DAC Review & Meeting the Genomic Data Demand

Jonathan Lawson, Broad Institute USA



SCAN ME

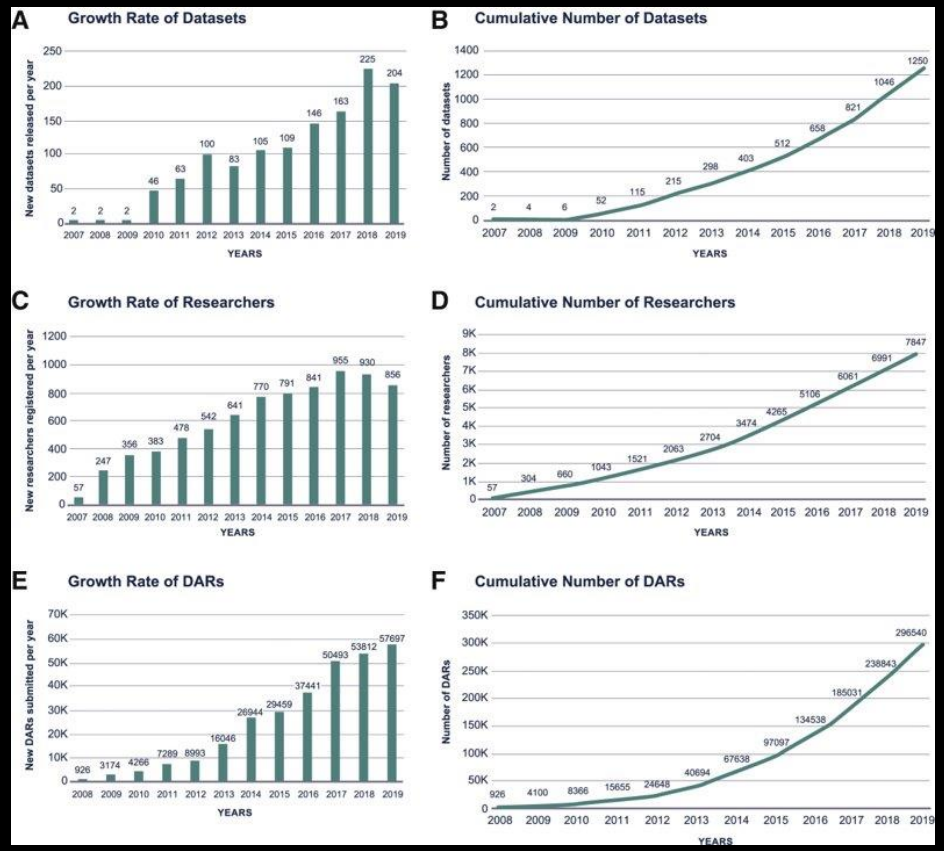
to learn more about today’s talk

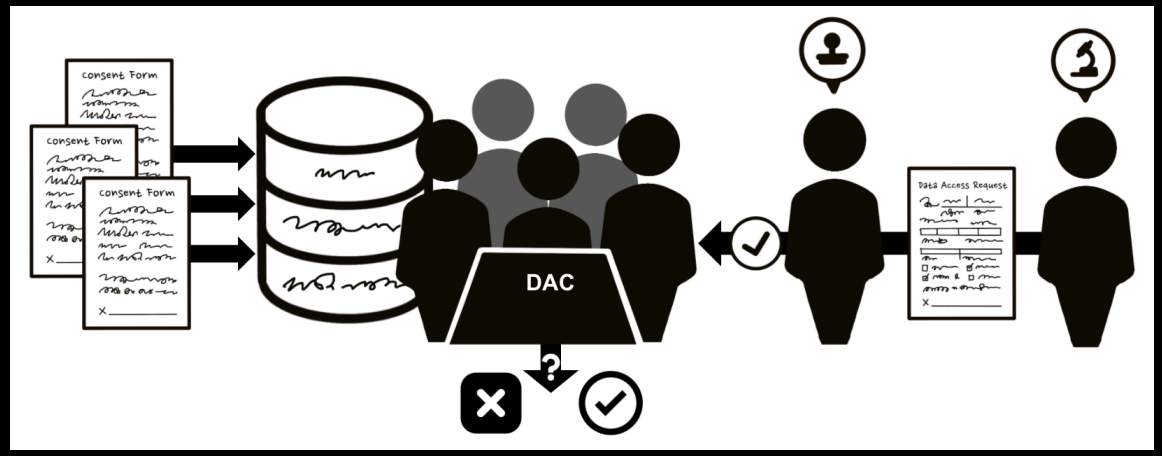
Outline

- What's the problem?
- What are we doing to solve it?
- Survey & interview with global DAC members
- What did we learn?

1. What's the problem?

A growing demand on highly manual data access processes





The Current State of Data Access Review

"Operational challenges of data access management limit researchers' ability to quickly, efficiently, and securely access data at scale"

Lawson, J. et al. The Data Use Ontology to streamline responsible access to human biomedical datasets, Cell Genomics 1. 10028 (2021)

"Current processes for data access will not be able to scale to meet the volume of access requests projected in the next few years without significant investment in human resources to cover exponential increase in tasks."

Cabili, M. N. et al. Empirical validation of an automated approach to data use oversight. Cell Genomics 1. 100031 (2021)

2. What are we doing to solve it?

GA4GH established the **Data Access Committee Review Standards** (DACReS) work group in 2020

Goal

Streamline the DAR process to expedite scientific discovery!

GA4GH Approved Standards

Date Access Committee Review Standards (DACReS) Policy

Leadership & Membership

Vasiliki Rahimzadeh, Baylor College of Medicine

Edward S. Dove, University of Edinburgh

Jonathan Lawson, Broad Institute

And over 50+ members from the GA4GH community

Survey & Interview with global DAC Members



Aim 1

Develop and pilot a DACReS Toolkit within the NHGRI AnVIL and GA4GH Driver Projects



Aim 2

Identify challenges and opportunities for implementing automated approaches to data access review, from the perspectives of domestic and international DAC members

Funded by the AnVIL Data Ecosystem Data Access Committee Review Standards Bioethics Supplement
Thank you NHGRI!

Survey Results Overview



Demographics & Institutional Representation



DAC Workflows



DAC Operational Statistics



Perceptions on Proposed Solutions

Demographics & Institutional Representation



National genomic consortia/databases represented



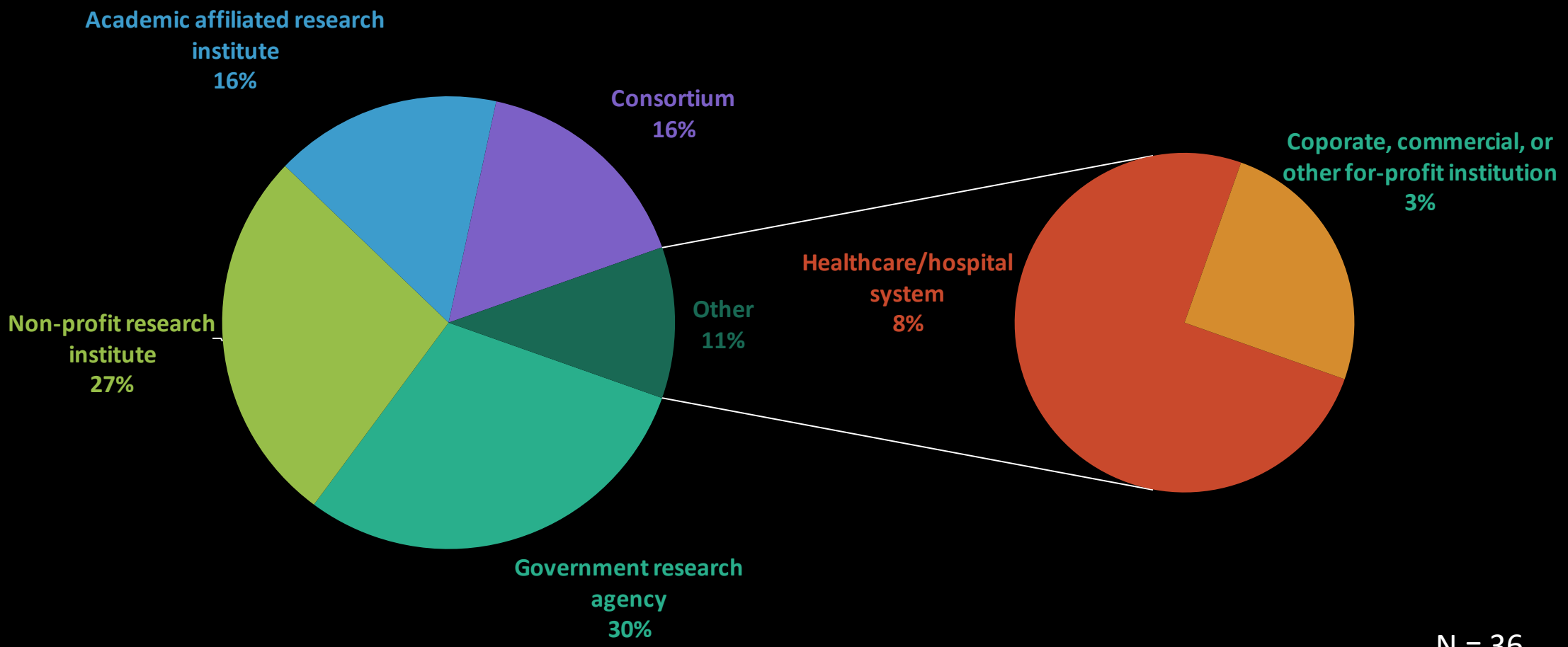
N = 35

DAC Demographics



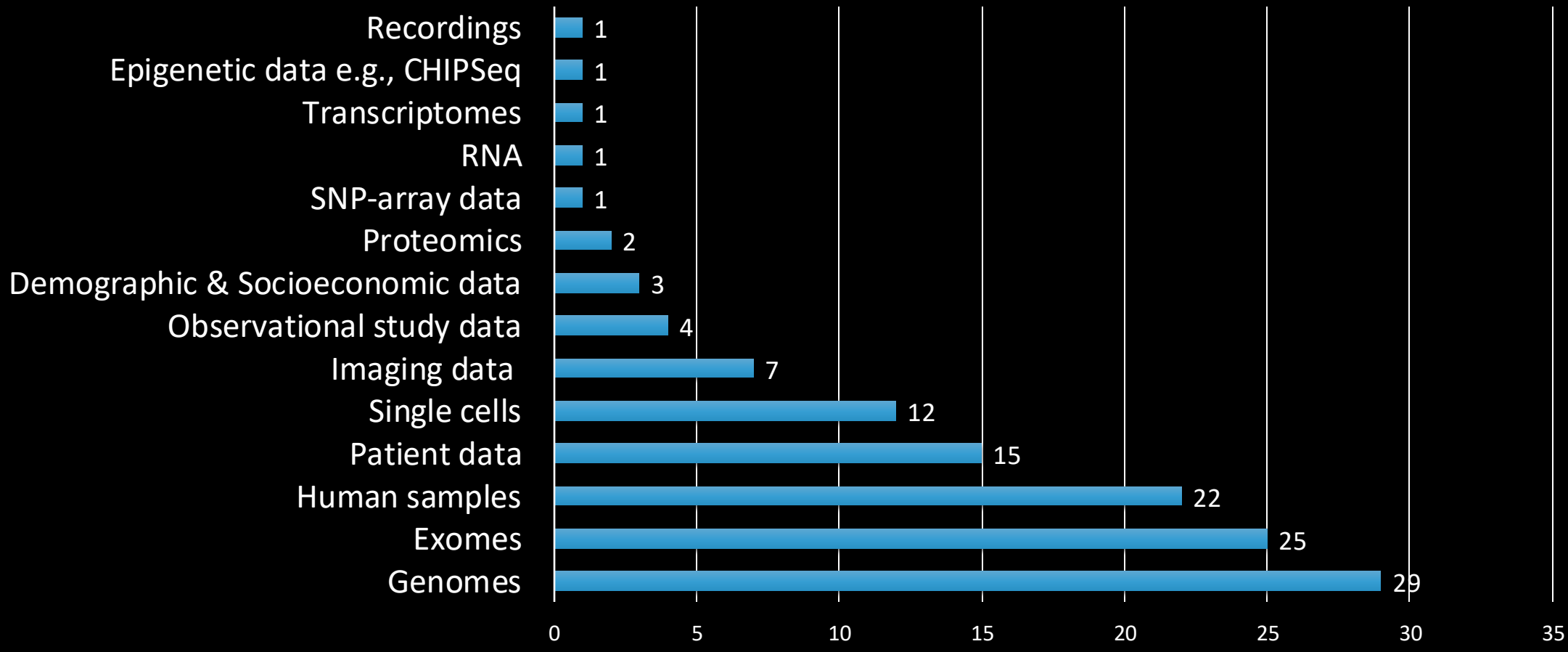
N = 35

What type of institution does your DAC serve?



N = 36

What data type(s) does your DAC manage?



How many datasets does your DAC manage?



N = 34

DAC Operational Statistics



Standard Operating Procedures



N = 33

Terms of Reference



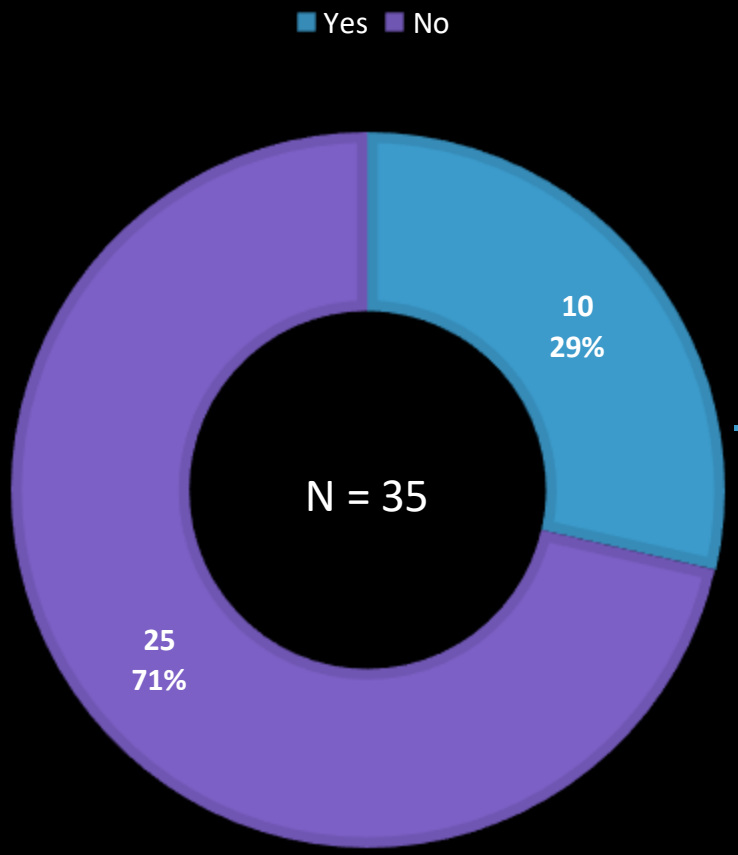
N = 35

I don’t know

Why did your DAC develop SOPs?

Need for **consistency and efficiency** in review process	30.77%
To foster **transparency and accountability**	28.57%
Need for **objective decision-making**	20.88%
Increased data access requests	8.79%
Increased datasets available	6.59%
Operations management at the institutional level	1.10%
Alignment to GA4GH DACReS Frameworks	1.10%
Data requests from unexpected third parties (e.g., private individuals)	1.10%
Legal duty	1.10%

Is your DAC work in your job description?



Government research agency	5
Non-profit research institute	3
Consortium	1
Corporate, commercial, or other for-profit institution	1

Is DAC work in your job description?

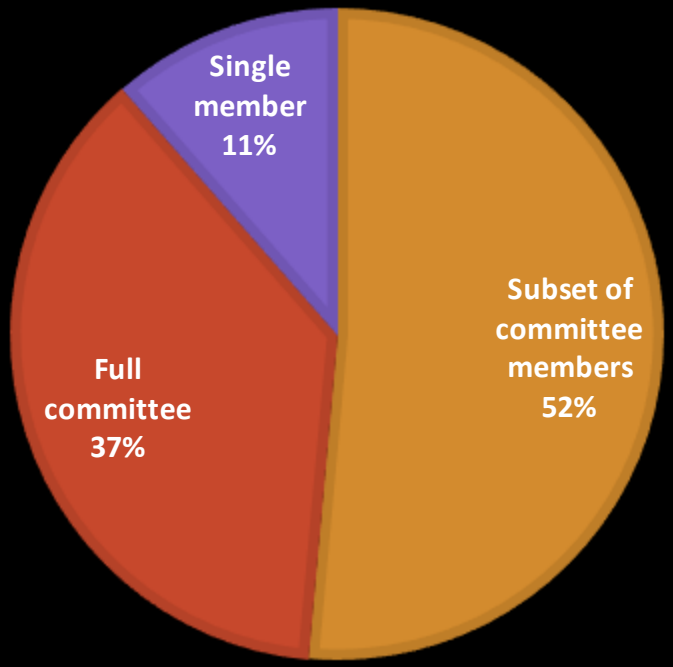
For those who answered "Yes"

- 4 people spend 1-2 hours weekly
- 2 people spend 50% of their time on DAC work
- 1 person spends 70% of their time on DAC work
- *Only one person dedicated 100% of their time to DAC work*
  - Managing >1000 datasets and >100 DARs per month

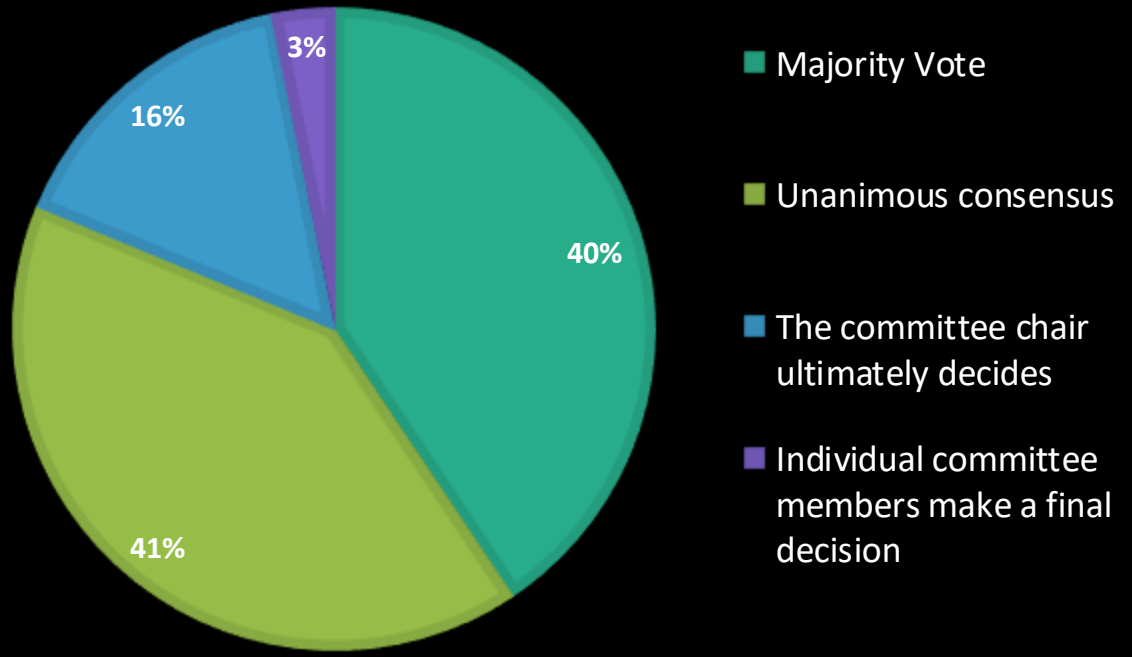
DAC Workflows



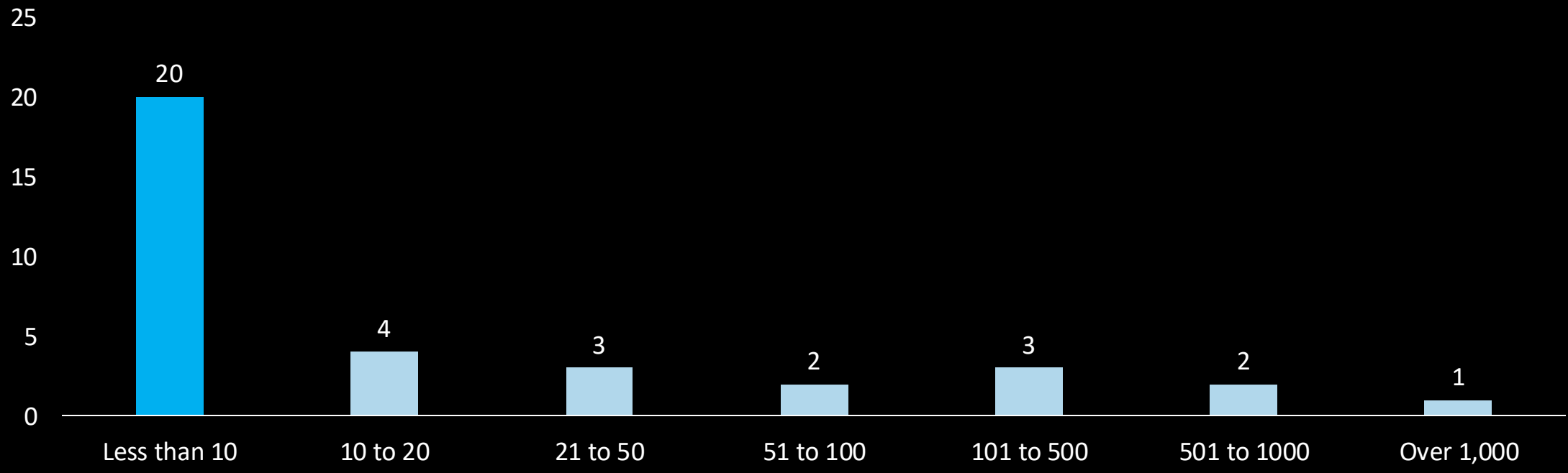
Who participates in reviewing DARs?



How does your DAC make a final decision on a DAR?



On average, how many new DARs does your DAC *receive* per month?



57% of DACs receive less than 10 DARs per month

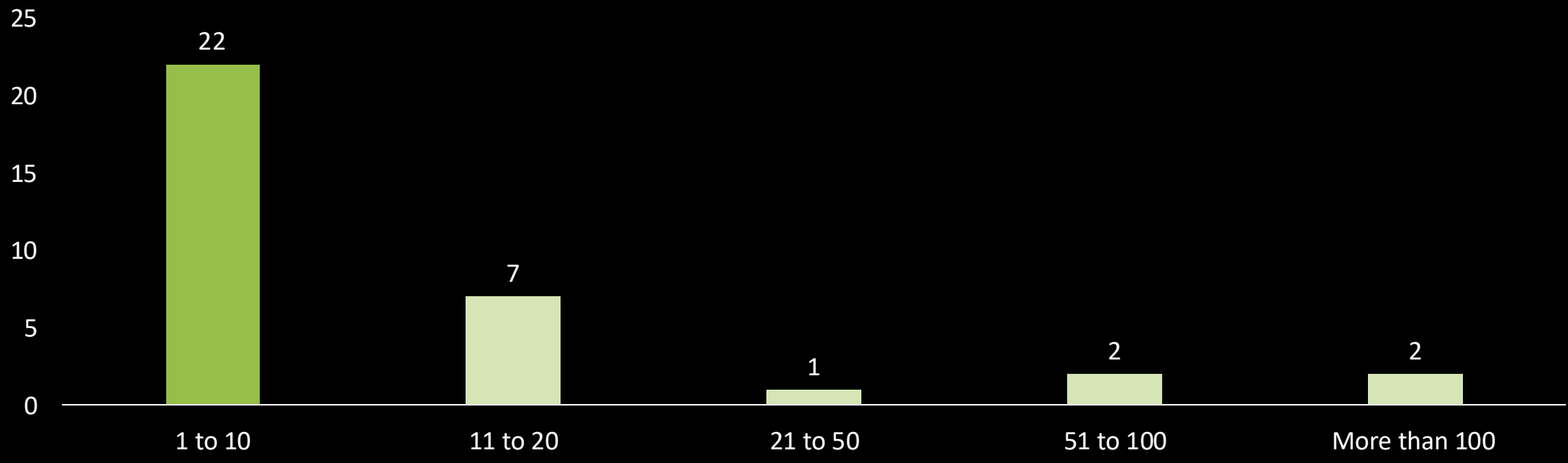
On average, how many new DARs does your DAC *receive* per month?

Less than 10 new DARs per month	
Non-profit research Institute	7
Consortium	5
Academic research institute	4
Government research agency	3
healthcare/hospital system	1

10-50 new DARs per month	
Government research agency	2
Academic research institute	2
Non-profit research Institute	2
Consortium	1

Over 50 new DARs per month	
Government research agency	5
Non-profit research Institute	2
Corporate, commercial, or other for-profit institution	1

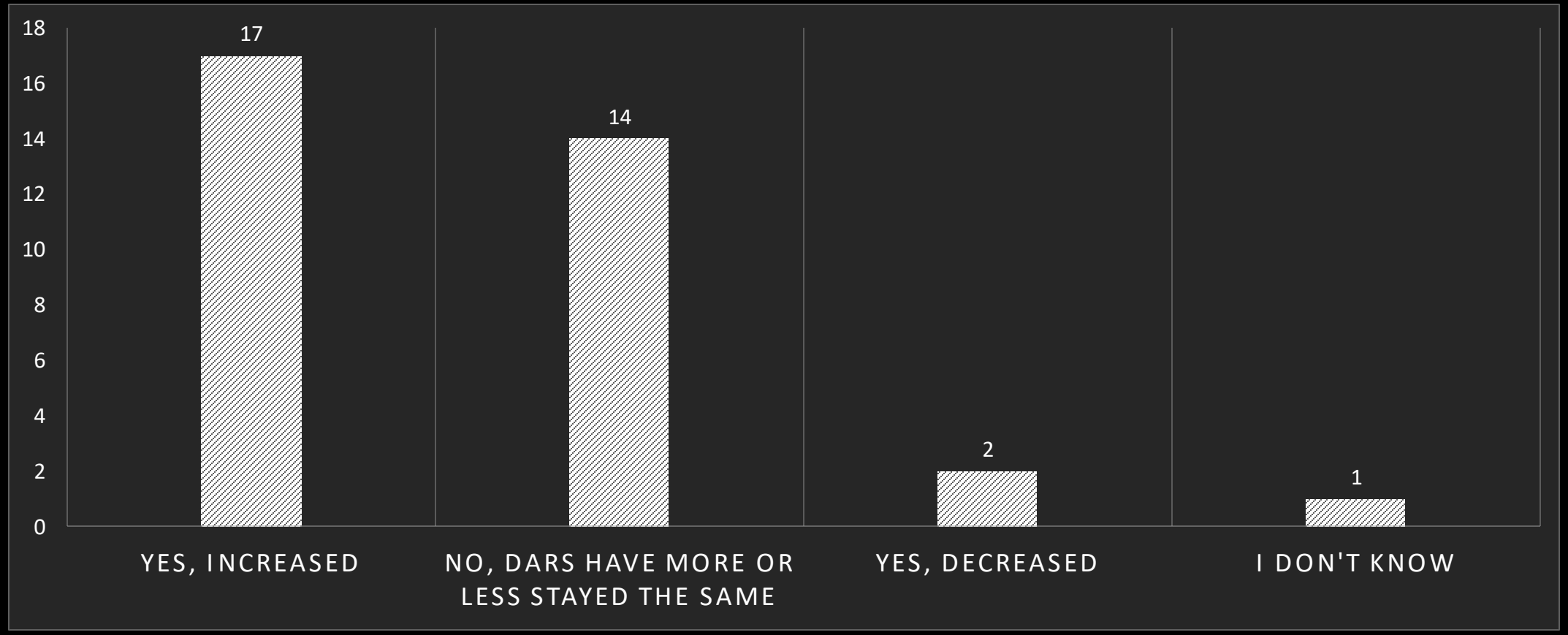
On average, how many DARs does your data access committee *review* per month?



N = 34

85% of DACs review less than 20 DARs per month

Have you perceived any changes to the number of DARs received during your service on the DACs?



N = 34

On average, how long does it take your DAC to review a DAR?

Less than a week	36% (12)
1 week	6% (2)
2 to 3 weeks	30% (10)
4 to 5 weeks	21% (7)
More than 6 weeks	6% (2)

N = 33

58% of DACs take a minimum of 2 weeks, 27% take at least a month

What are the causes of DAR review delays?

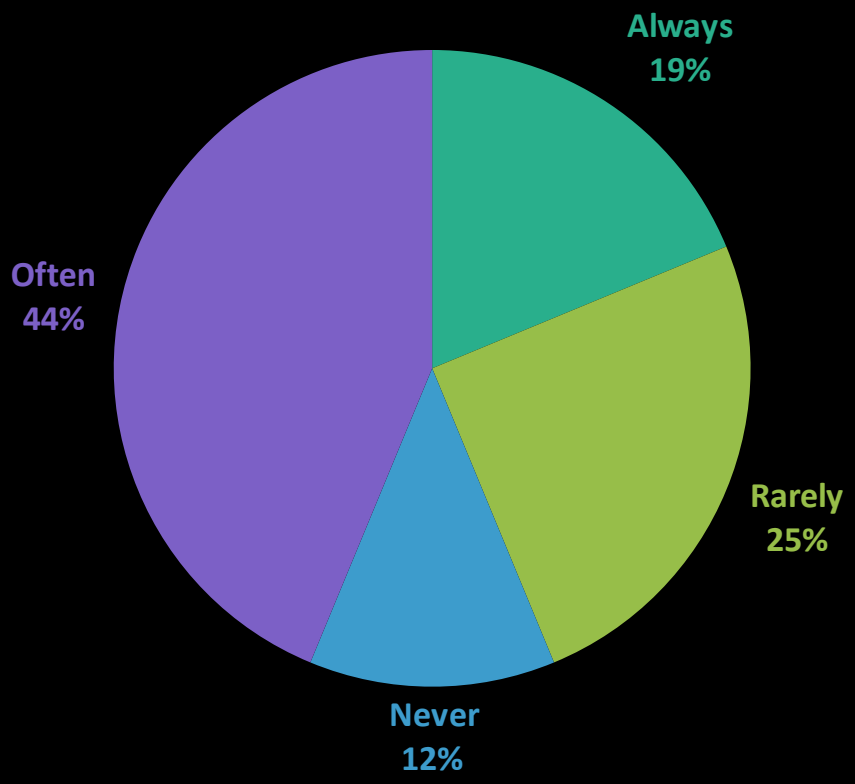
1. Missing data/incomplete information in the DAR submitted by researcher	**36.76% (25)**
2. Verifying researcher's identity and bona fides	**17.65% (12)**
3. Different interpretations of data sharing terms/allowable data uses among members of the data access committee	**14.71% (10)**
4. Verifying institution's legitimacy/trustworthiness	**11.76% (8)**
5. Ambiguous data sharing language in the original (underlying) consent form	5.88% (4)
5. Issues with scientific feasibility	5.88% (4)
7. Negotiating a data access agreement and discussing legal implications	2.94% (2)
8. No delay experienced	1.47% (1)
9. Submission long before a scheduled DAC meeting	1.47% (1)
10. DAC members are slow to react	1.47% (1)

When reviewing a DAR, how much time on average do you spend confirming the researchers' proposed uses comply with the datasets' use terms/limitations?

Less than 25% of the time	39.39% (13)
25 to 50% of the time	36.36% (12)
51 to 75% of the time	15.15% (5)
More than 75% of the time	9.09% (2)

N = 33

How often, if ever, does your DAC consult the original consent form when assessing the data use terms of a new access request?



Perceptions on Proposed Solutions



What are the benefits of using software to conduct data access reviews?

1. Record-keeping/audit-ability	**27.27% (30)**
2. Ensure consistency in review process	**22.73% (25)**
3. Reduce processing time	**20% (22)**
4. Improve interoperability	**17.27% (19)**
5. Save expenses	8.18% (9)
6. There are no benefits	2.73% (3)
7. Provide a second-opinion on our DAR decisions	1.82% (2)

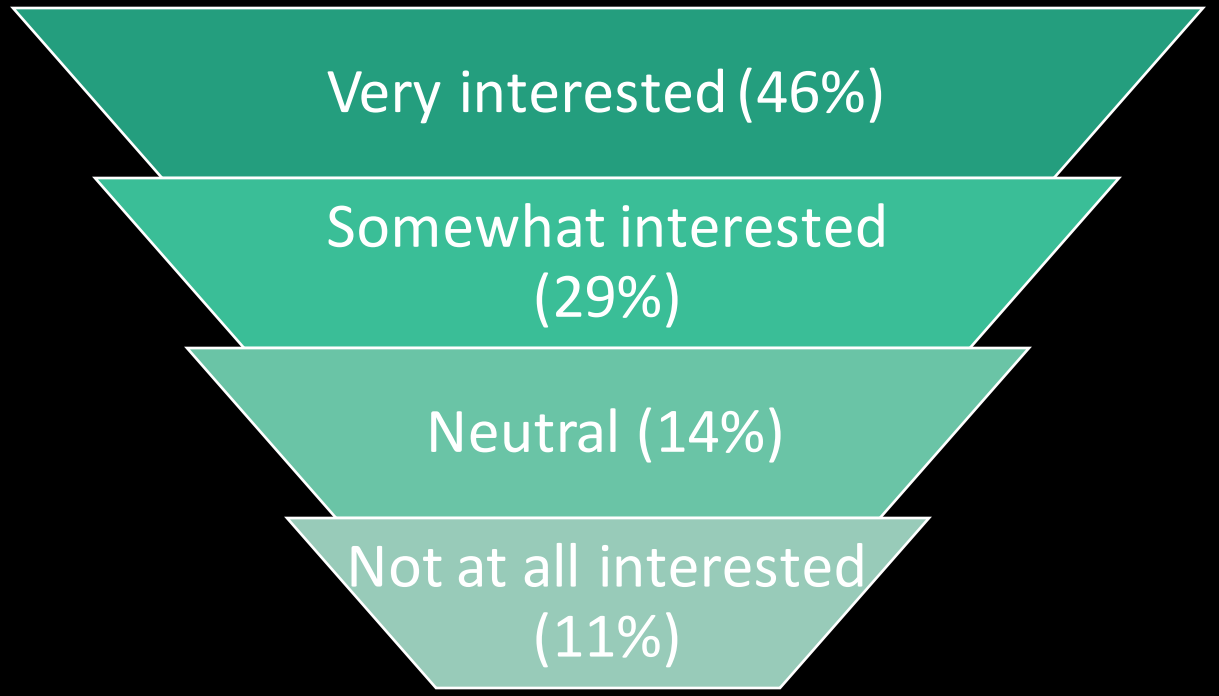
What benefits, if any, are there in having permitted uses for your datasets in a standardized format?

1. Reduce processing time	**22.39% (15)**
1. Ensure consistency in the review process and the results	**22.39% (15)**
2. Have a clear, shared understanding of the meaning of the data use categories (minimize misunderstanding among members of the data access committee, but also between the data access committee and requesters)	**20.90% (14)**
3. Improve interoperability	**16.42% (11)**
4. Reduce member workload	**14.93% (10)**
There are no benefits	2.99% (2)

What are the potential challenges of using software to conduct data access reviews?

1. Technically too complicated, difficult, or annoying to use/learn	**29.17% (14)**
2. Efficiency gains are insignificant or non-existent	**20.83% (10)**
3. Security Concerns	**18.75% (9)**
4. Concerned system will be inaccurate	14.58% (7)
5. All data access committee work should be conducted by humans alone, without any software or digital support	6.25% (3)
6. Lack of availability of such software	4.17% (2)
7. There are no challenges	4.17% (2)
8. Cannot be applicable to our diverse datasets	2.08% (1)

How interested would you be in a system* like DUOS?

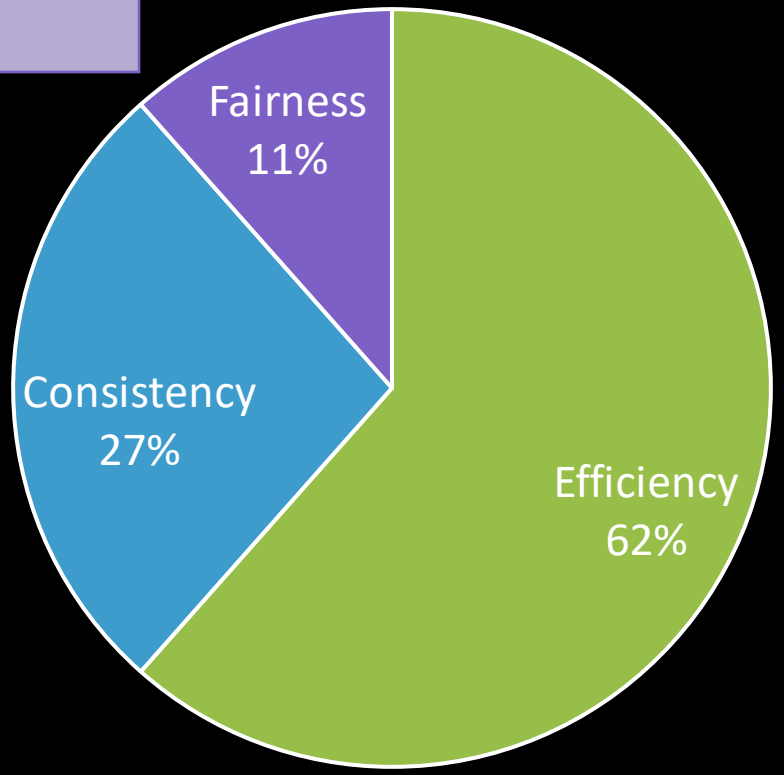


75% of respondents express interest in (semi-)automated systems

*A system that 1) checks the compatibility between data use terms (per the consent form) and intended uses (described in the data access request) and 2) displays these results to the data access committee to make their final decision.

Why are you interested?

- Increase transparency
- Highlight potential biases



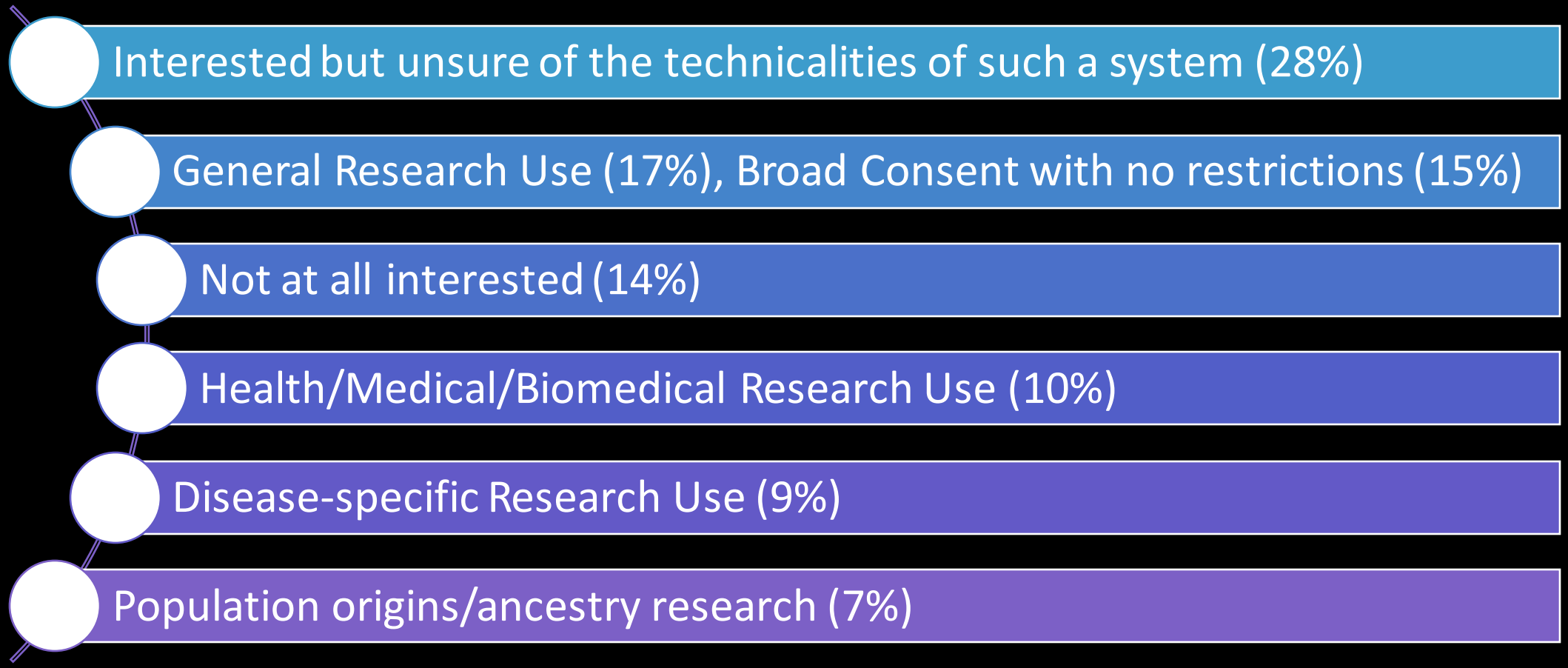
- Expedite & optimize review process
- Reduce workload
- Make review process easier for the requester
- Improve interoperability

- Help standardizing review criteria
- Reduce human error
- Help reviewers with less expertise in the field

Why are you not interested? Well… it depends

No efficiency gains (e.g., no data use terms in the consent forms, small sized DAC – software will only make DAC review more complicated)	**28.57% (4)**
Concerns about accuracy due to a lack of specificity and context	**21.43% (3)**
Do not have the authority to decide whether or not to use software for DAC review	14.29% (2)
Inconsistent with our institutional data access policy	7.14% (1)
Concerns about misunderstandings/misinterpretations by requesters (distrust in requesters)	7.14% (1)
Human expertise is needed	7.14% (1)
System would not work for our datasets (discrepancies in data use languages of a system and our datasets are too big)	7.14% (1)
Legal regulations prohibiting from adopting such software	7.14% (1)

Would you be interested in using a system that automatically approves/rejects data access requests without requiring human review for the following data use types?



Why are you interested in a system that automatically approves/reject DARs?

Efficiency (save time – more time for complicated cases, easy access)	35.71% (5)
Automation	21.43% (3)
Streamlining the review process (especially in relatively simple, straightforward cases e.g., broad use limitations)	14.29% (2)
Increase data access requests	7.14% (1)
Standardization	7.14% (1)
Improve interoperabilty	7.14% (1)
Record-keeping/auditability	7.14% (1)

Why are you NOT interested in a system that automatically approves/reject DARs?

Concerns about accuracy and reliability (due to lack of specificity and context)	22.22% (4)
Human expertise is needed	22.22% (4)
Contradictory to consents obtained from donors	11.11% (2)
Doubts in feasibility of automation	5.56% (1)
Distrust in data requesters' ability to understand data use conditions and local/social considerations	5.56% (1)
No efficiency gains (small-sized DAC - too little volume)	5.56% (1)
Undermining transparency (black box solution)	5.56% (1)
Legal restrictions	5.56% (1)
Concerns about public receptibility	5.56% (1)
Existence of in-house system (already in use)	5.56% (1)
Inconsistent with institutional policies	5.56% (1)

What did we learn?

Opportunities and challenges of hybridized human-machine governance systems to enhance quality, consistency, and fairness in data access decisions

Challenges to manual oversight for data access management and re-use

Problem of scale

Demand for access to highly integrated datasets is driving the need for realtime access management

Problem of resource constraints

Human data use oversight is costly, sometimes inconsistent and time-intensive

Problem of variation

Analysis of survey data suggest that DAC structure, membership, and decision-making vary widely and globally

Conclusion

Our survey of DAC members showed strong support for standardization, and the pressing need for efficient, consistent, and fair data access review. We also found that there is general support for automated approaches to reduce processing time and DAC member workload. Respondents were not uniform about their major concerns regarding such automation.

Thank you!

What will be the future scope of work for data access committees? Who will staff DACs in the future and what additional expertise, if any, will they require?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Should DACs consider harms to communities when evaluating requests for data access?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

In addition to standardizing data use terms using an ontology, what else can (or should) we standardize? What would be the necessary components that need to be standardized for more efficient, interoperable data access and sharing?

Top

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

⚠ When survey is active, respond at **pollev.com/vasilikir505**

What new tools should the DACReS working group prioritize developing next to improve the quality and effectiveness of DAC review?

0 done

0 underway

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Q&A / Discussion

1. What will be the future scope of work for data access committees? Who will staff DACs in the future and what additional expertise, if any, will they require?

2. Are DACs best equipped to consider both individual and group harms of access events and the data science products generated therefrom?

Q&A / Discussion

3. In addition to standardizing data use terms using an ontology, what else can (or should) we standardize? What would be the necessary components that need to be standardized for more efficient, interoperable data access and sharing?

4. What should the hybridized human-ADS (automated decision support) approaches look like? What will be the roles of human DACs be in the future?