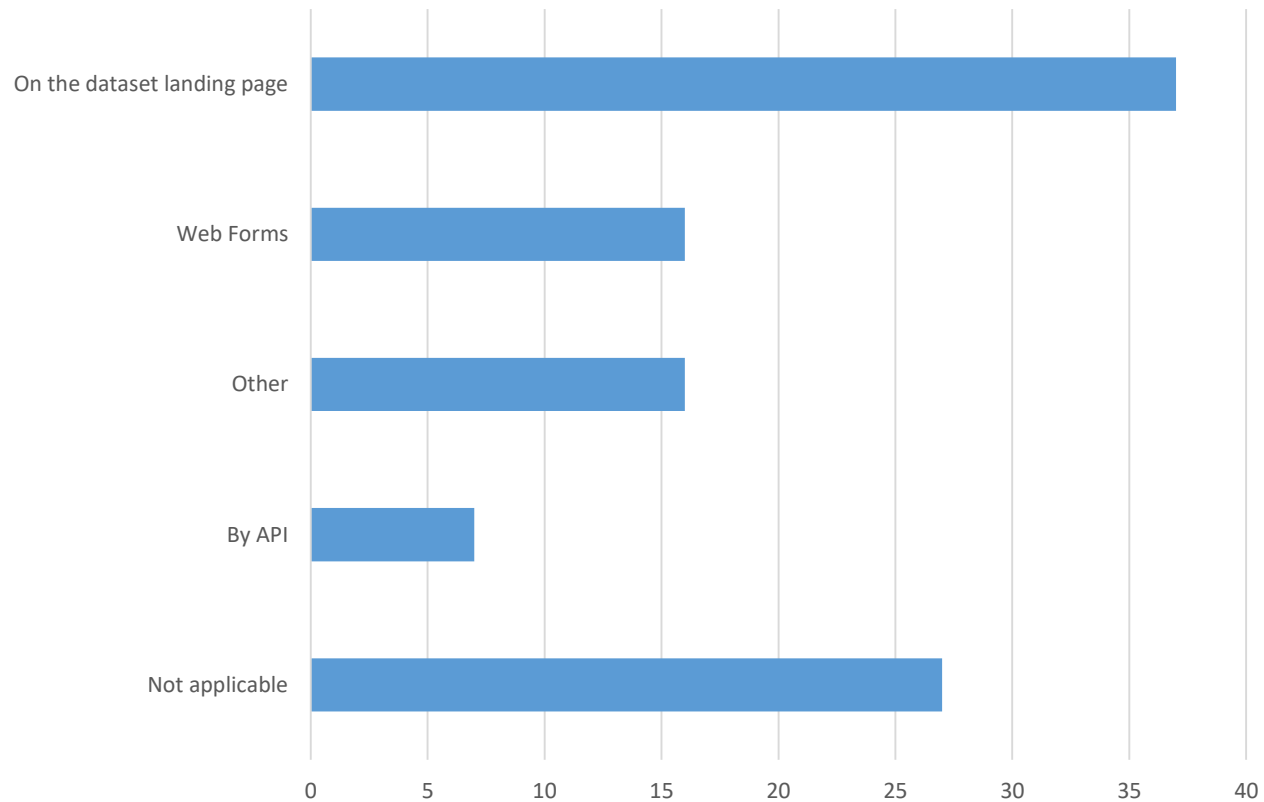
# Other Metrics

- Number of deposits
- PlumX Metrics
- Requests from Apache logs
- Altmetrics
- Information and data only
- Unique users downloading data products per time unit
- Project specific metrics
- Social media mentions
- User status (logged in or anonymous)

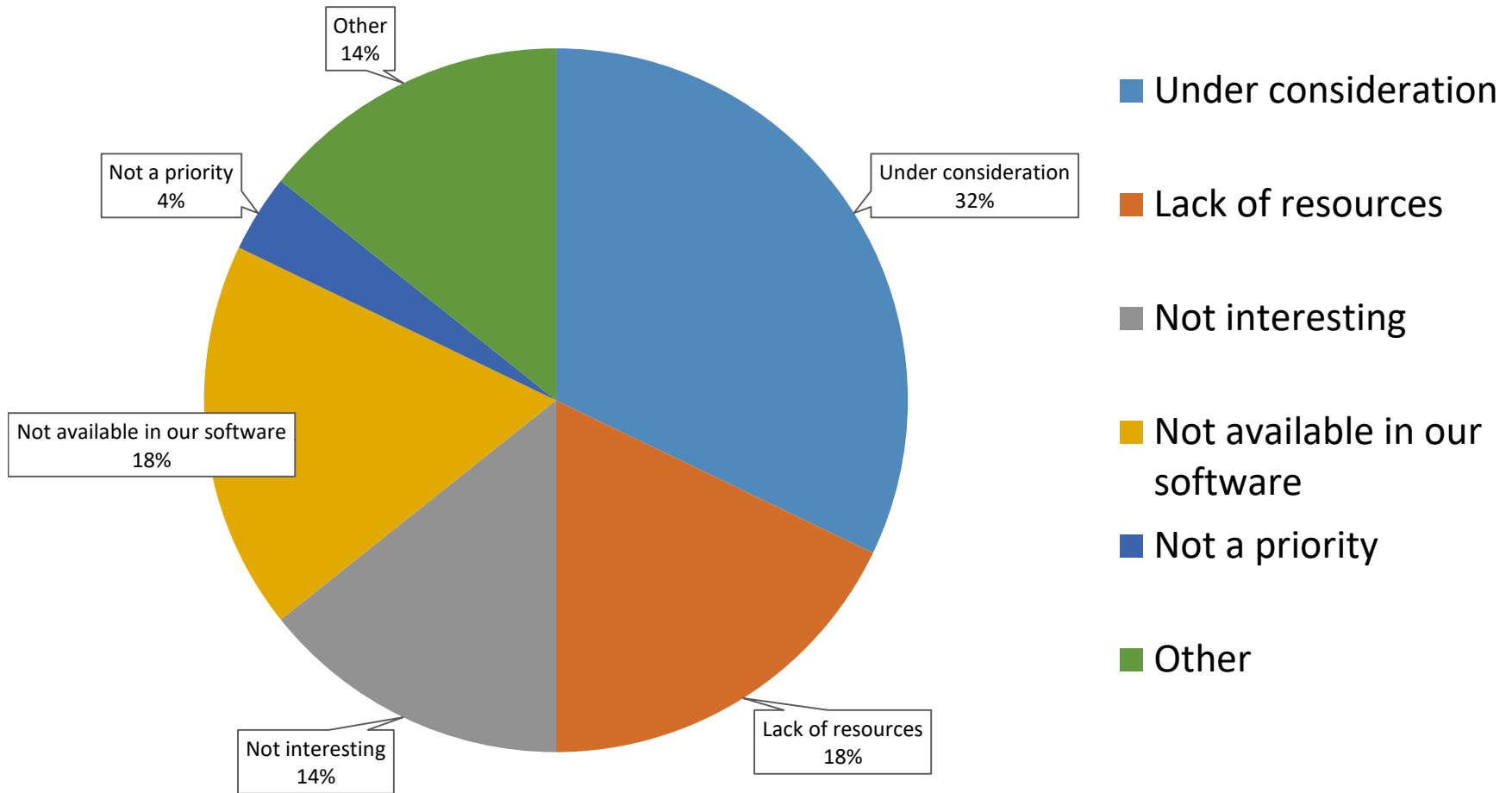
# Summary – Current Practices

- Current Practices
  - 89% tracking data downloads, 66% tracking metadata
  - 54% receive requests for metrics
  - Institutional repositories receive more requests for metrics
  - Downloads & Views tracked by most but not necessarily exposed
  - Citations and Download Volume not tracked as much

# Usage Metrics Dissemination Methods



# Reasons for not exposing Metrics



# Reasons other than lack of time/knowledge

I'm not convinced of their accuracy yet.

Never been asked to.  
Not funded for that.

Must be checked due to GDPR.

Our repository is a young effort and usage is not high enough to publicly display...yet

End users don't care about these

Problem of definition what constitutes usage regarding research data

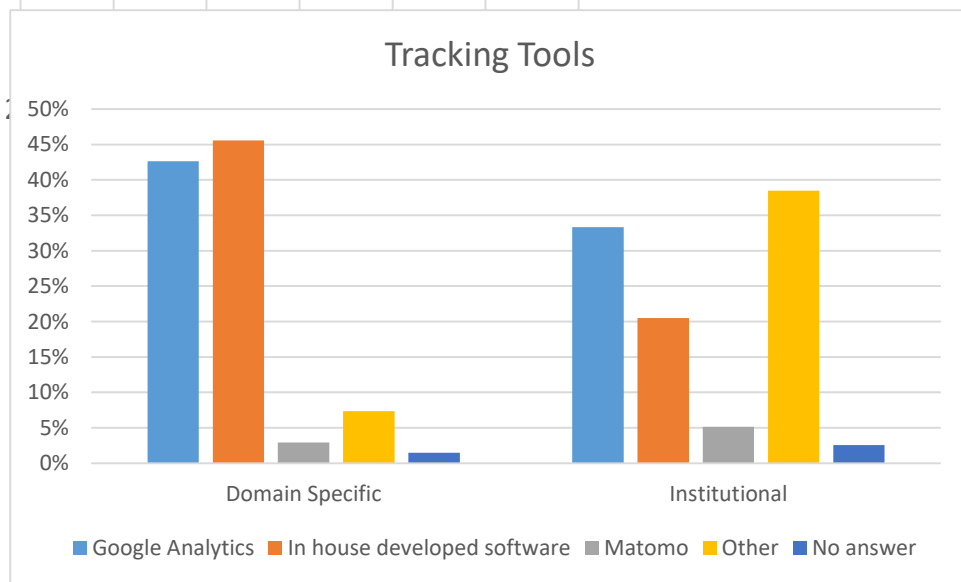
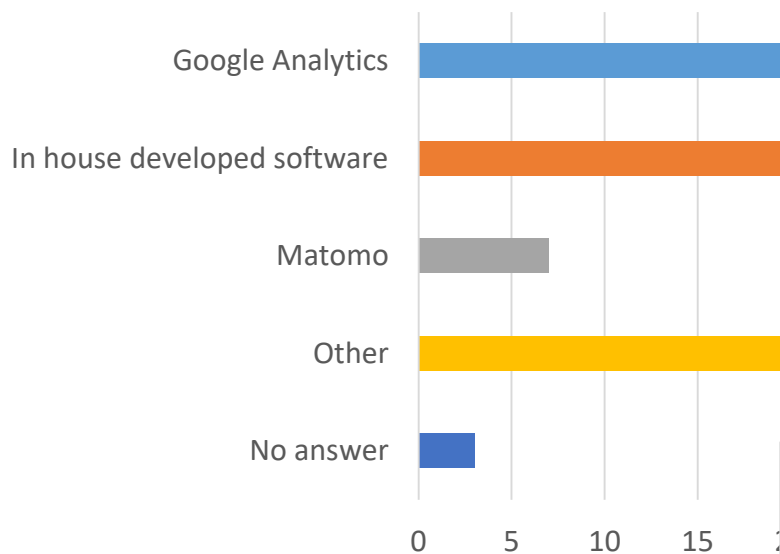
We ... have since moved to Google Analytics and have not yet worked on exposing them

No time, no place, no use

We don't want to appear to criticize undownloaded content.



# Tracking Tools / Software



# Tracking Tools / Software : Other

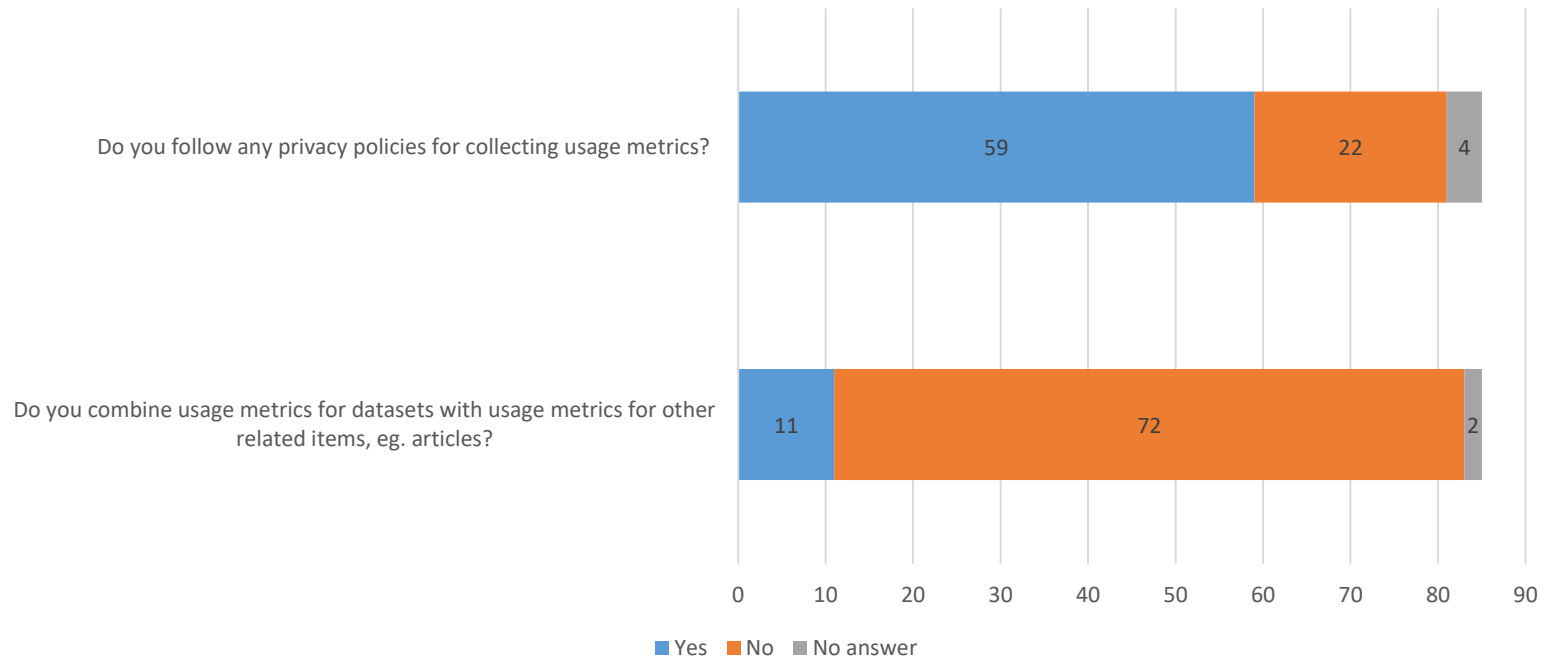
- Apache/web server logs
- Dspace Add-ons
- AWStats
- Analytics that comes with Digital Commons
- EzPAARSE
- Feeds from bibliographic indexes
- Hotjar
- Kibana and CRM
- OA Statistics service VZG
- PlumX Metrics
- Samvera-Hyrax out of the box analytics
- Whatever Figshare is using
- Zabbix

# Summary – Current Practices

- Current Practices

- 89% tracking data downloads, 66% tracking metadata
- 54% receive requests for metrics
- Institutional repositories receive more requests for metrics
- Downloads/Views tracked by most but not necessarily exposed
- Citation and Download Volume not tracked as much
- About a third of respondents considering implementation
- Lack of time/resources the most common reason for not exposing
- Some not convinced of the value of usage metrics
- Off-the-shelf and bespoke methods both popular
- Google Analytics commonly used

# Metrics Policies



# COUNTER Code of Practice for Research Data

Are you aware of the COUNTER CoP for Research Data?

Yes - 39

No - 45

Do you plan to standardize your usage metrics against this code of Practice?

Yes - 13

Maybe - 24

2

4

Maybe - 35

6

- COUNTER Code of Practice is relatively well-known – for its young age (46%)
- 90% considering implementation (20% definitely considering)

# COUNTER CoP Compliance Issues

If you do not plan to standardize your usage metrics against this Code of Practice – why not?

"Because metrics are mainly a marketing instrument for the repository and it's maybe not worth the effort"

"Limited resources and uncertainty about repository roadmap"

"Looks like a lot of work"

"COUNTER is too journal article focused, and doesn't seem relevant to data"

"We'd be delighted if DSpace supported COUNTER; don't have resources to do it ourselves."

"bepress handles this for us"

"Only one entry of the COUNTER metrics fits what we are doing now: Total Searches, Result Clicks and Record Views by Month and Platform - currently we record searches by type of information. Our query volume bears little resemblance to citation references for data actually used. "

"Can't easily find my way at the COUNTER site"

# COUNTER CoP Compliance Issues

If you do not plan to standardize your usage metrics against this Code of Practice – why not?

"Because metrics are mainly a marketing instrument for the

re  
m  
ef

- Lack of skills, time and/or resources to implement
- Software platform controls what is possible
- Not a priority
- COUNTER is not about data

"L  
and uncertainty  
about repository  
roadmap"

"Looks like a lot of  
work"

"COUNTER is too journal  
article focused, and  
doesn't seem relevant

resources to do it  
ourselves."

"bepress handles this  
for us"

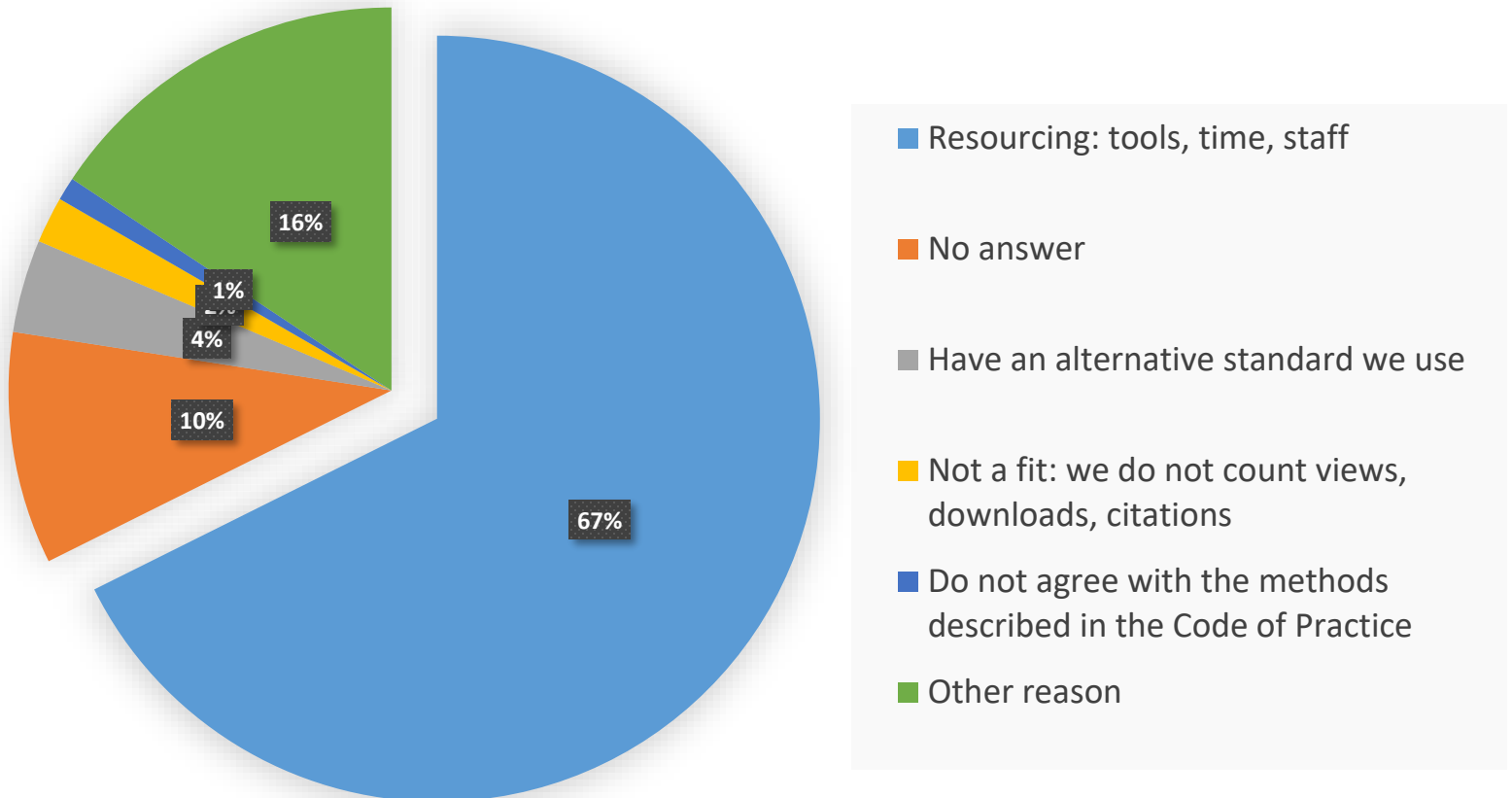
"Only one entry of the  
COUNTER metrics fits what  
we are doing now: Total

resemblance to citation  
references for data actually  
used. "

"Can't easily find my  
way at the COUNTER  
site"

# COUNTER CoP Compliance Issues

What do you see as barriers that stand between your repository and implementation of comparable usage metrics?





# COUNTER CoP Compliance Issues

What do you see as barriers that stand between your repository and implementation of comparable usage metrics?

