Long-term Data Preservation Task Force

Data Sharing and Preservation session

EOSC Symposium

26 | 09 | 2023 by Hervé L'Hours & Roxanne Wyns

Agenda

- Introductory presentations
  1. Introducing federation in the context of EOSC and how long term preservation fits into it
     - Hervé L'Hours, Repository and Preservation Manager at UK Data Archive, LTDP-TF co-chair
     - Roxanne Wyns, Innovation Manager at KU Leuven – LIBIS, LTDP-TF co-chair
  2. Data federation and preservation in food composition data
     - Karl Presser, Founder and Managing Partner of Premotec GmbH
- Interactive session on LTDP recommendations
  - Looking into a selection of the task force recommendations
  - Poll + gathering comments

  This is a hybrid conference:
  - People in the room should use the mic to ask questions
  - TF colleagues are monitoring the online chat for questions
- Wrap-up and closing

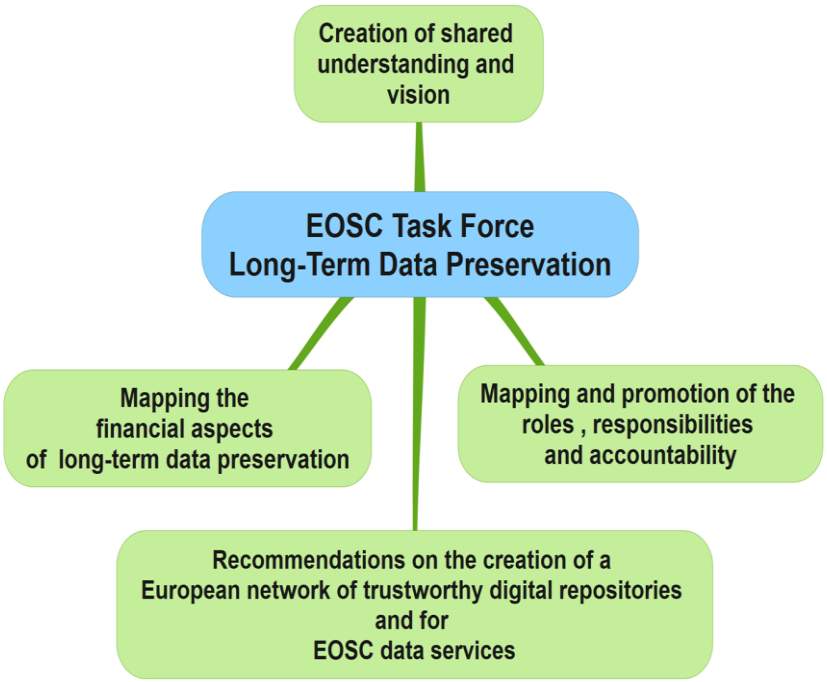
eosc | Long-term Data Preservation Task Force

Preservation in the context of EOSC

Data Sharing and Preservation session, EOSC Symposium

26 | 09 | 2023 by Hervé L'Hours & Roxanne Wyns

Task force structure



Co-chairs:
- Hervé L'Hours (UK Data Archive)
- Roxanne Wyns (KU Leuven)

EOSC-A liaison: Ignacio Blanquer (BoD)
EOSC-A support officer: Celia Alvarez

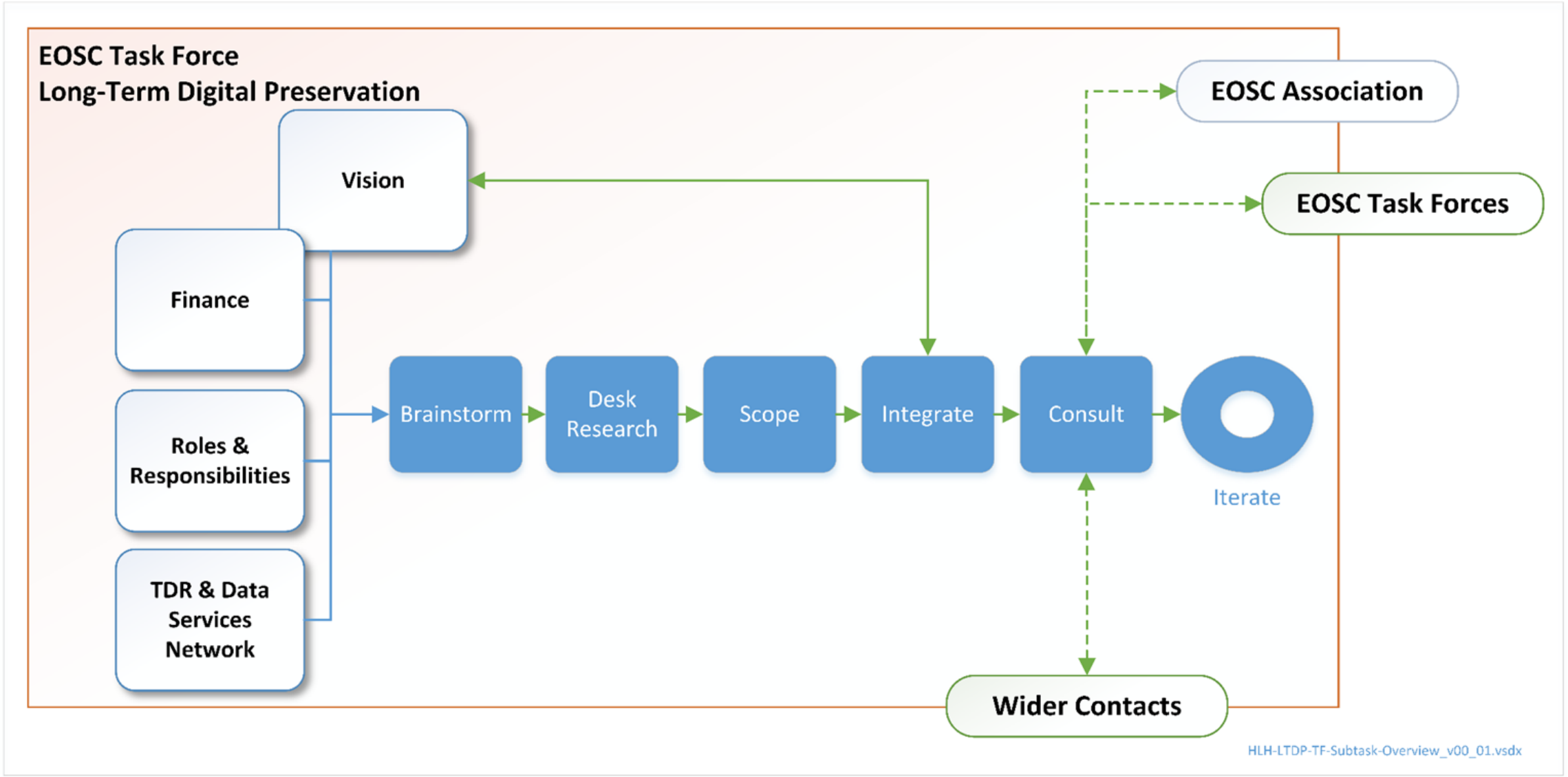
Shared Understanding and Vision
Subgroup lead: Chris De Loof (Belnet)
Pirjo-Leena Forsström (CSC), Rui Fernandes (C4G), Gerardo Ganis (CERN), Ville Tenhunen (EGI), Andras Holl (MTA), Maciej Brzeźniak (PSNC), Lara Lloret Iglesiasc (CSIC), Riccardo Smareglia (INAF), Eileen Gibney (UCD)

Mapping and promotion of roles
Subgroup lead: Mariusz Majdański (IG PAS)
Andrea Lammert (DKRZ), Sangeetha Shankar (DLR), Christian Cuciniello (EC), Marcello Maggi (INFN), Florina Piroi (TU WIEN), Mojib Wali (TU GRAZ), Sabine Crépé-Renaudin (CNRS), Bregt Saenen (Science Europe)

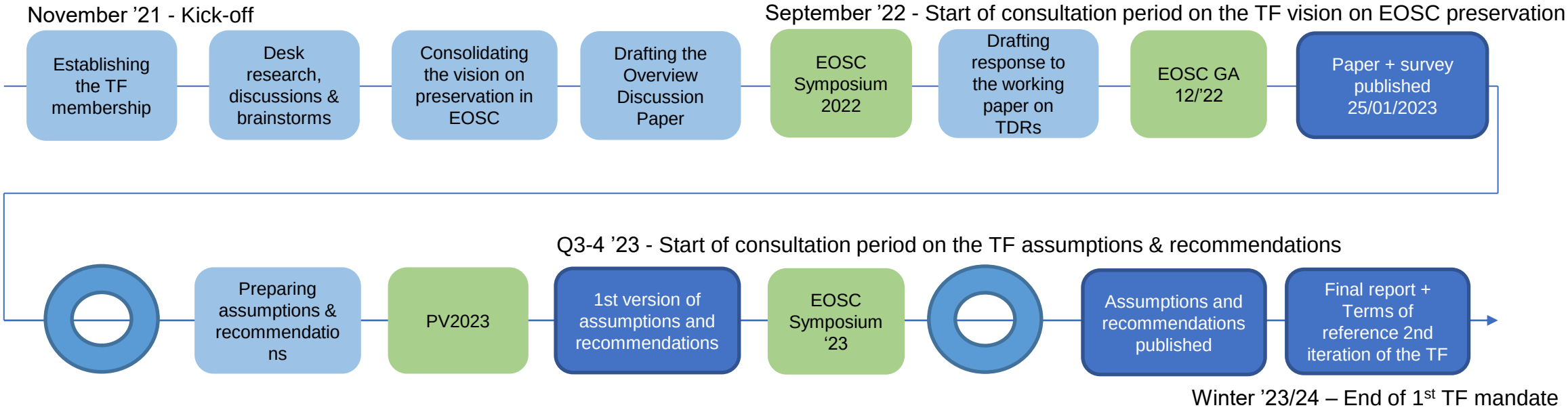
Mapping of financial aspects
Subgroup lead: Paul Stokes (JISC)
Jean-Yves Nief (CC-IN2P3), Jiri Novacek (CEITEC), Toni Andreu (EATRIS), Martina Stockhause (IPCC DDC)

Recommendations TDR
Subgroup lead: Didi Lamers (Radboud university)
Olesea Dubois (Sciences Po Paris), S. Venkataraman (OpenAire), Pierre-Yves Burgi (OLOS.swiss), Lluís Anglada (CSUC), Cécile Cavet (Univ. Paris), Ingrid Dillo (DANS), David Antos (CESNET), Draženko Celjak (SRCE), Matthew Viljoen (EGI)

Task force approach



Task force timeline



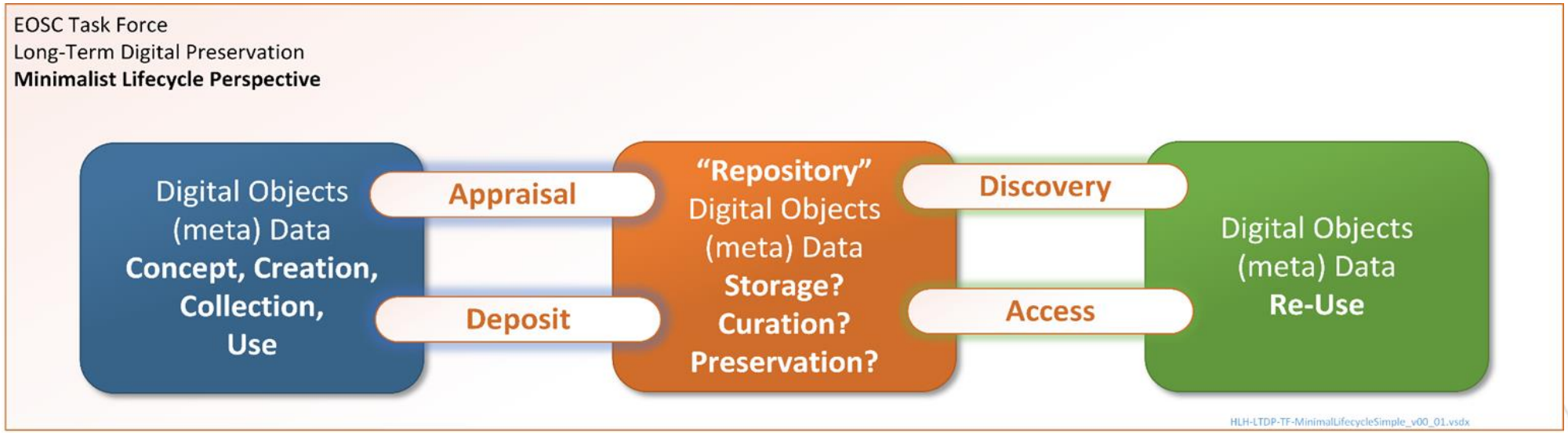
TF Outputs

- Overview discussion paper https://doi.org/10.5281/zenodo.7516259 + survey
- Response to the working paper “Toward a European network of FAIR-enabling Trustworthy Digital Repositories”. https://doi.org/10.5281/zenodo.7568399
- Preservation in the context of EOSC. Sustainable repositories curating digital objects from a long-term FAIR enabling perspective, PV2023 (pending publication)
- LTDP-TF Recommendations Consultation (pending publication)

Preservation in its wider context

Wide range of assumptions and perspectives around digital preservation

Contextualise preservation within the wider research life cycle and across the federated context of scientific infrastructure (meta)data services



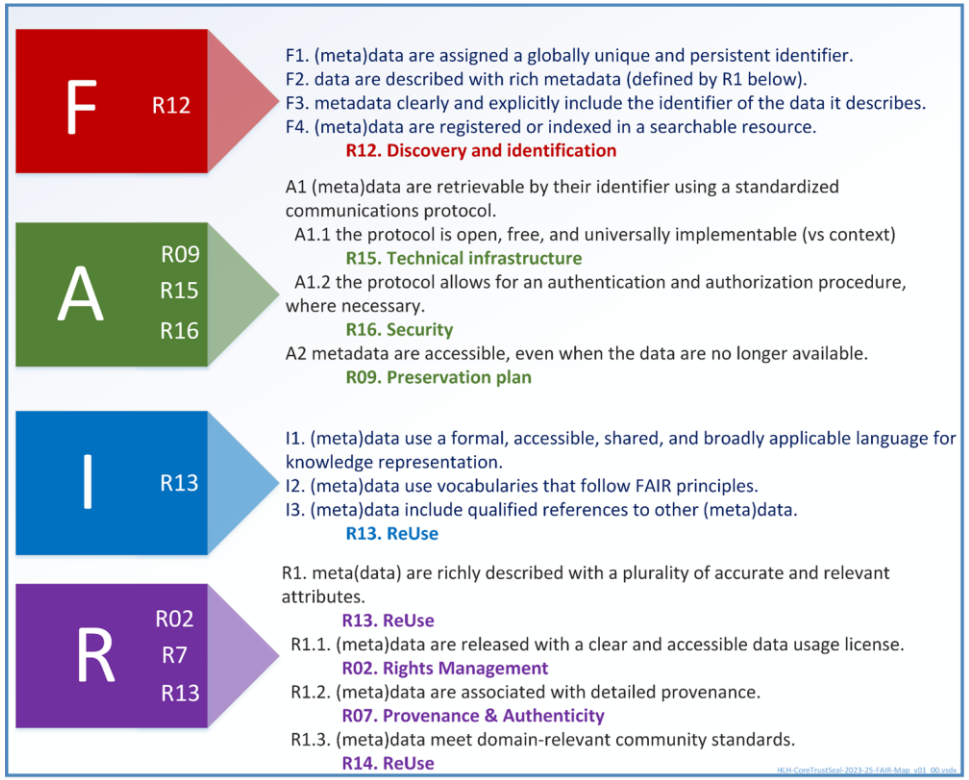
TRUST & FAIR in the context of the federation

Trustworthy Digital Repositories cover transparency, responsibility, sustainability and technology

Digital objects with a high potential for reuse are represented by FAIR

The need for data and metadata to *remain* FAIR is not explicit in the FAIR principles

FAIR-enabling repositories can ensure that digital objects remain FAIR over time through preservation



Assertions and recommendations

Repositories and (meta)data services

The EOSC is made up of a wide range of services directly or indirectly handling data and or metadata. Not all self-identify as ‘repositories’ and not all repositories offer active preservation. The TF has identified that addressing the broader range of (meta)data services in the ecosystem is necessary to deliver EOSC.

Levels of Care

Assumptions and interpretations of what constitute ‘preservation’ vary. The TF has identified that transparently defining the service tiers offered by different (meta)data services, including repositories is necessary to deliver EOSC. These range from simply providing access to retained digital objects, to full active preservation for the long term.

Recommendations formulated to be

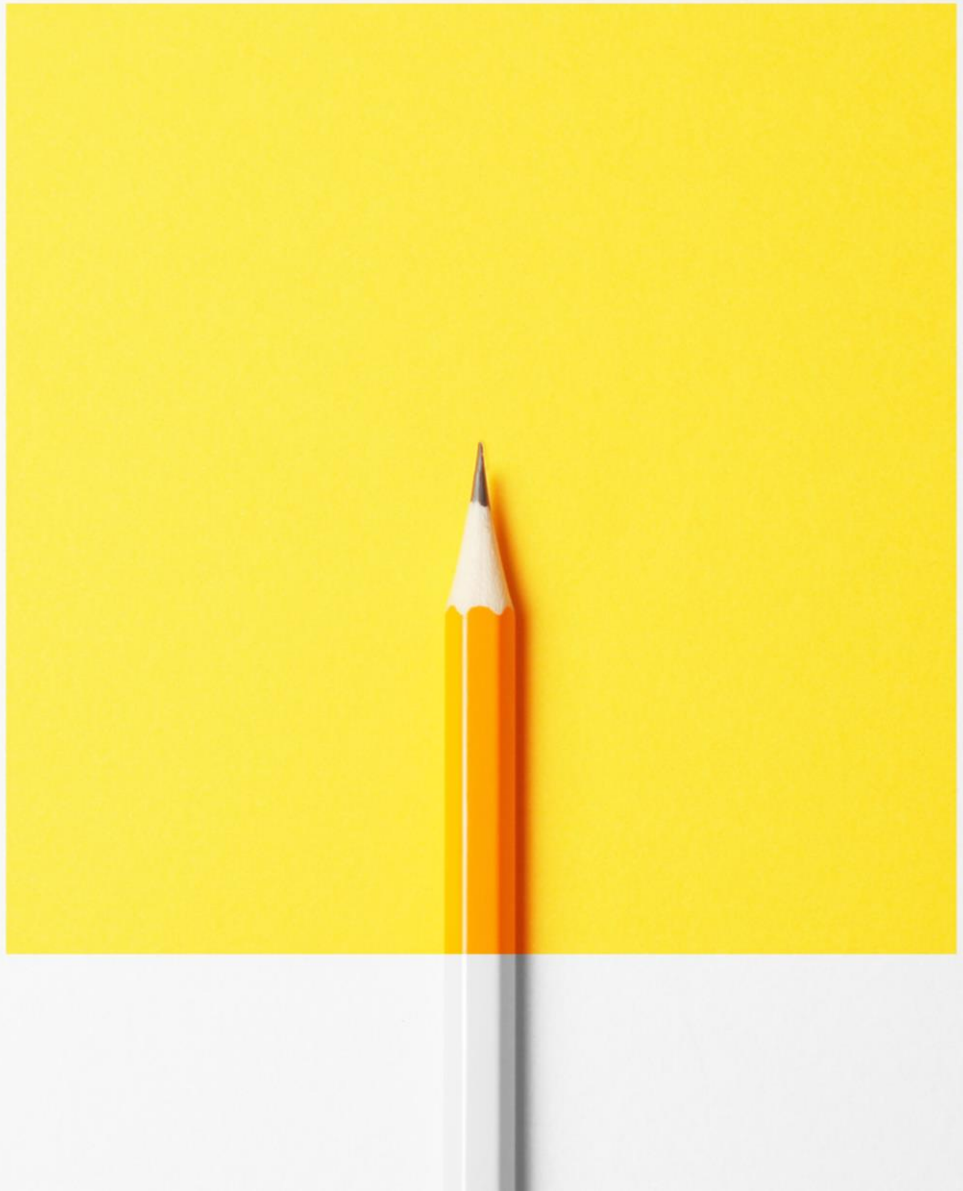
- Time independent
- Generally applicable to influence policies and provide guidance

Consultation

Consultation via

- EOSC Symposium → focussed on sub-set of recommendations (agreement scale + comment)
- Webinar sessions → focussed on sub-set of recommendations (agreement scale + comment)
- Feedback template → possible to provide feedback on all assertions and recommendations

Results of consultation period to be used to revise assertions and recommendations and inform the final report



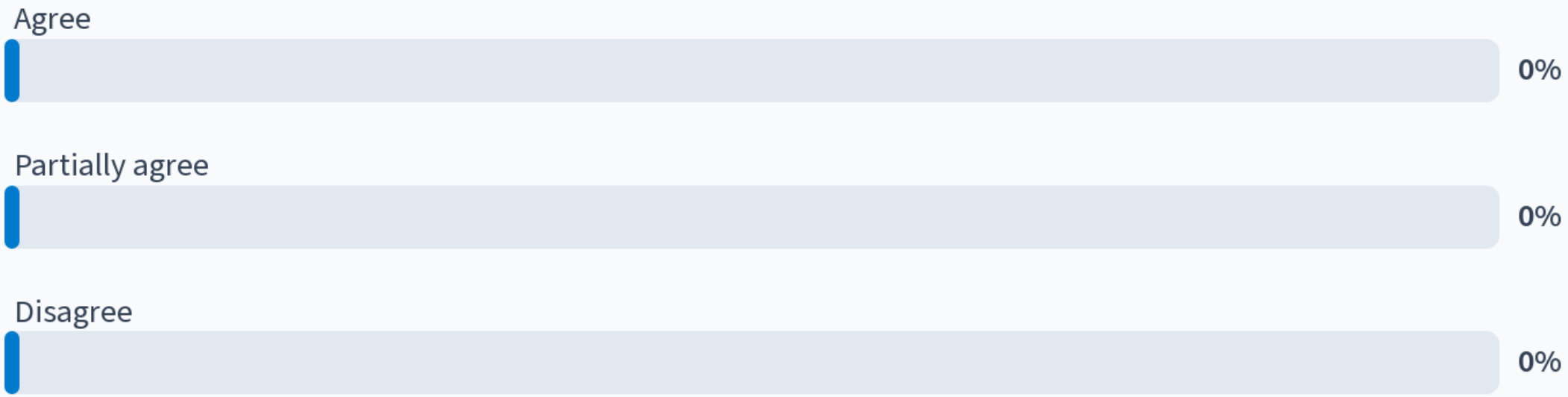
Assertions and Recommendations for Optimal Preservation Outcomes

We love to get your feedback!

Go to https://pollev.com/roxannewyns154

- Response options:
  - Agree
  - Partially agree
  - Disagree
- Leave your comments

@Data services, including repositories should specify all of the levels of care they apply to objects within their collection, including through registry metadata



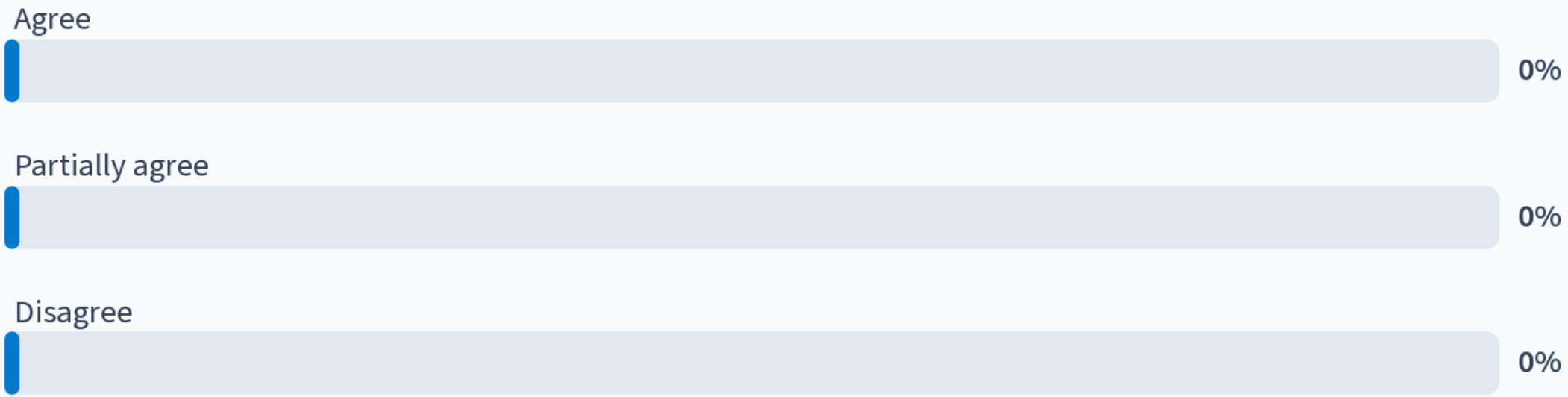
Topic: Curation & Preservation Levels - @19

Comments @Data services, including repositories should specify all of the levels of care they apply to objects within their collection, including through registry metadata

Nobody has responded yet.

Hang tight! Responses are coming in.

@Digital objects should include metadata that specify their level of care and the timeframes or criteria for reappraisal of the level of care



Topic: Curation & Preservation Levels - @20

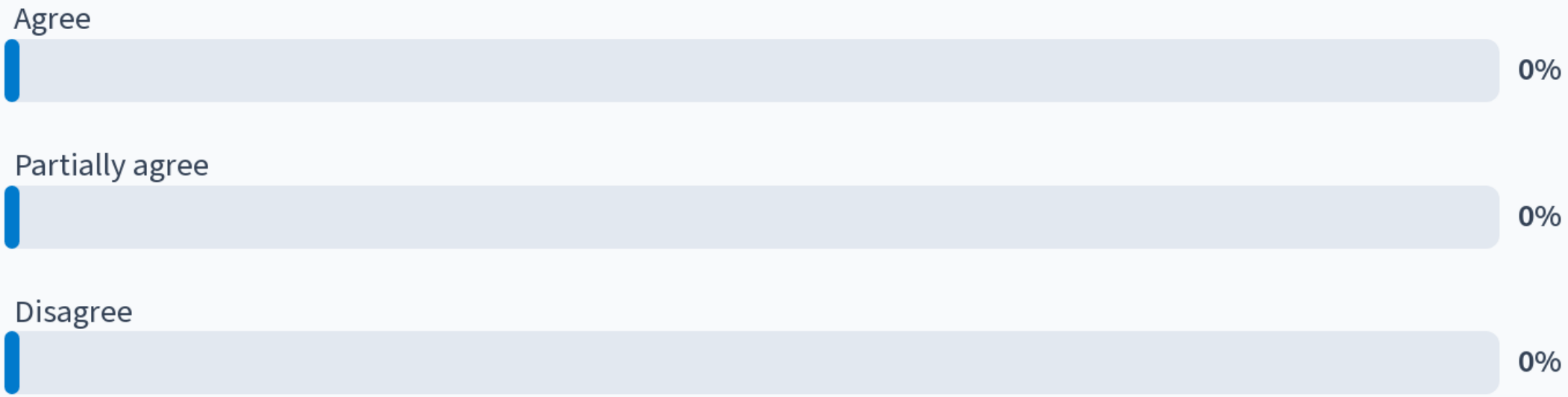
Comments @Digital objects should include metadata that specify their level of care and the timeframes or criteria for reappraisal of the level of care

Nobody has responded yet.

Hang tight! Responses are coming in.

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

@FAIR-enabling practices undertaken by data services should be made transparent to users and funders to increase trust in services



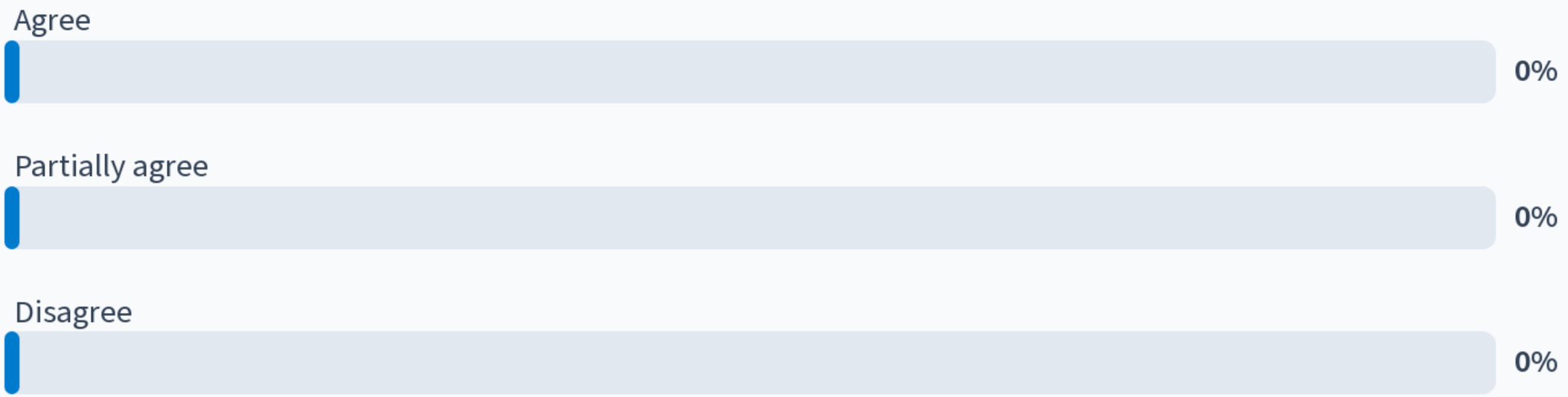
Topic: FAIR + Time - @24

Comments @FAIR-enabling practices undertaken by data services should be made transparent to users and funders to increase trust in services

Nobody has responded yet.

Hang tight! Responses are coming in.

@ Standards and guidance should be developed, coordinated and maintained to provide full lifecycle information on preservation to researchers and preservation practitioners



Topic: Standards, Assessment, Certification & Knowledge - @ 48

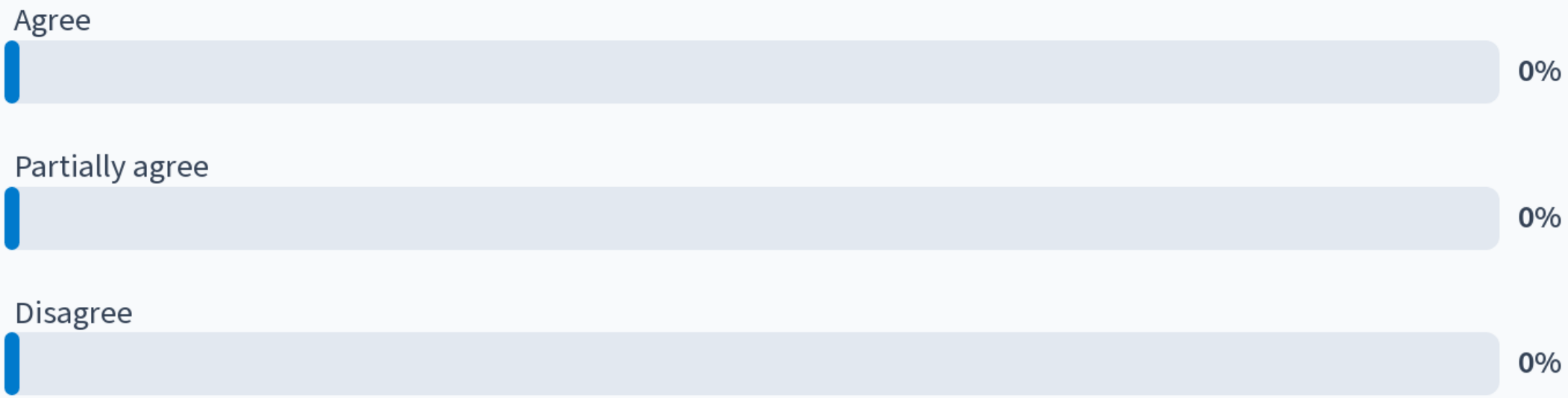
Comments @ Standards and guidance should be developed, coordinated and maintained to provide full lifecycle information on preservation to researchers and preservation practitioners

Nobody has responded yet.

Hang tight! Responses are coming in.

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

@Retention and reappraisal decisions and timescales, including guaranteed preservation timescales, should be transparent



Topic: Retention, Appraisal & Re-Appraisal - @61

Comments @Retention and reappraisal decisions and timescales, including guaranteed preservation timescales, should be transparent

Nobody has responded yet.

Hang tight! Responses are coming in.

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

Topic: Digital Objects, Research Projects, Repository and (meta)data Services - @16

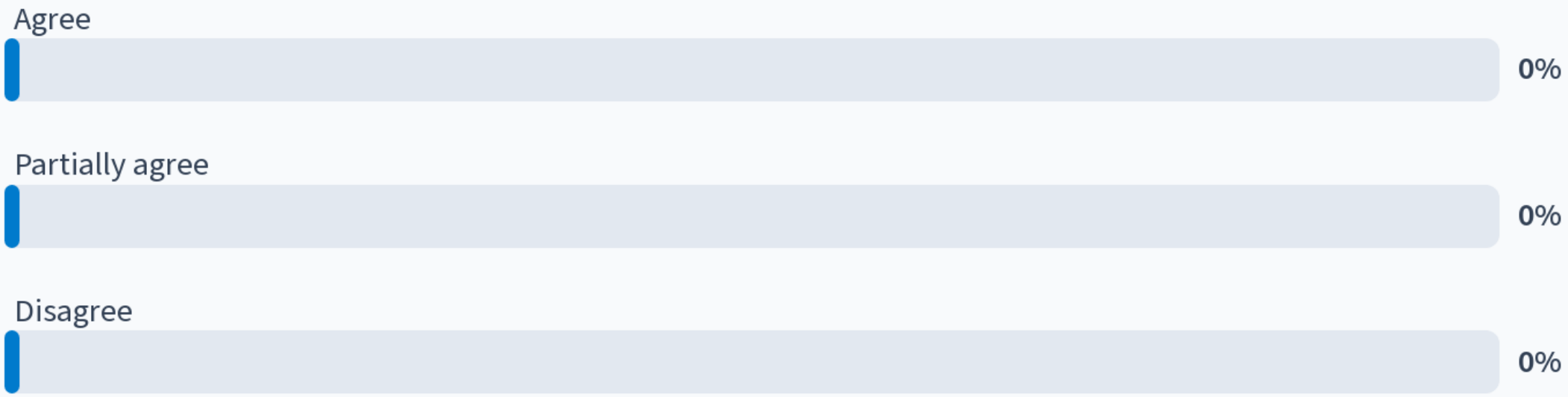
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Comments @Minimum criteria for acceptable storage practices in different scenarios should be defined as a foundation for all tiers of retention, curation and preservation services

Nobody has responded yet.

Hang tight! Responses are coming in.

@Data and metadata creators, collectors and reusers, including researchers, should develop the knowledge and skills at a general level and within their own disciplinary and domain area of expertise, so that their actions are preservation-aware



Topic: Roles and Responsibilities - @90

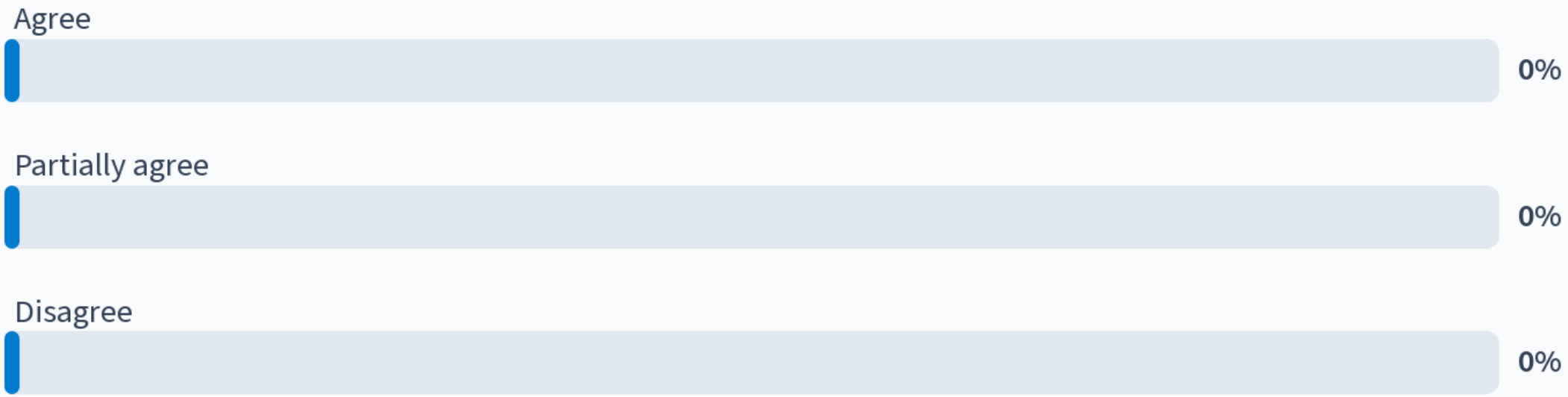
Comments @Data and metadata creators, collectors and reusers, including researchers, should develop the knowledge and skills at a general level and within their own disciplinary and domain area of expertise, so that their actions are preservation-aware

Nobody has responded yet.

Hang tight! Responses are coming in.

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

@Full lifecycle costs should be assessed, including the unique costs of preservation using methodologies that include the potential costs of inaction



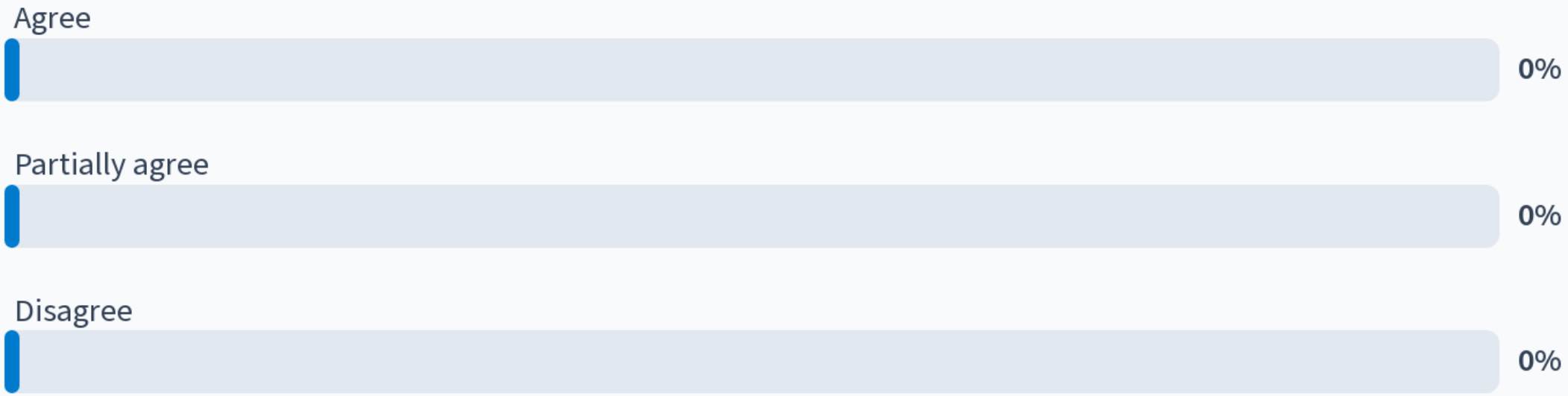
Topic: Finance & Funders - @102

Comments @Full lifecycle costs should be assessed, including the unique costs of preservation using methodologies that include the potential costs of inaction

Nobody has responded yet.

Hang tight! Responses are coming in.

@Other ongoing work to define repository and data service characteristics and expectations exist and should be encouraged and supported



Topic: Trustworthy Digital Repositories (TDR) and other (meta)data Services - @34

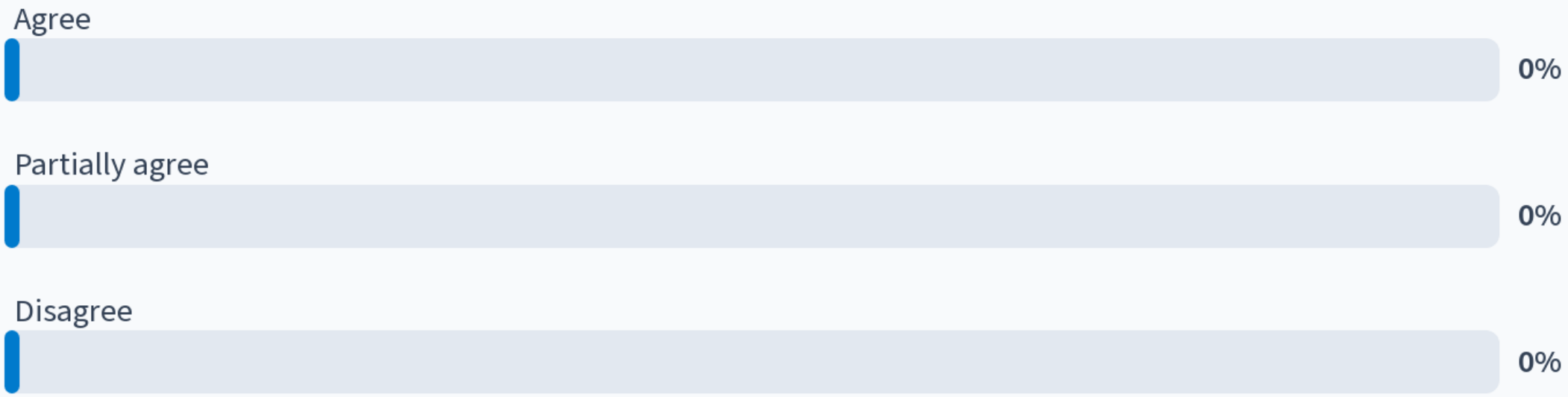
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Comments @Other ongoing work to define repository and data service characteristics and expectations exist and should be encouraged and supported

Nobody has responded yet.

Hang tight! Responses are coming in.

@To maintain clarity and alignment these other efforts should map and crosswalk their own criteria to the CoreTrustSeal. Any reductions, additions or variations versus the CoreTrustSeal should be documented and explained ...



Topic: Trustworthy Digital Repositories (TDR) and other (meta)data Services - @35

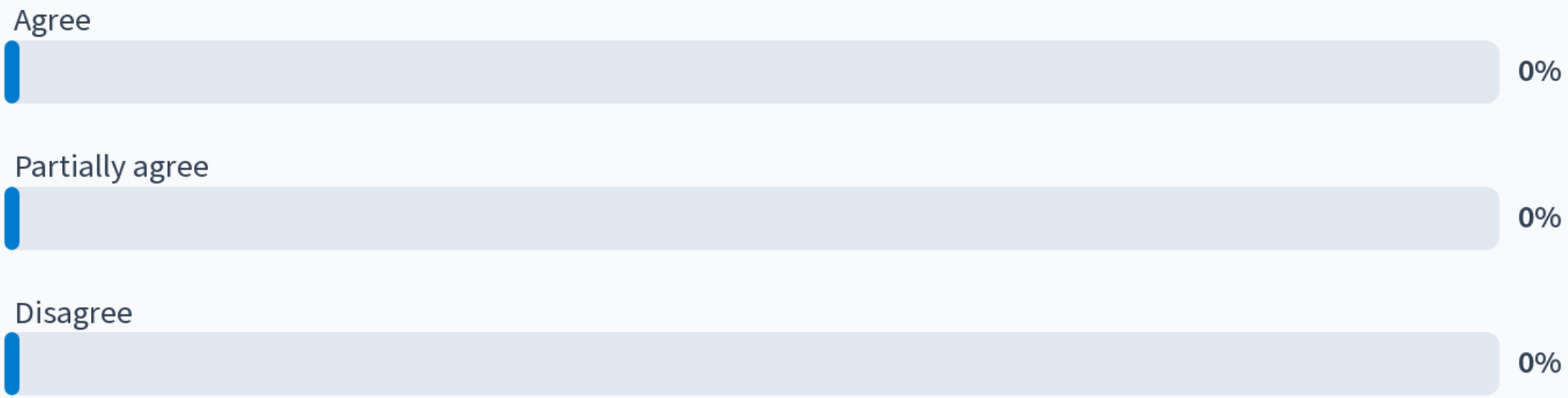
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Comments @To maintain clarity and alignment these other efforts should map and crosswalk their own criteria to the CoreTrustSeal. Any reductions, additions or variations versus the CoreTrustSeal should be documented and explained ...

Nobody has responded yet.

Hang tight! Responses are coming in.

@Living and machine actionable data management plans should form the basis of continuity through the research data lifecycle



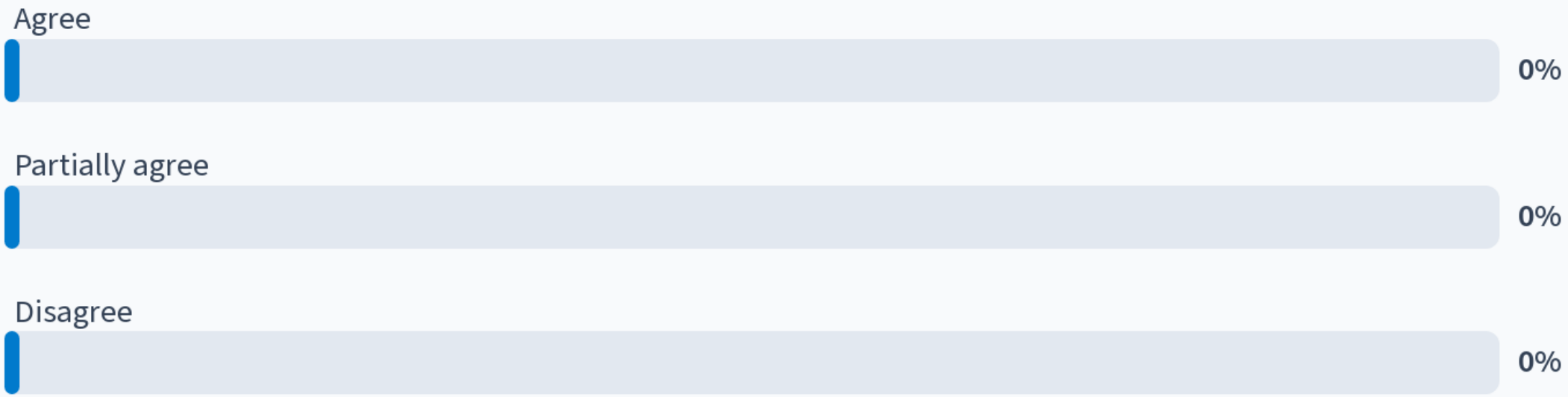
Topic: Transparency of Services, Artefacts and Levels of Care for Objects - @67

Comments @Living and machine actionable data management plans should form the basis of continuity through the research data lifecycle

Nobody has responded yet.

Hang tight! Responses are coming in.

@Registries of digital objects should align metadata relating to retention periods, appraisal periods and levels of curation and preservation



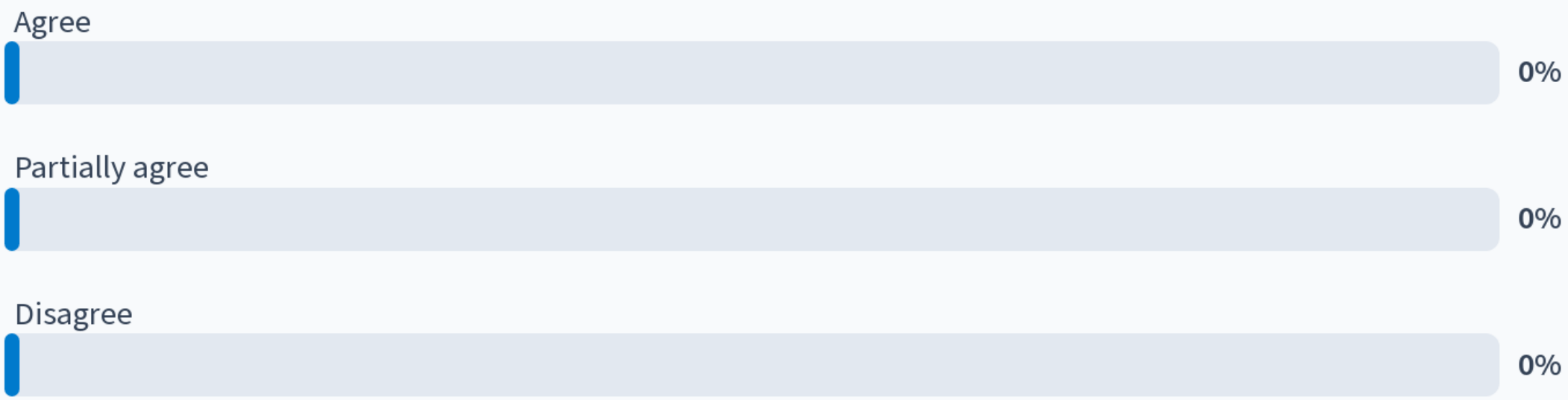
Topic: Transparency of Services, Artefacts and Levels of Care for Objects - @69

Comments @Registries of digital objects should align metadata relating to retention periods, appraisal periods and levels of curation and preservation

Nobody has responded yet.

Hang tight! Responses are coming in.

@Efforts by repositories and other data services to share transparent information about their functions, activities and objects should be rewarded by targeted investment towards improvement



Topic: Outcomes, Judgement and Gatekeeping - @53

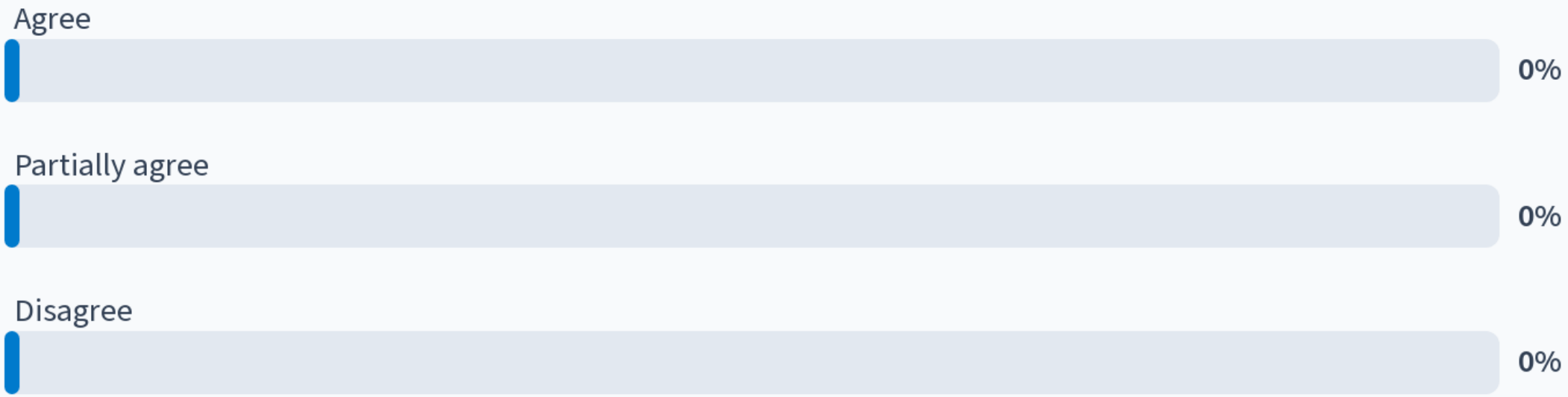
Comments @Efforts by repositories and other data services to share transparent information about their functions, activities and objects should be rewarded by targeted investment towards improvement

Nobody has responded yet.

Hang tight! Responses are coming in.

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

@No data service or digital object should be unnecessarily excluded from any part of the EOSC



Topic: Outcomes, Judgement and Gatekeeping - @55

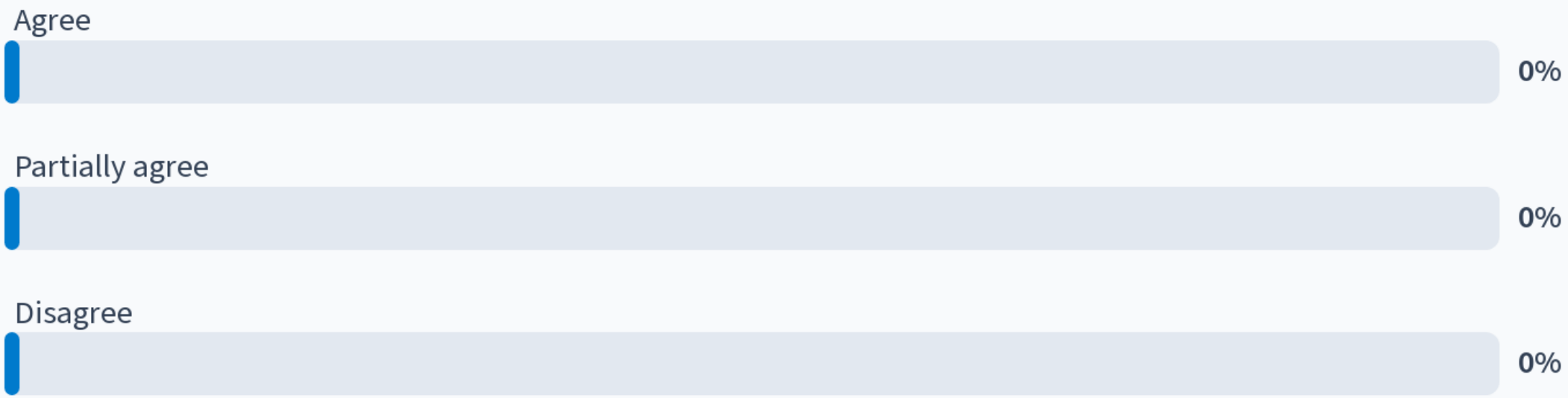
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Comments @No data service or digital object should be unnecessarily excluded from any part of the EOSC

Nobody has responded yet.

Hang tight! Responses are coming in.

@ Unique preservation functions and activities should be defined alongside functions and activities that apply to all (meta)data services



Topic: Preservation Systems, Actions & Outcomes - @29

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

Comments @ Unique preservation functions and activities should be defined alongside functions and activities that apply to all (meta)data services

Nobody has responded yet.

Hang tight! Responses are coming in.

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

Wrap-up

eosc | Long-term Data Preservation Task Force

THANK YOU

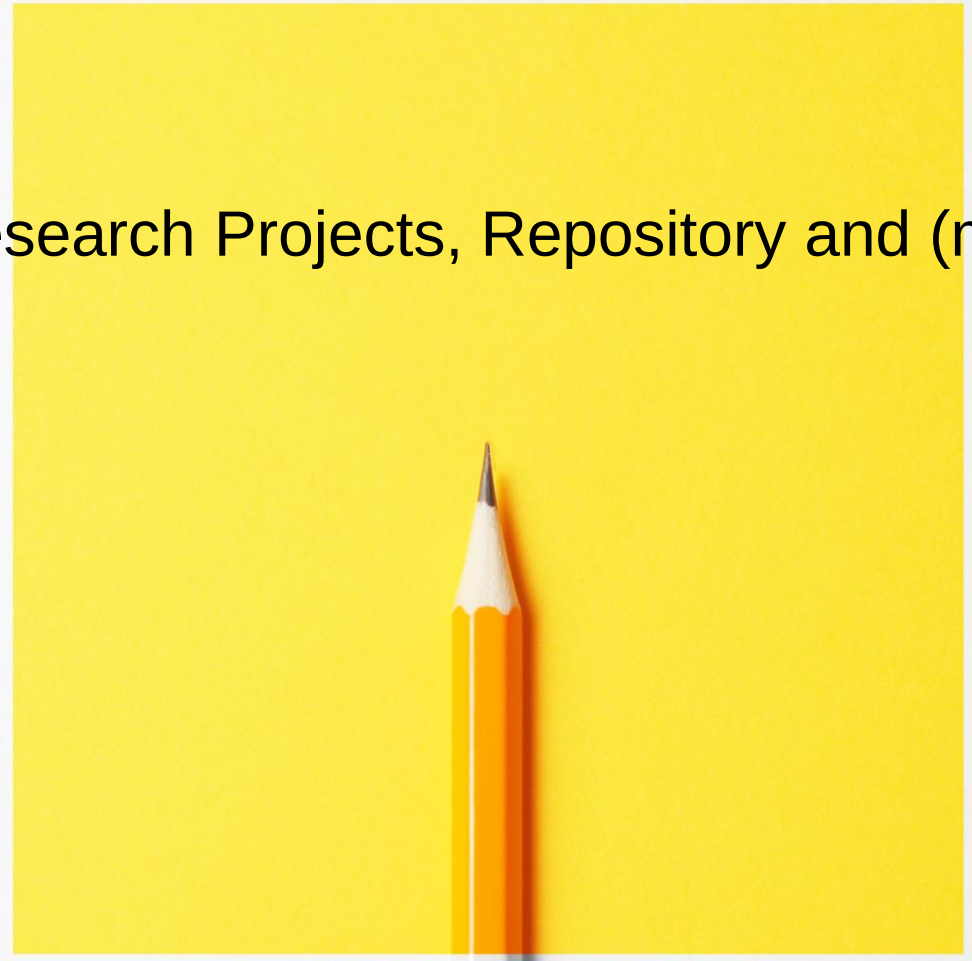
26 | 09 | 2023 by Hervé L'Hours & Roxanne Wyns

FULL OVERVIEW OF REQUIREMENTS PER HEADING

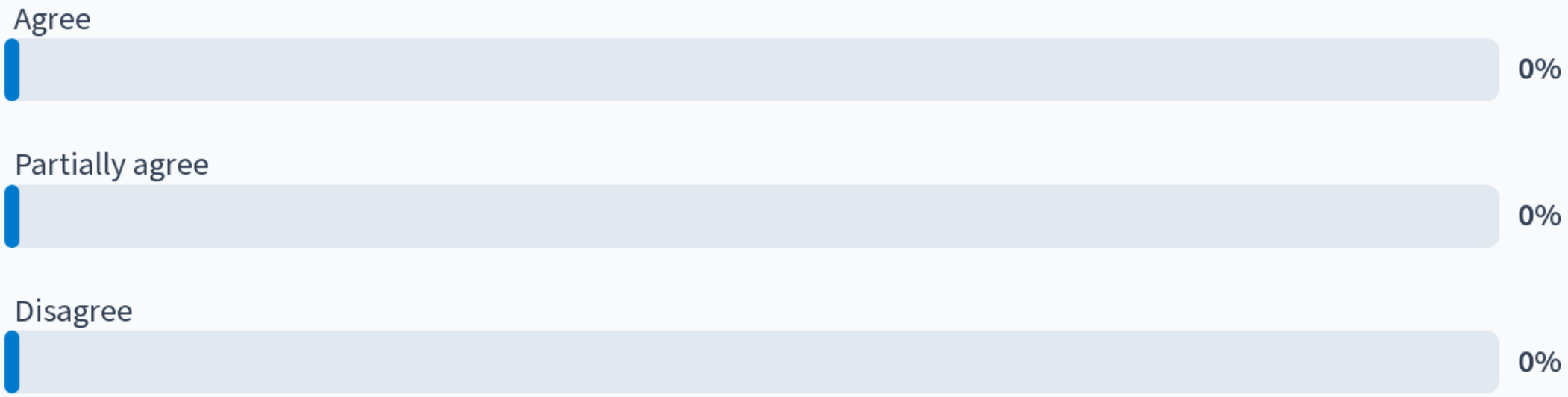
Favourites to discuss. Select 5 topics of preference

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

Digital Objects, Research Projects, Repository and (meta)data Services

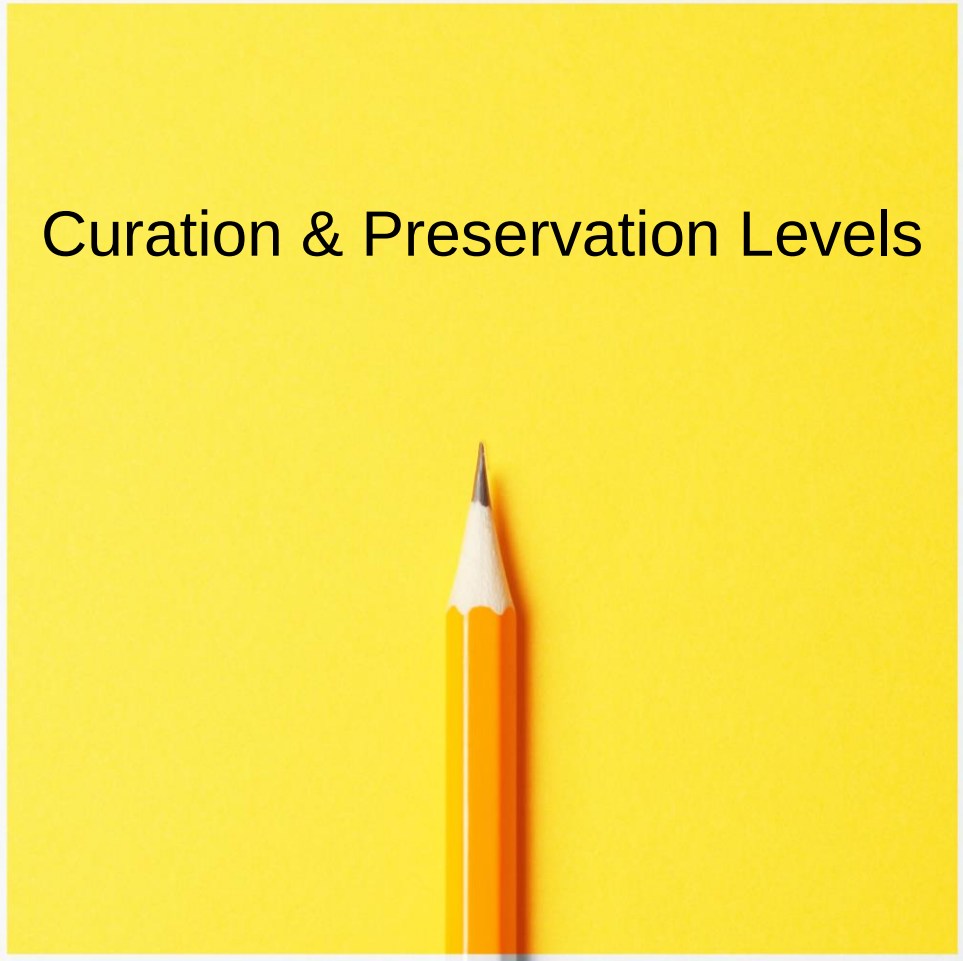


@Additional work is needed to define different types of digital object systems that offer (meta)data services, including the functions and activities they require and the levels of care they provide for data and metadata

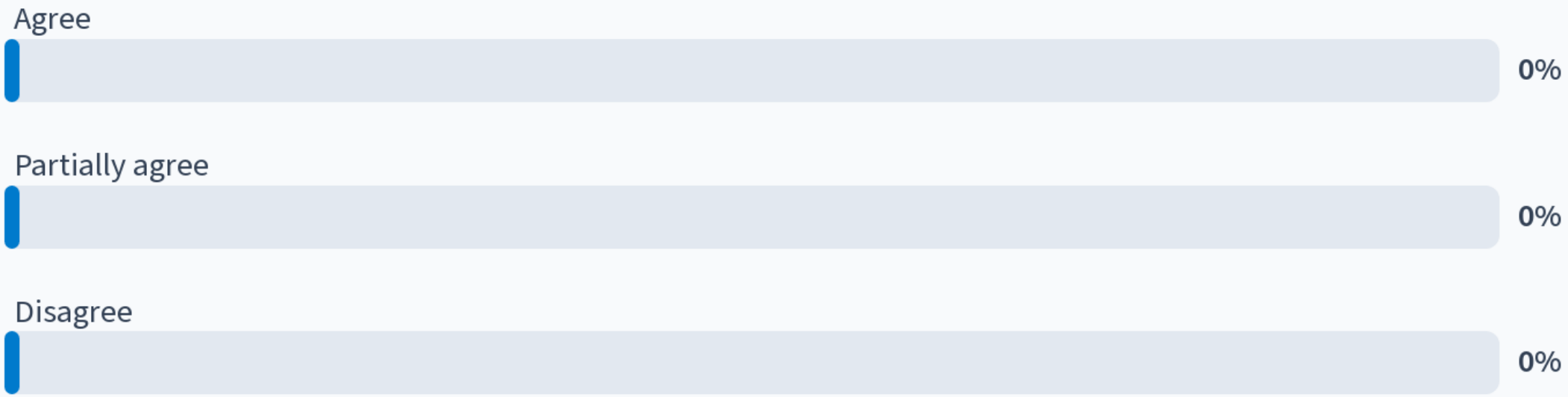


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

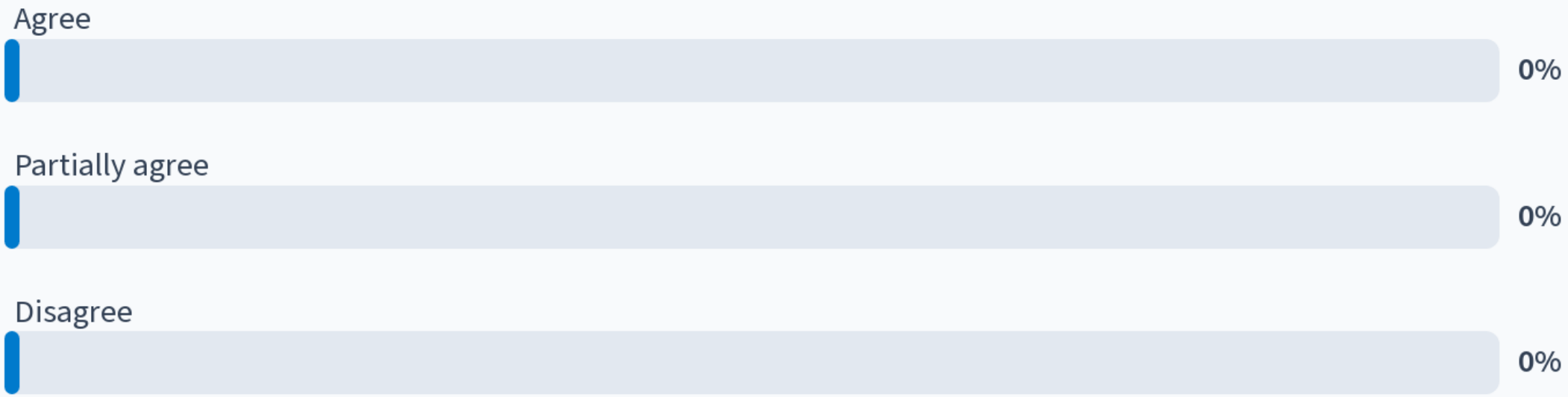
Curation & Preservation Levels



@Data services, including repositories should specify all of the levels of care they apply to objects within their collection, including through registry metadata

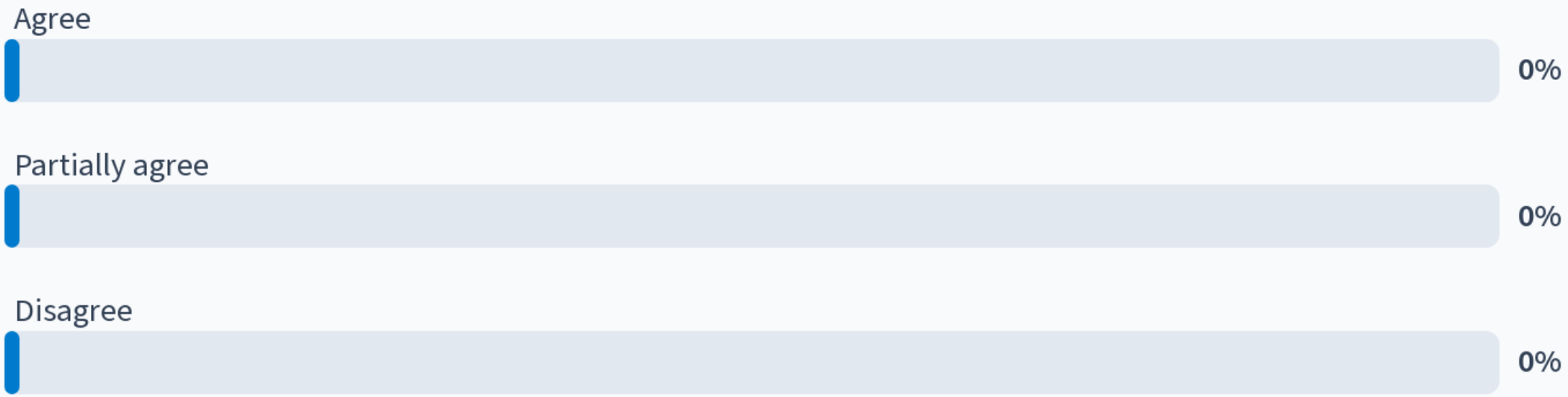


@Digital objects should include metadata that specify their level of care and the timeframes or criteria for reappraisal of the level of care

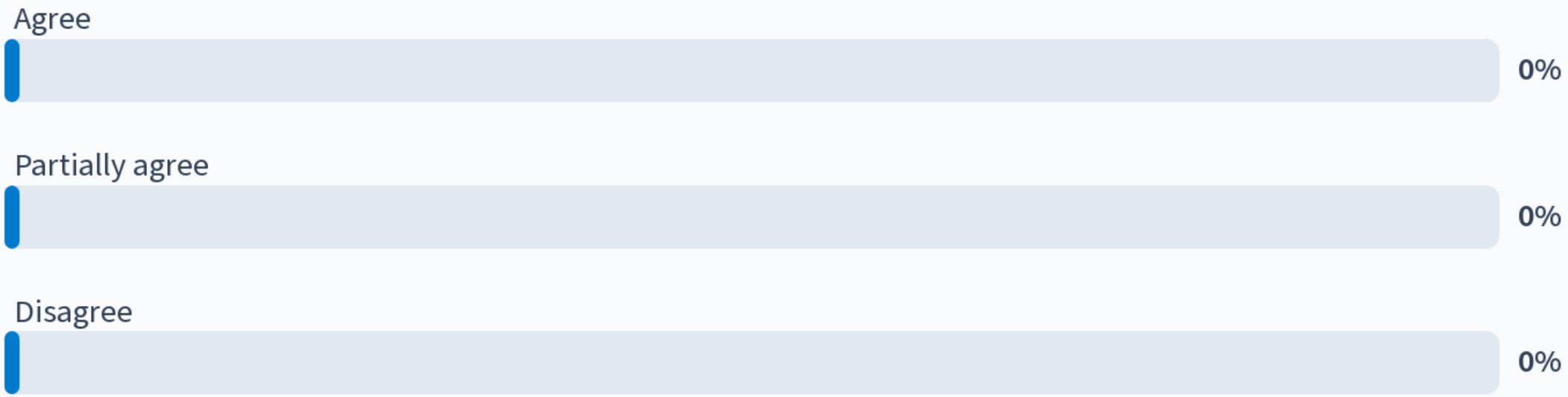


FAIR + Time

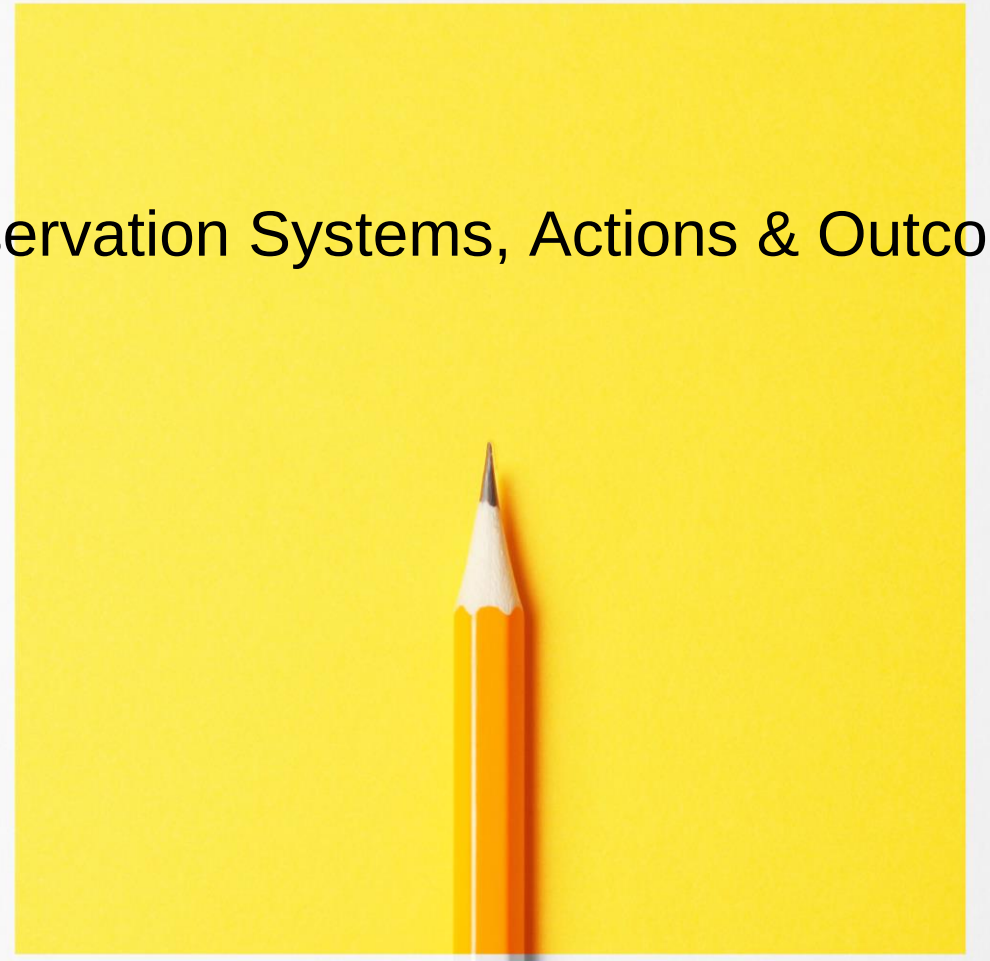
@FAIR-enabling practices to be undertaken by data services should be defined



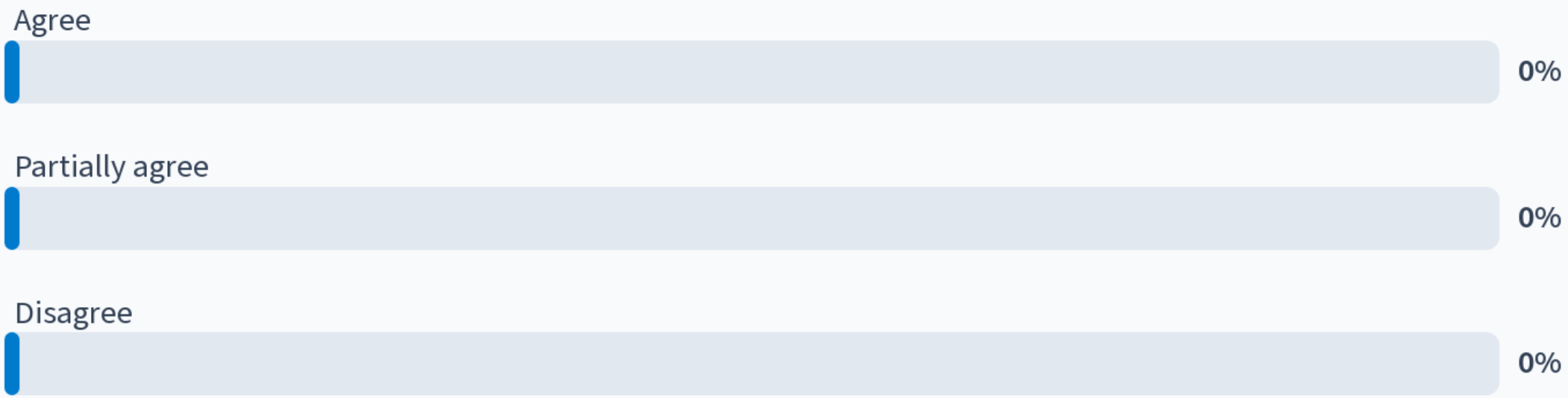
@FAIR-enabling practices undertaken by data services should be made transparent to users and funders to increase trust in services



Preservation Systems, Actions & Outcomes

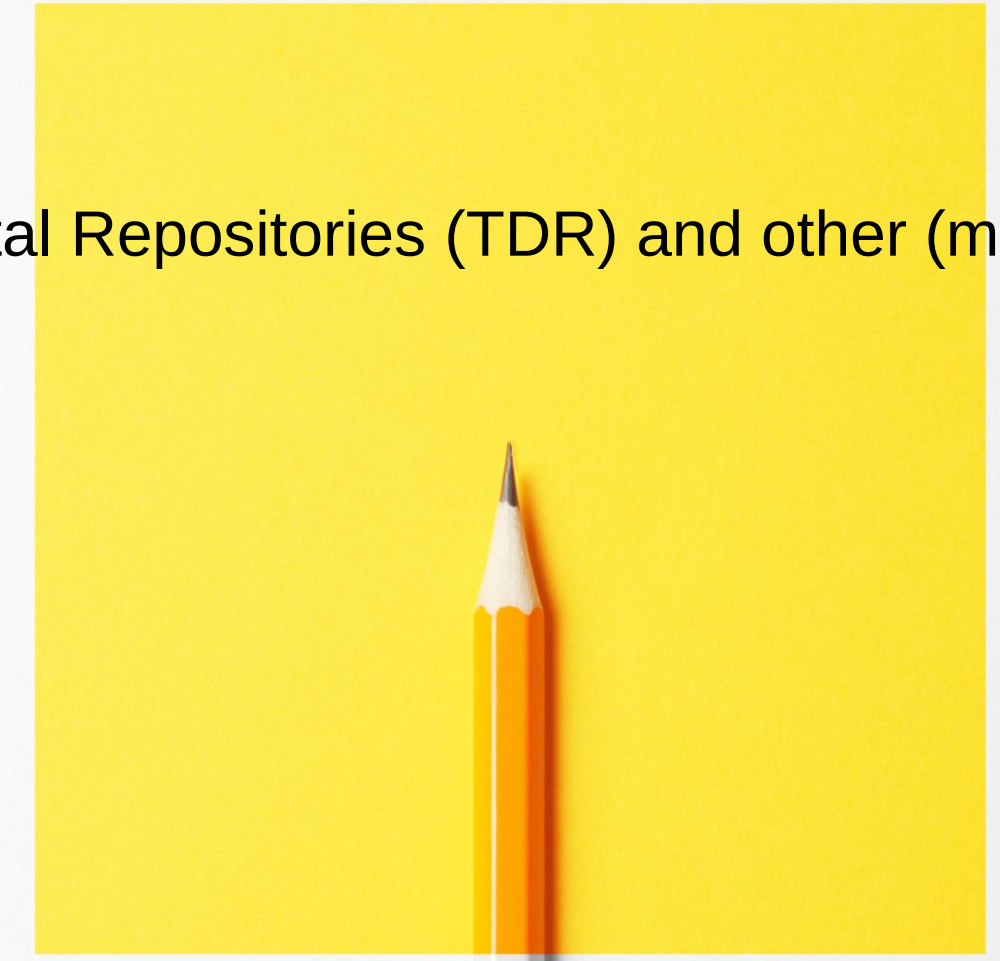


@ Unique preservation functions and activities should be defined alongside functions and activities that apply to all (meta)data services

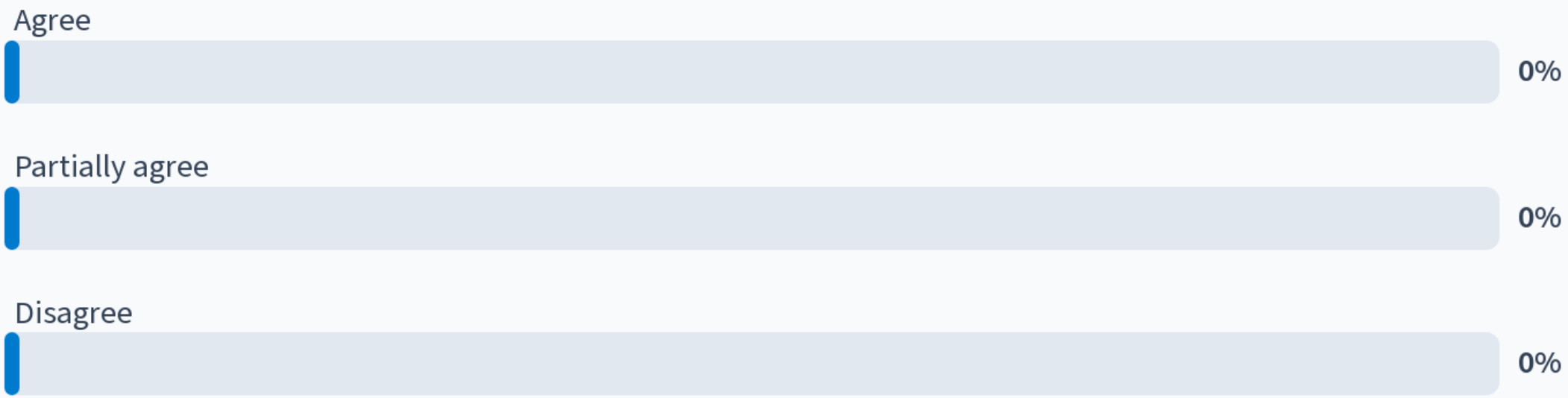


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Trustworthy Digital Repositories (TDR) and other (meta)data Services

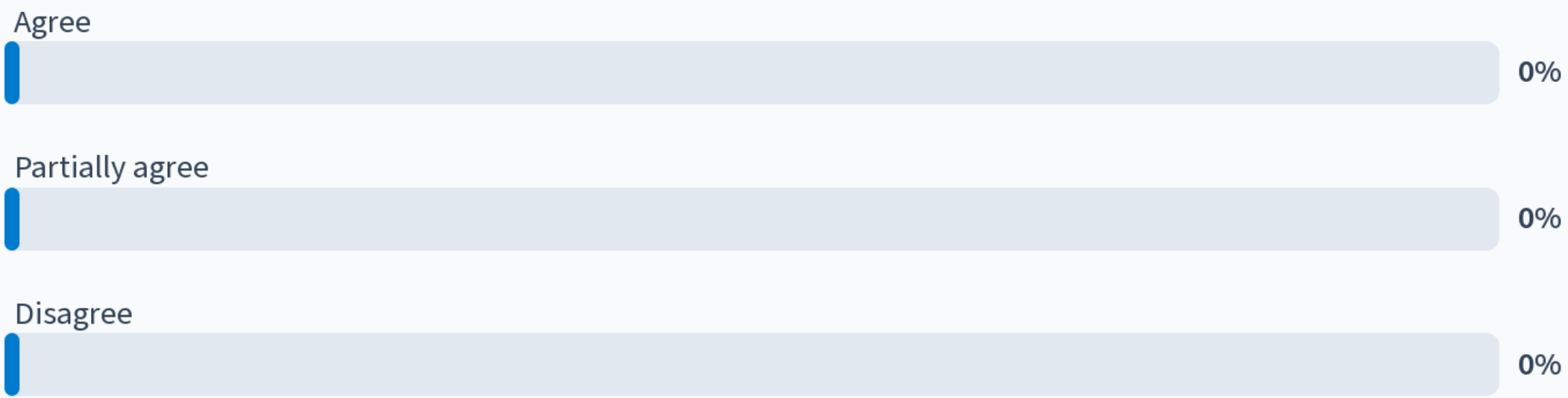


@Other ongoing work to define repository and data service characteristics and expectations exist and should be encouraged and supported



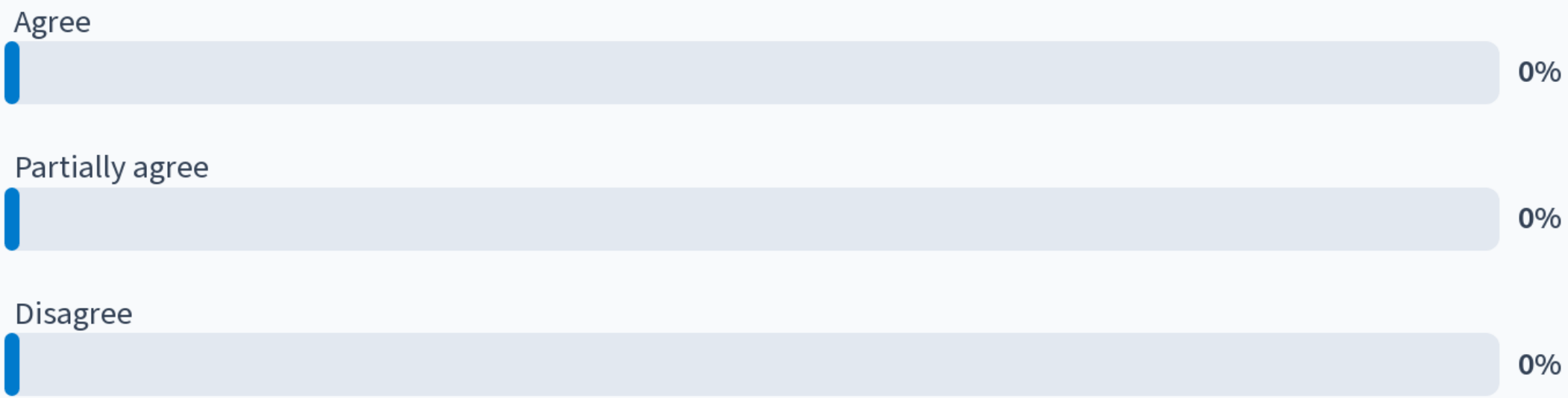
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

@To maintain clarity and alignment these other efforts should map and crosswalk their own criteria to the CoreTrustSeal. Any reductions, additions or variations versus the CoreTrustSeal should be documented and explained ...



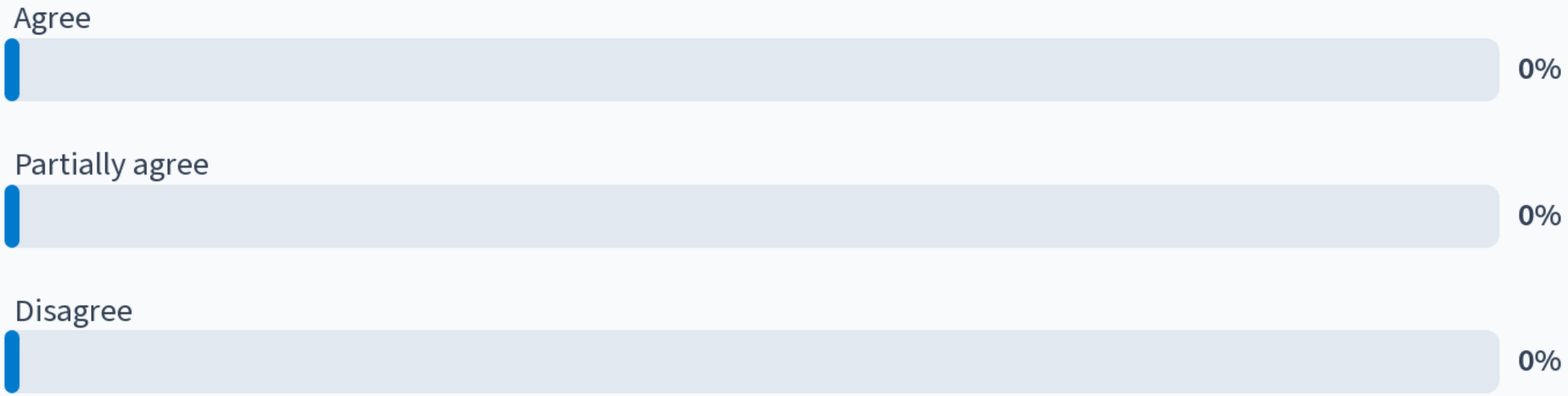
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

@Efforts to define more specific domain/disciplinary criteria, or criteria that define expectations for specific types of digital objects should adopt the CoreTrustSeal requirements where possible, and elaborate around them where necessary



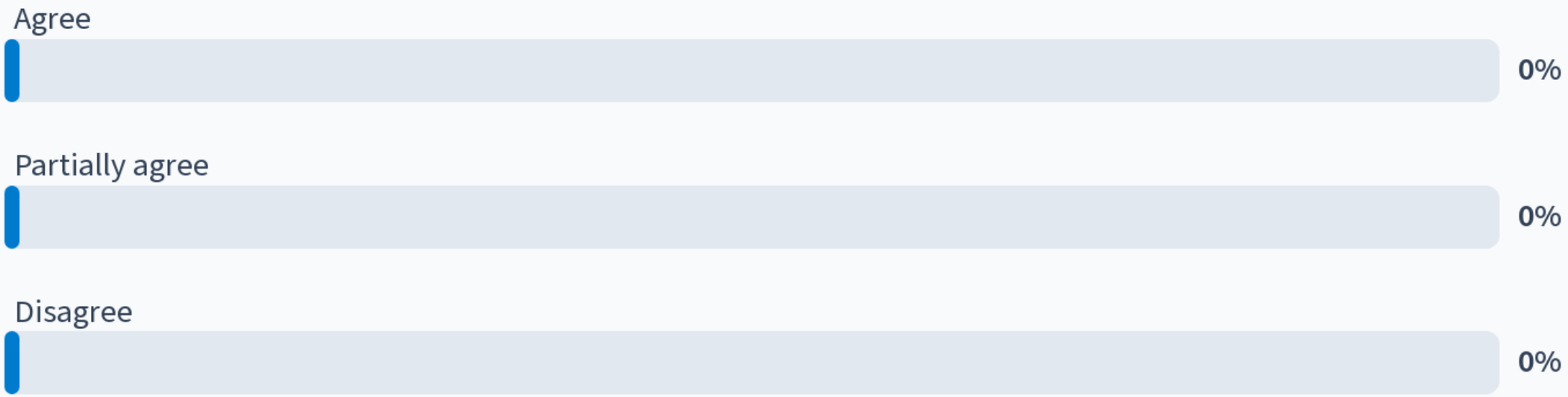
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

@Additional work is needed to define different types of digital object systems, the functions and activities they undertake and the levels of care they provide for data and metadata



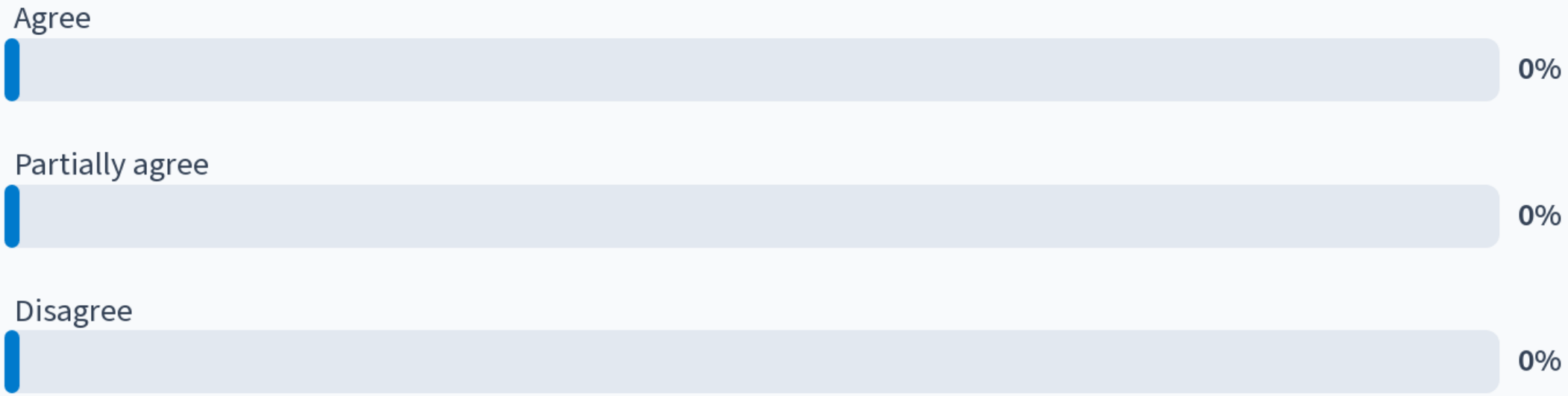
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

@Roles and responsibilities including for complex partnerships, third party relationships and outsourcing should be understood and transparent

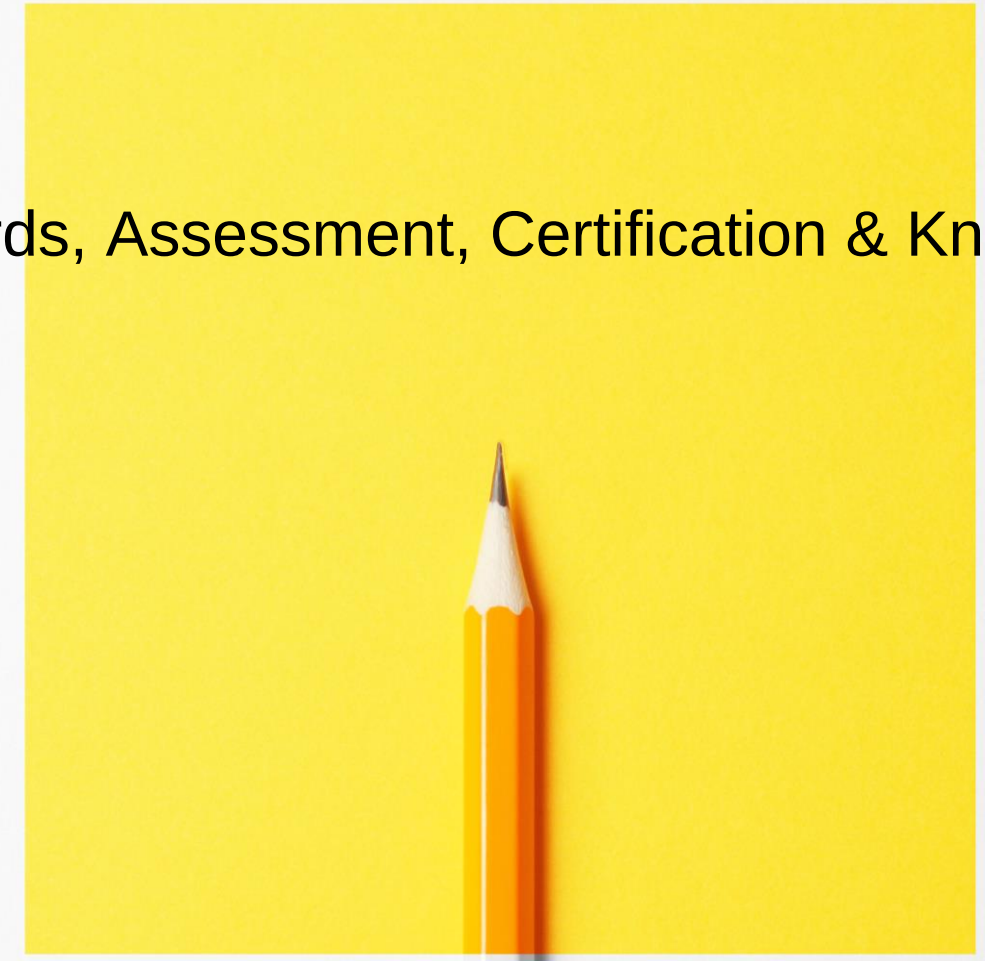


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

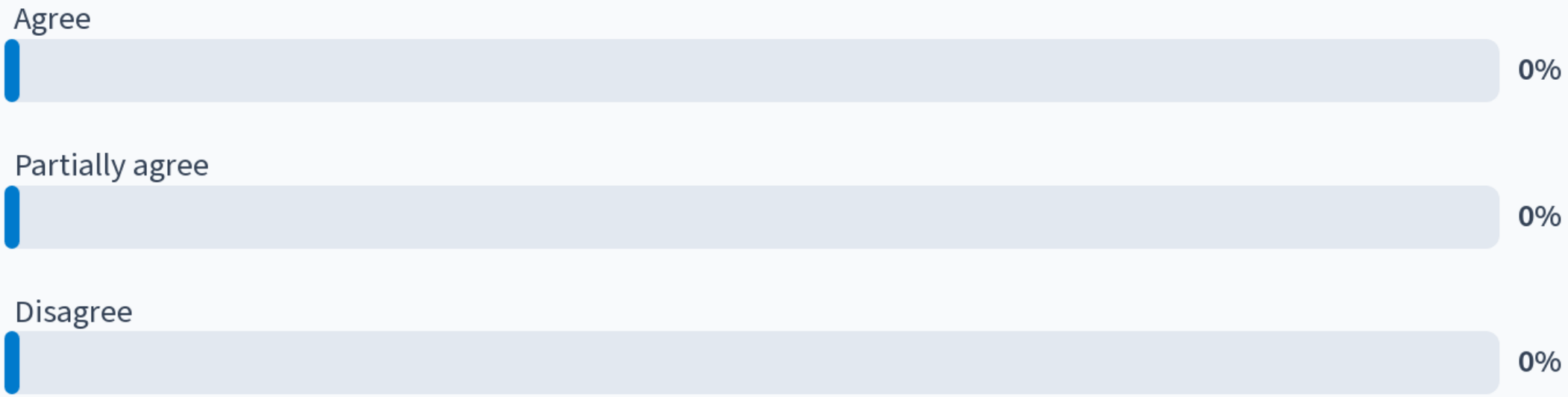
@Technical repository service providers' (storage providers, ARCHIVER etc) portfolio of service offerings should be clear and comparable for client end-users



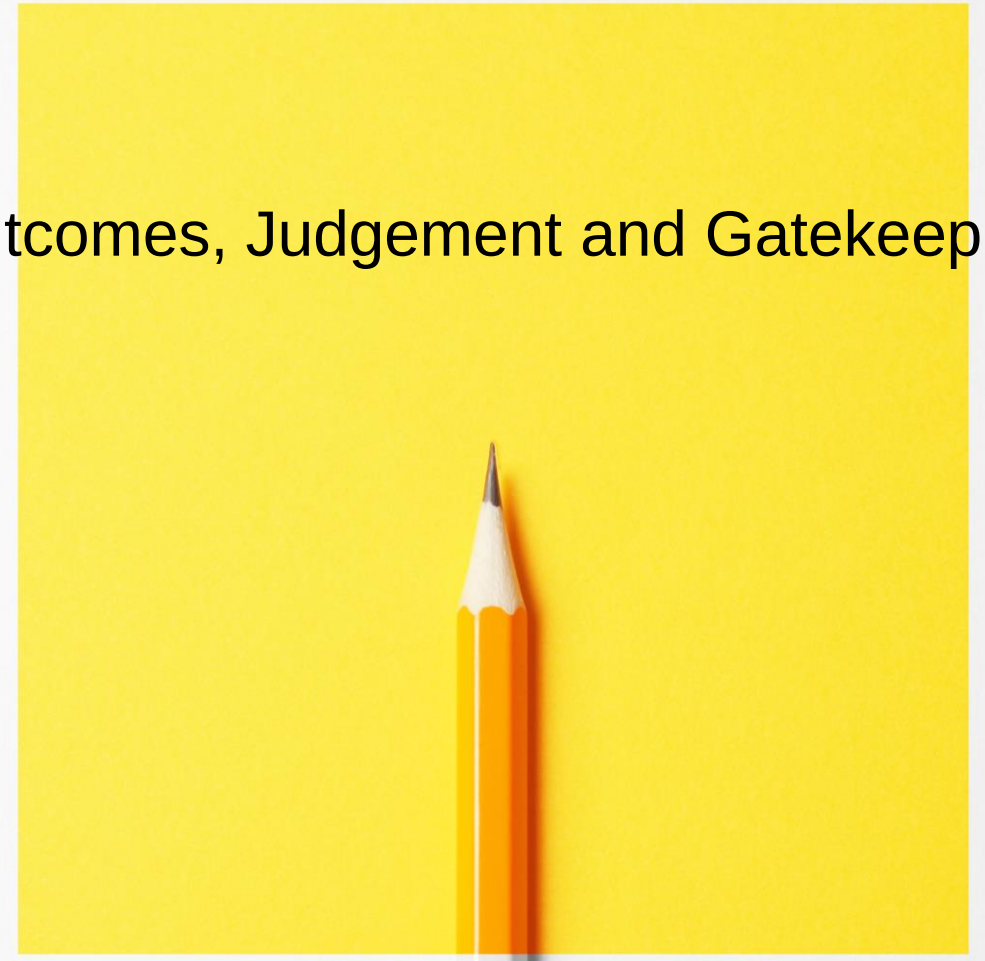
Standards, Assessment, Certification & Knowledge



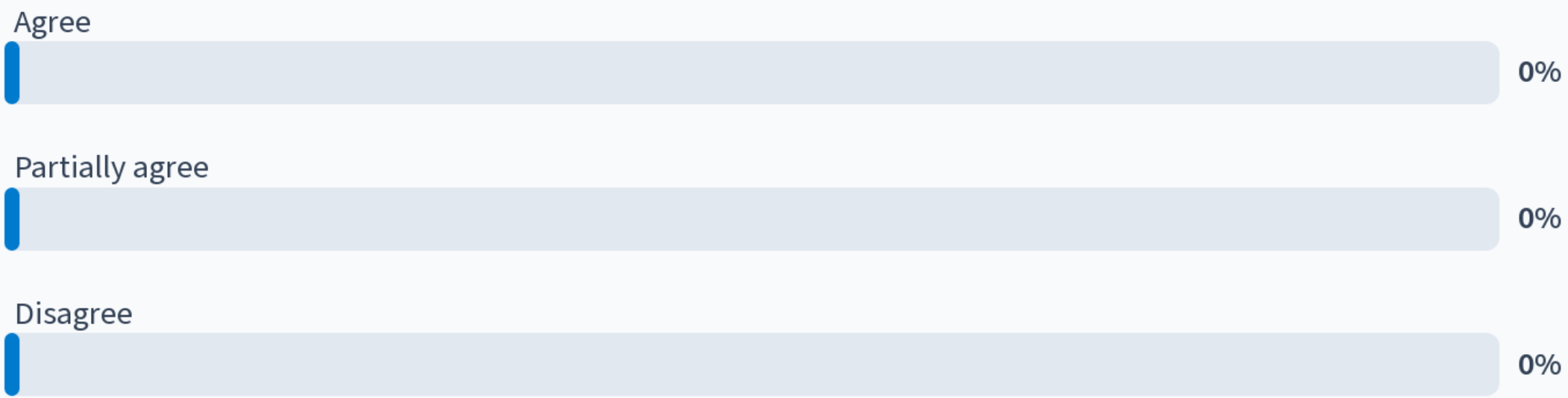
@ Standards and guidance should be developed, coordinated and maintained to provide full lifecycle information on preservation to researchers and preservation practitioners



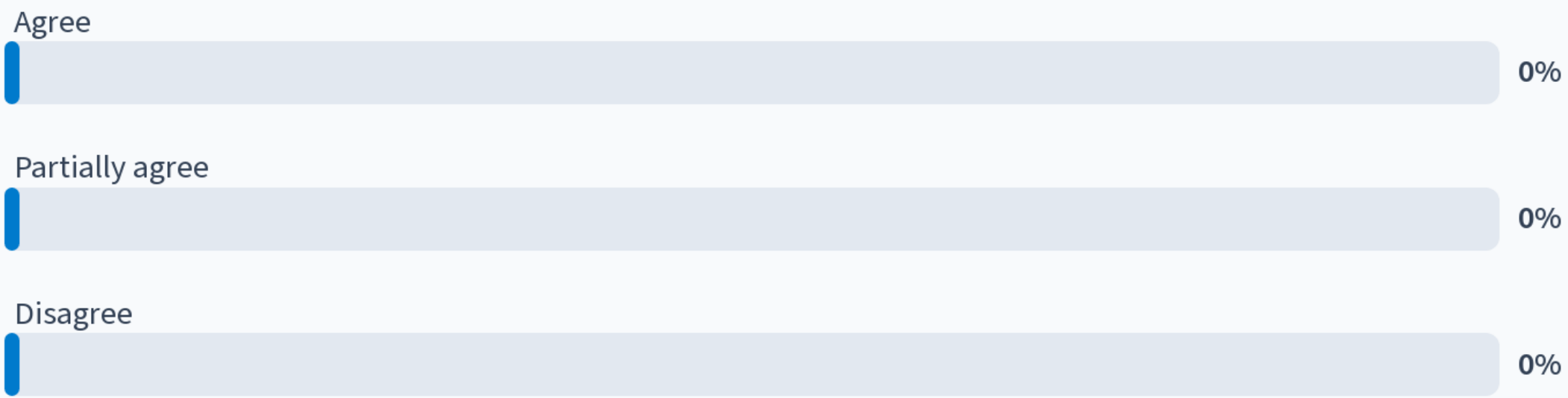
Outcomes, Judgement and Gatekeeping



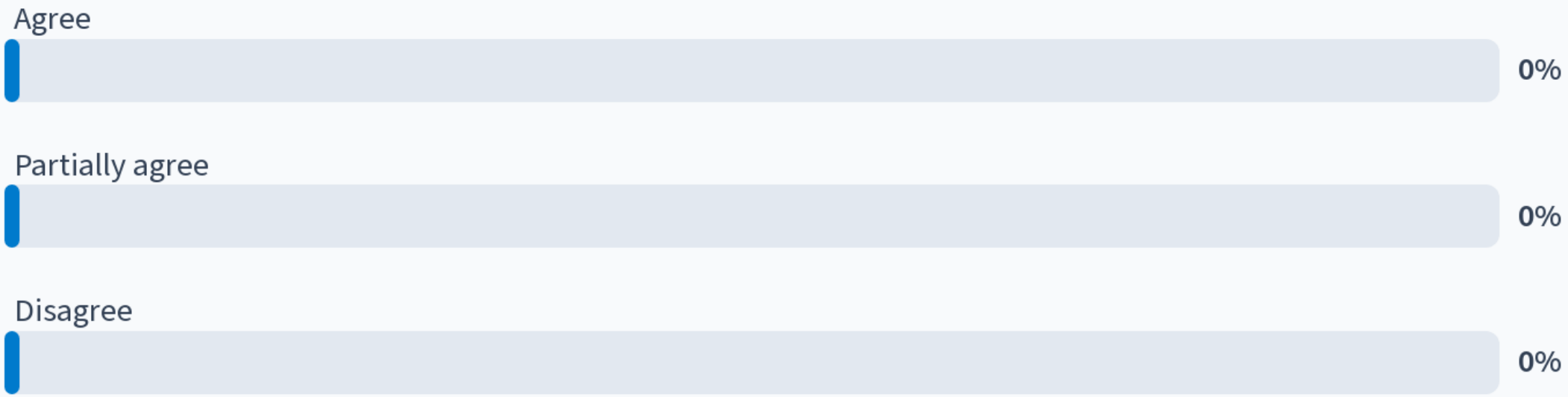
@Digital Objects and the services that enable their FAIRness, deposit, storage, curation, access and preservation should be supported in transparent efforts to use assessment as a route to improvement



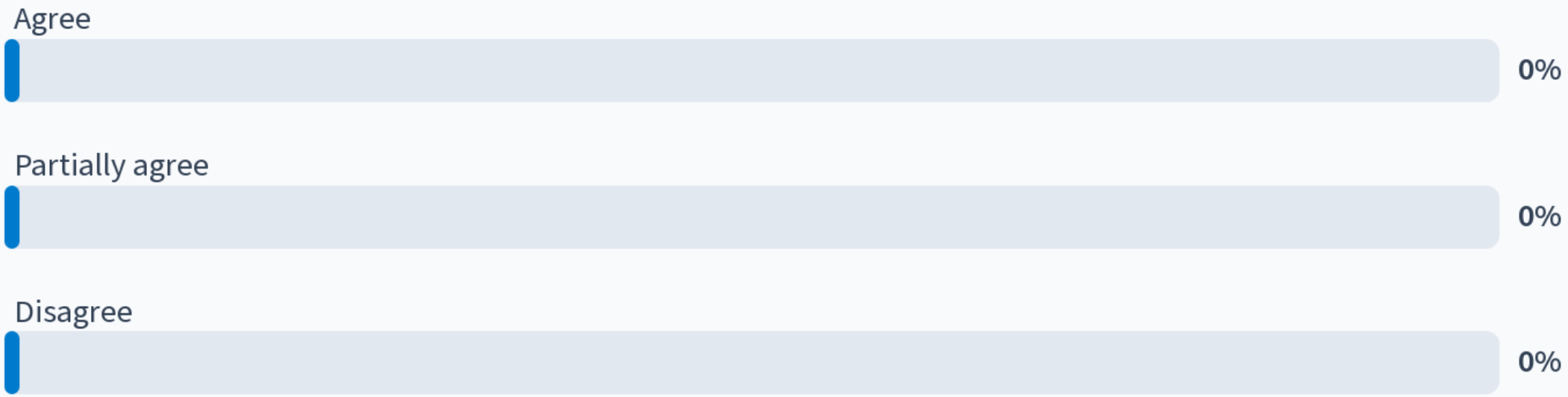
@Efforts by repositories and other data services to share transparent information about their functions, activities and objects should be rewarded by targeted investment towards improvement



@No data service or digital object should be unnecessarily excluded from any part of the EOSC

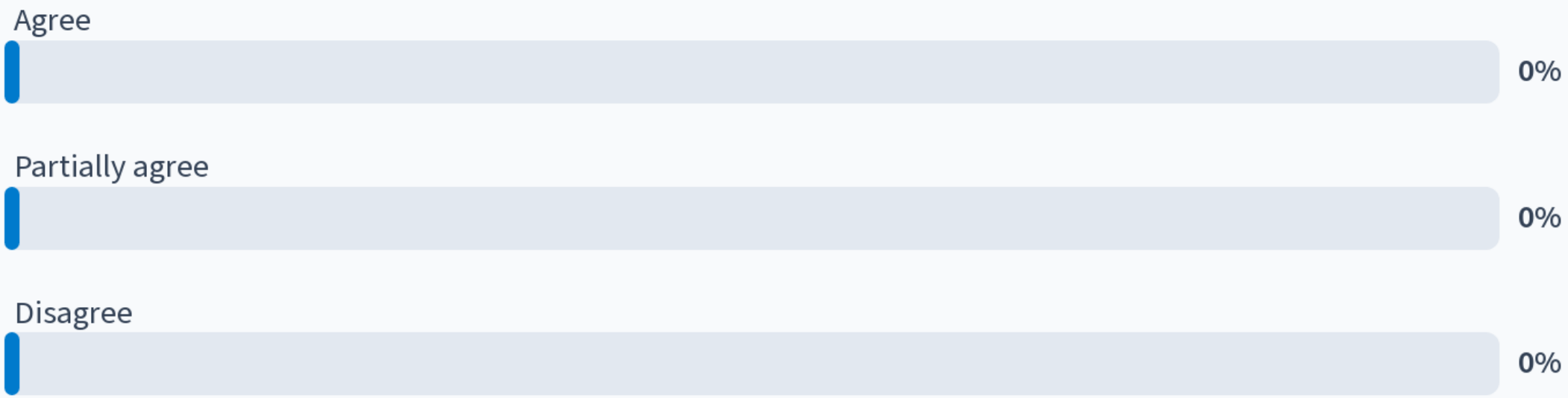


@Retention and reappraisal decisions and timescales, including guaranteed preservation timescales, should be transparent

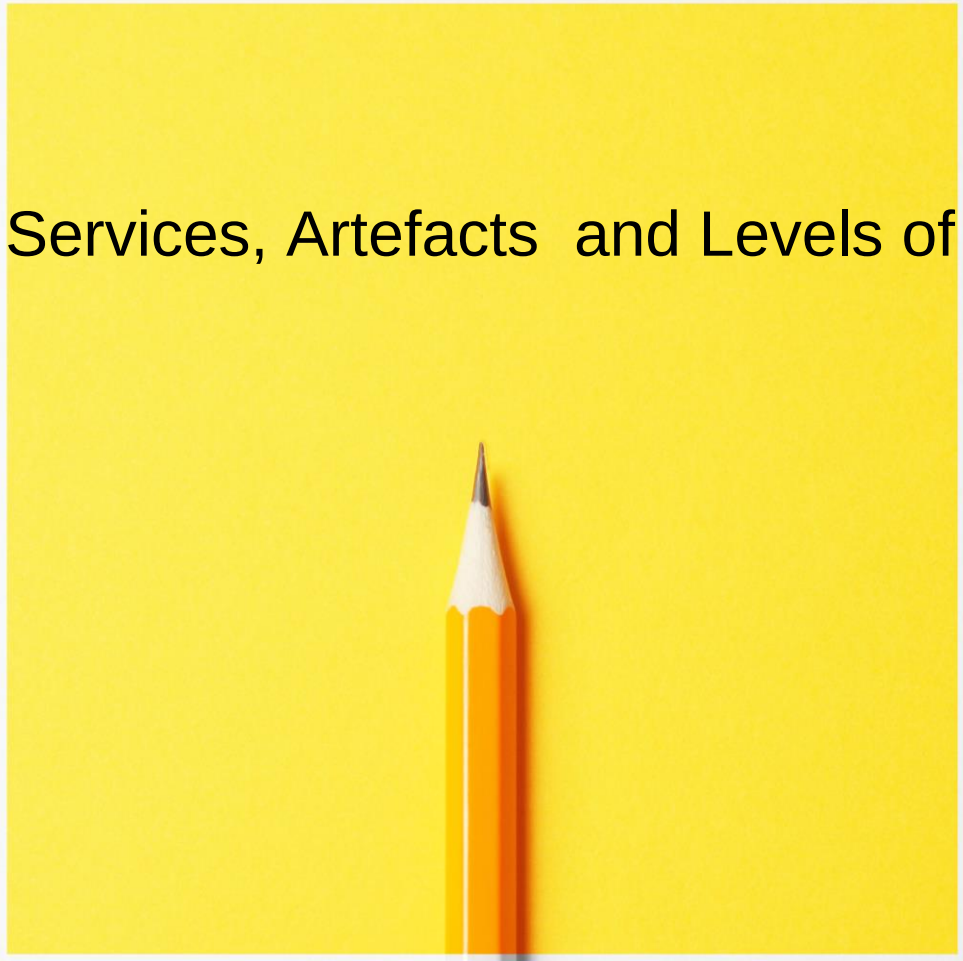


Retention, Appraisal & Re-Appraisal

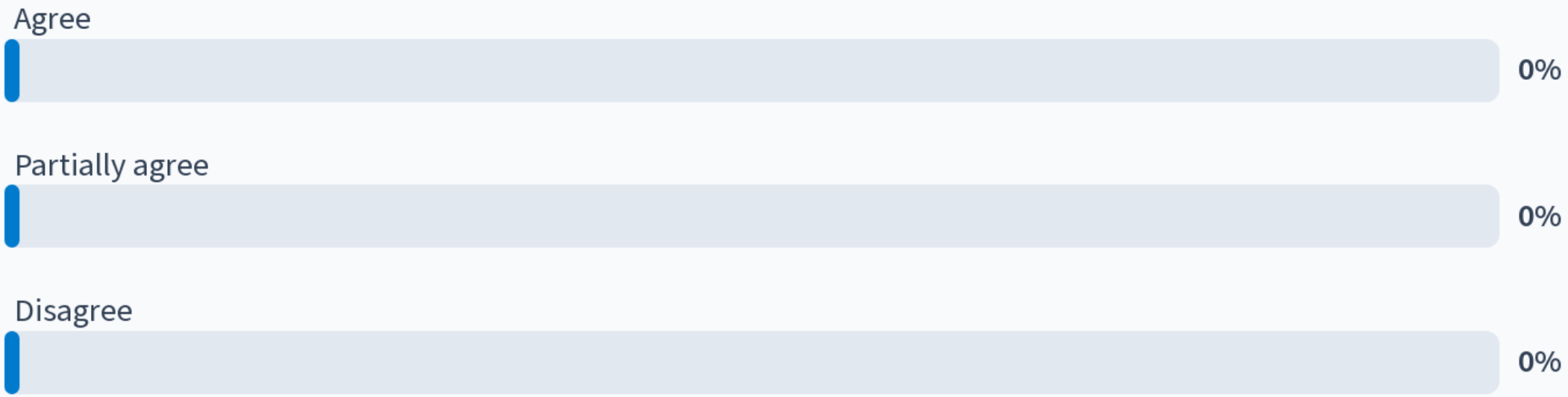
@Retention and reappraisal decisions and timescales, including guaranteed preservation timescales, should be transparent



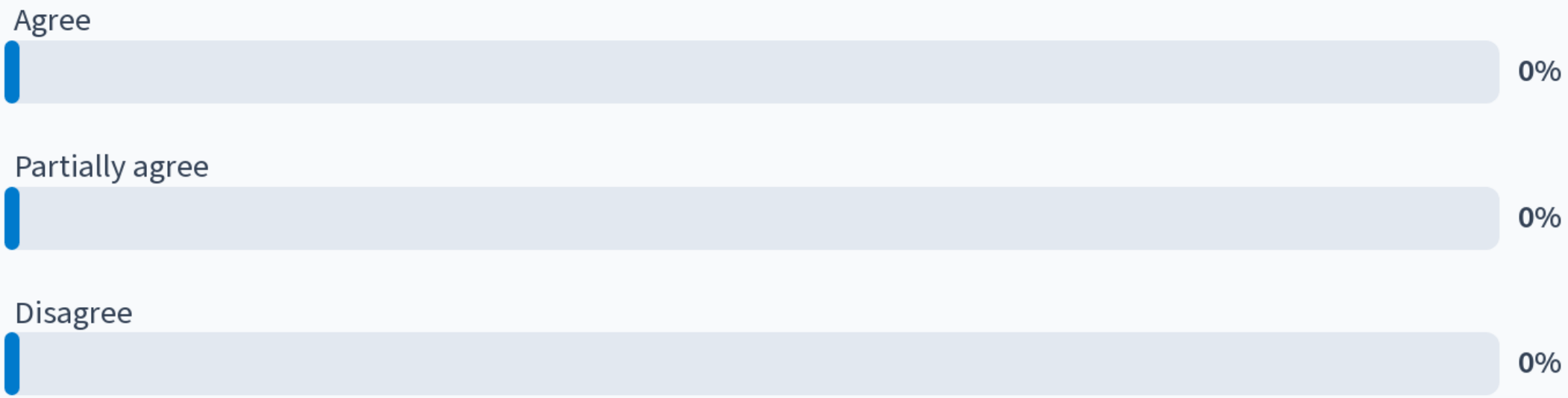
Transparency of Services, Artefacts and Levels of Care for Objects



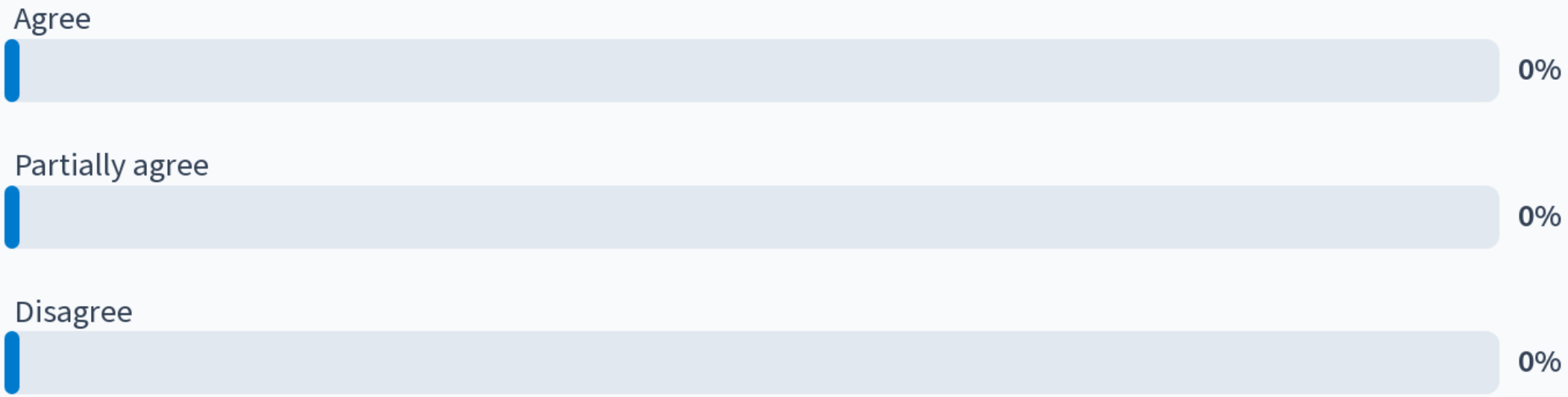
@Each service should be transparent about the levels of service provided



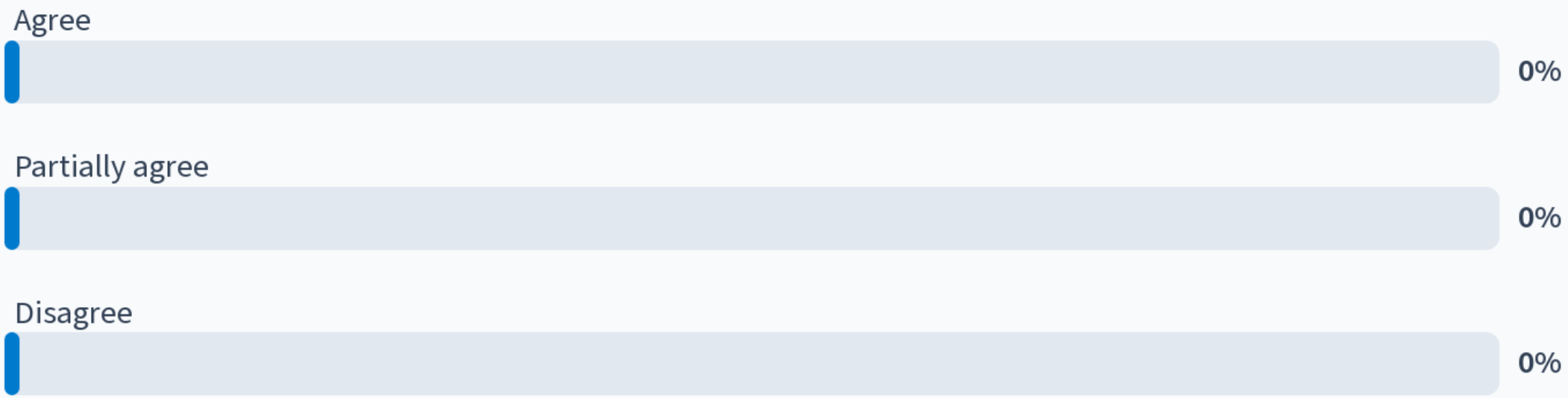
@Each object should have a defined level of care associated with the service(s) that take responsibility for them



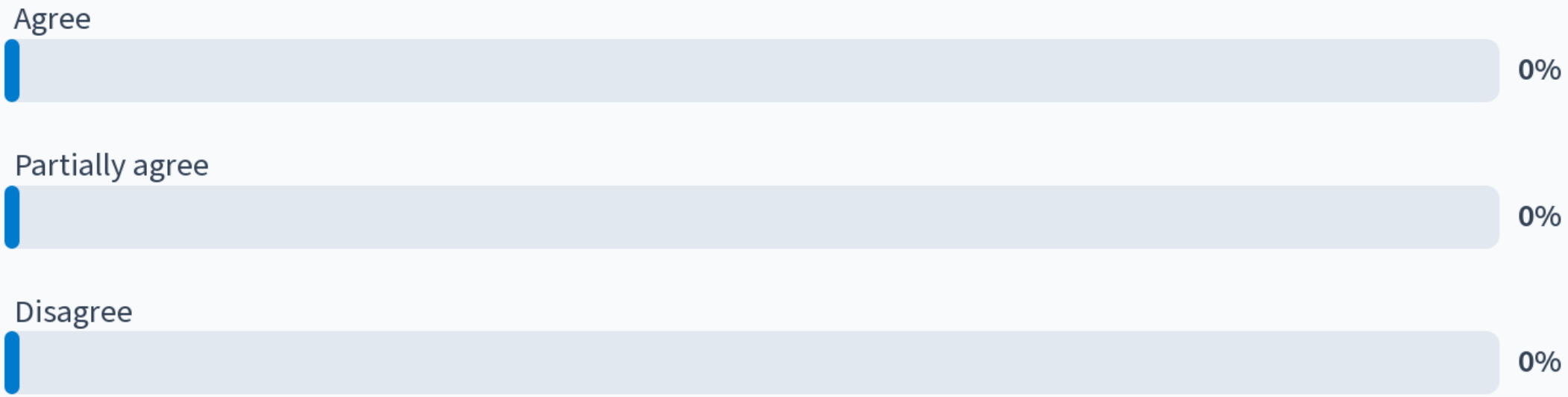
@Living and machine actionable data management plans should form the basis of continuity through the research data lifecycle



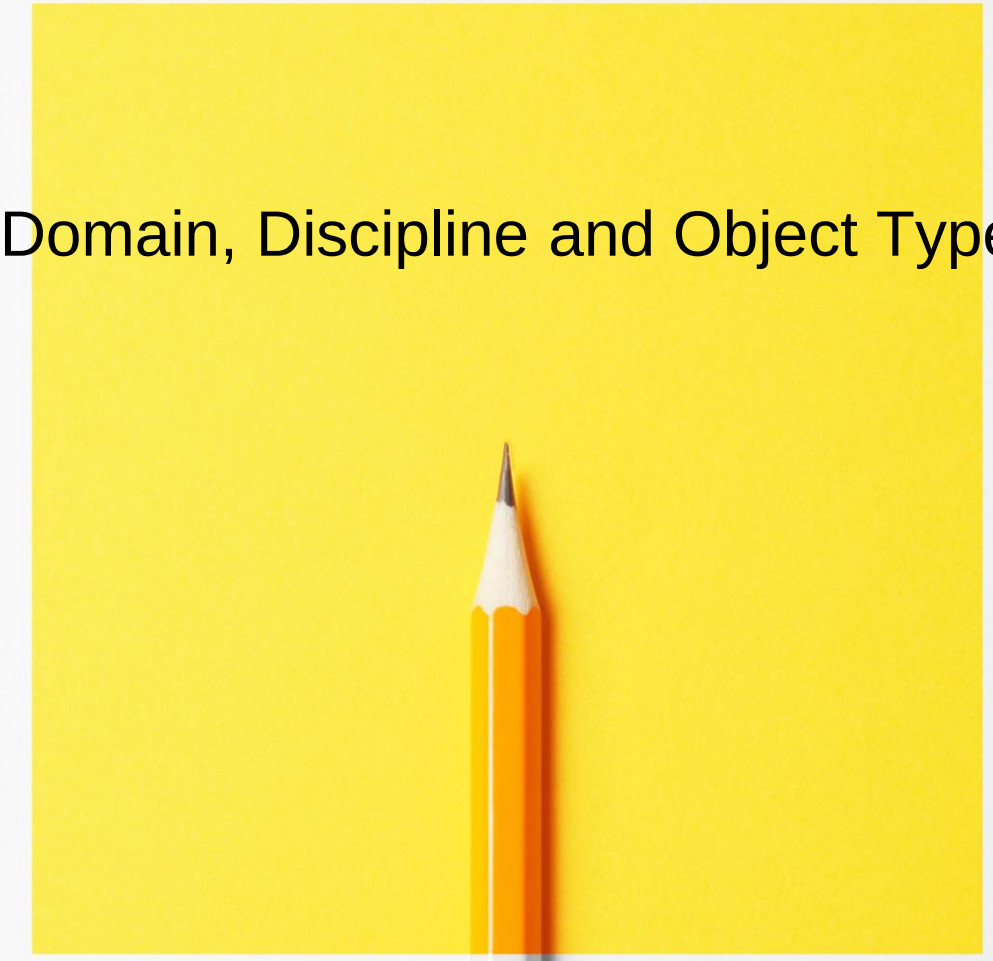
@Registries of repositories and other data services should align service level metadata and supporting information



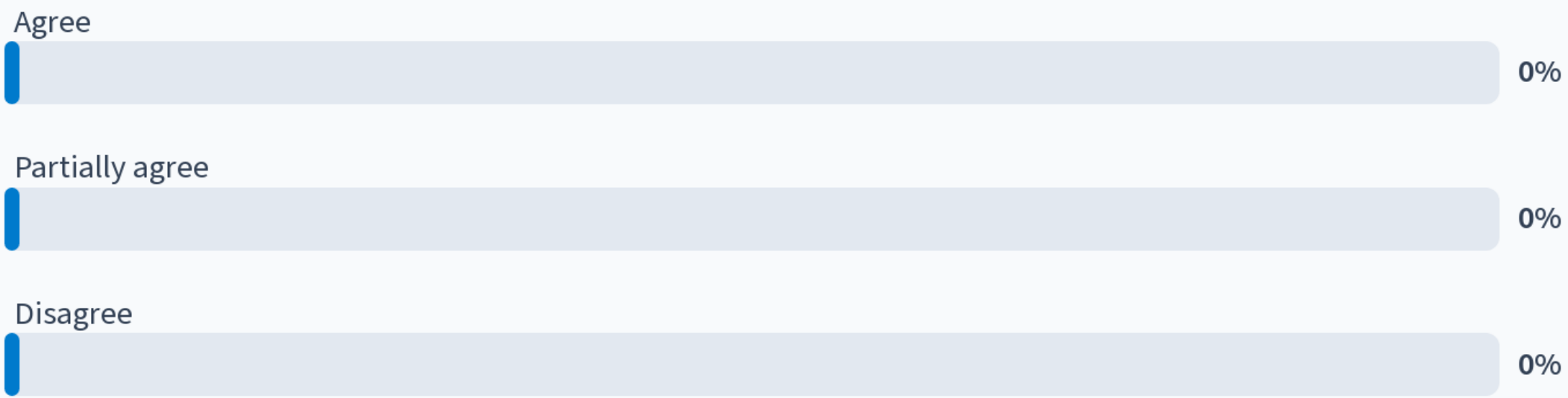
@Registries of digital objects should align metadata relating to retention periods, appraisal periods and levels of curation and preservation



Generic versus Domain, Discipline and Object Type Specific Issues



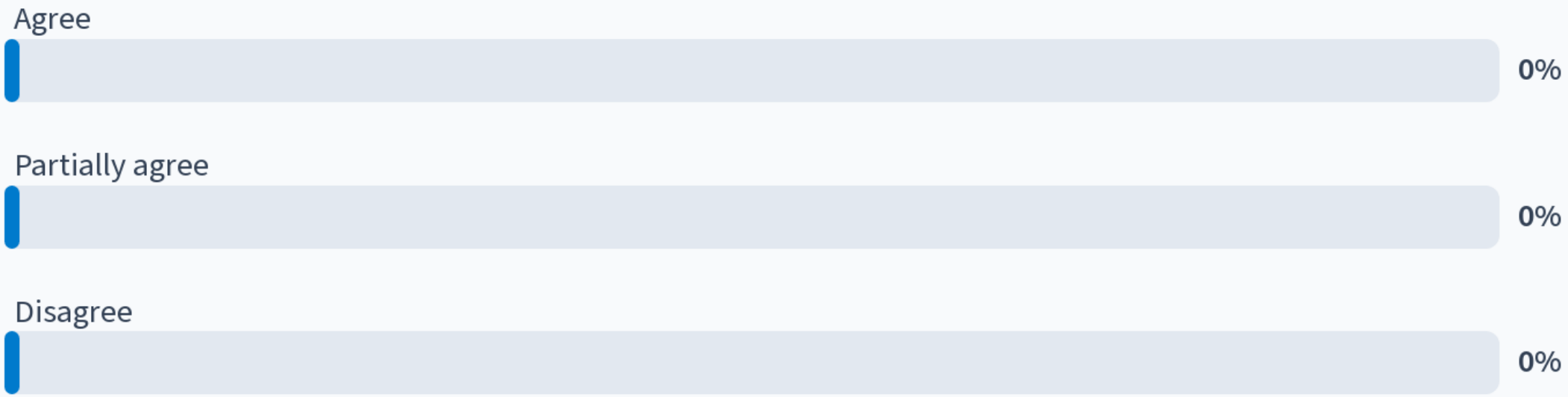
@Addressing the challenges of metadata and interoperability in and across scientific domains and disciplines must be supported in further investment to identify granular needs for specific types of digital objects and disciplines



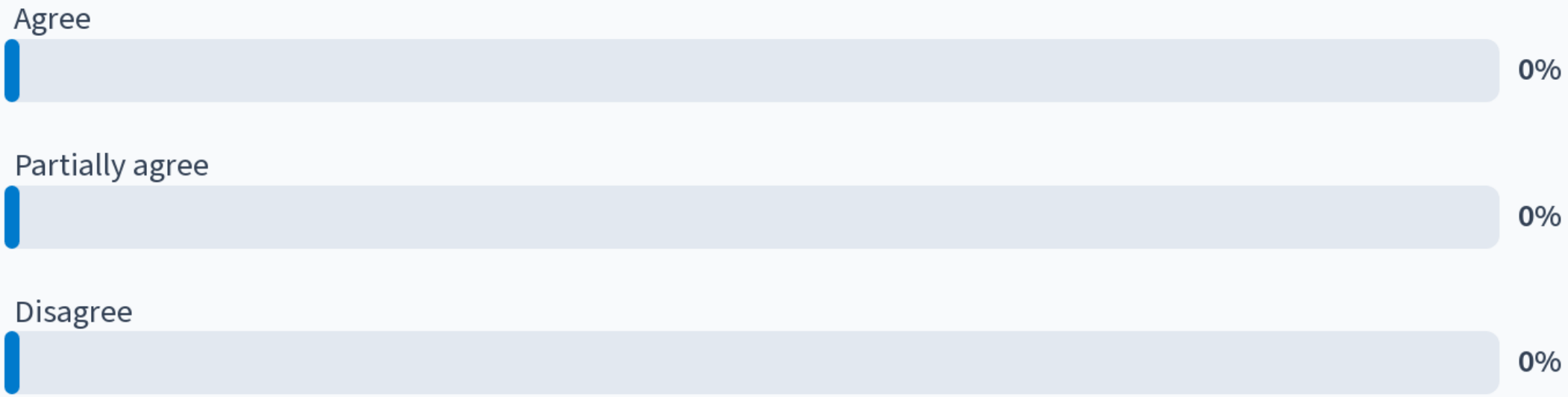
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Roles and Responsibilities

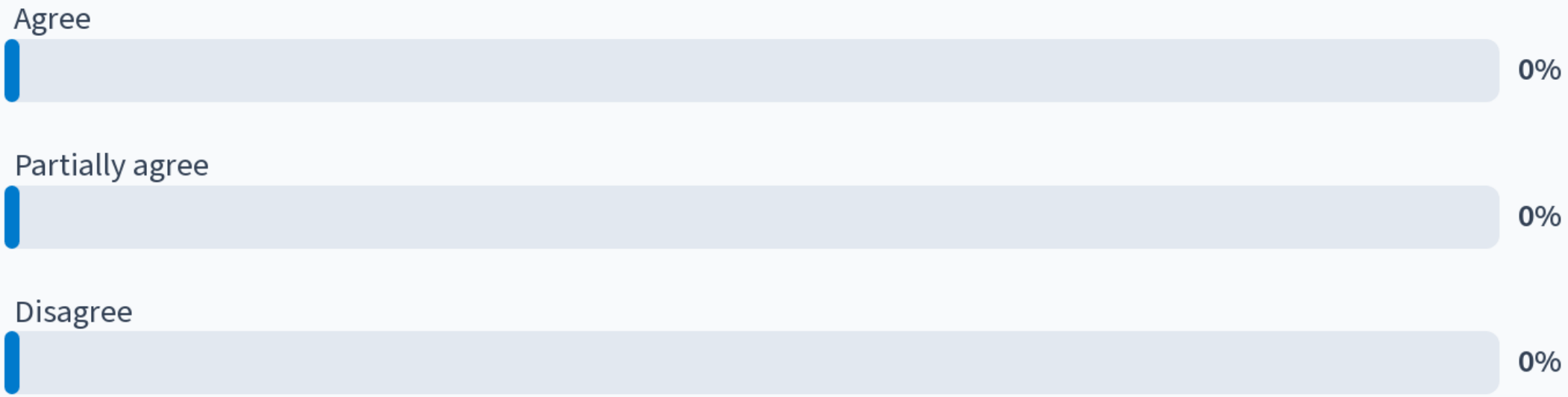
@Where responsibility is distributed, accountability should remain clear, including accountability for (meta)data loss or destruction



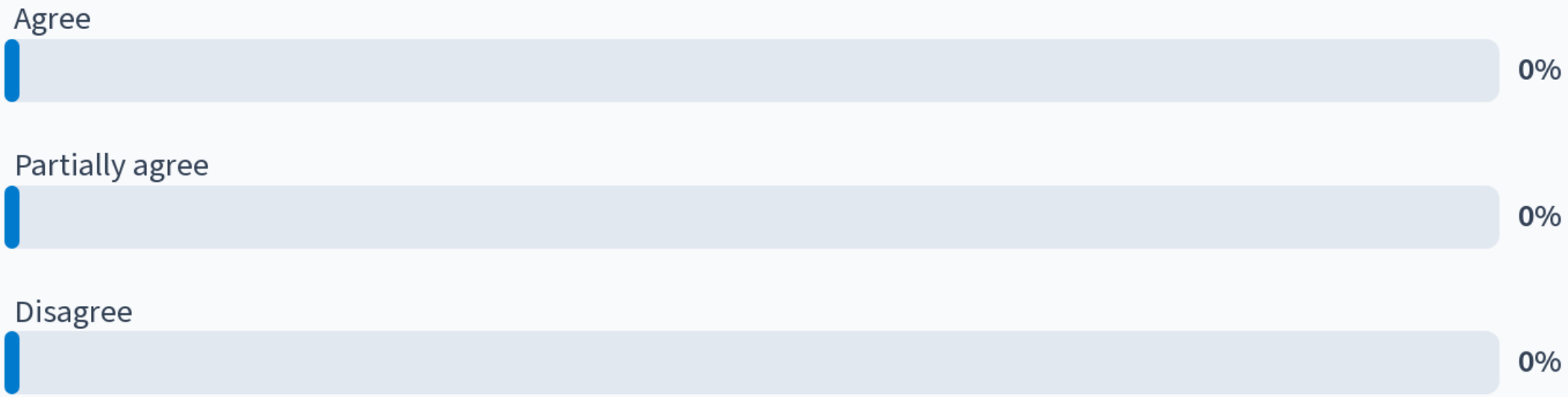
@Digital object management outcomes, including preservation, should be integrated into a roles and responsibilities framework that integrates all actors and actions



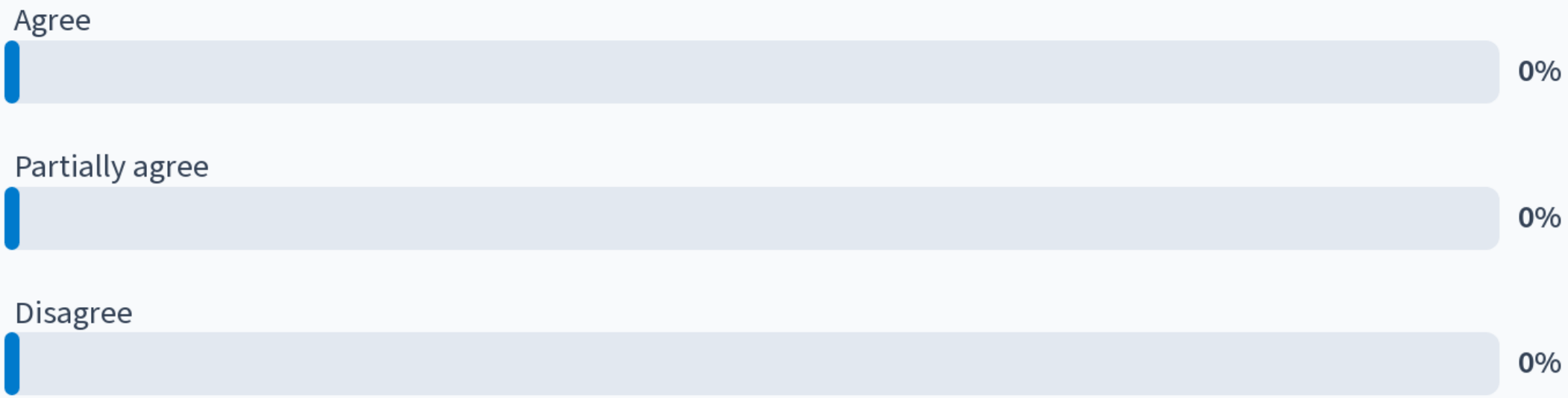
@The roles and responsibilities framework should be aligned with clear process models that meet the needs of different stakeholder communities



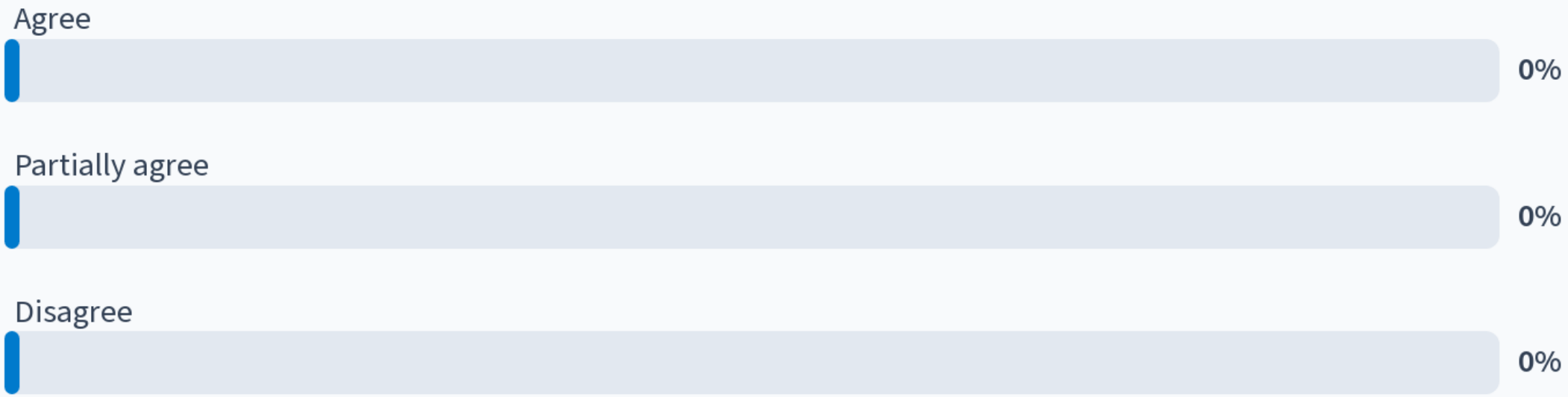
@Defined roles should be in place to take responsibility from the point of conception to ensure that preservation actions are considered throughout the life cycles



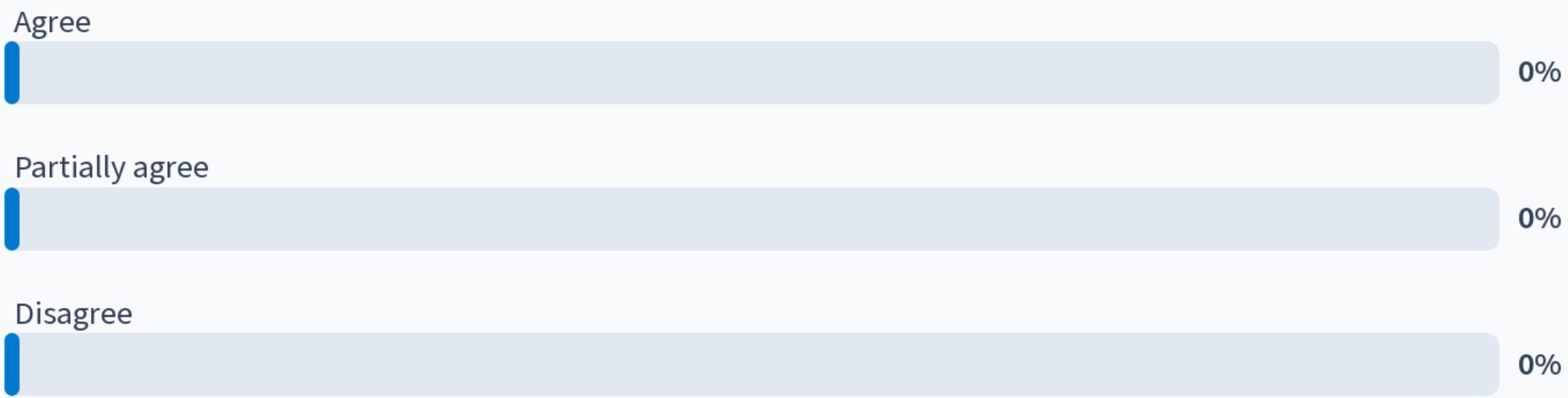
@Risk analysis approaches should be used to identify when in the lifecycle it is appropriate to take preservation actions



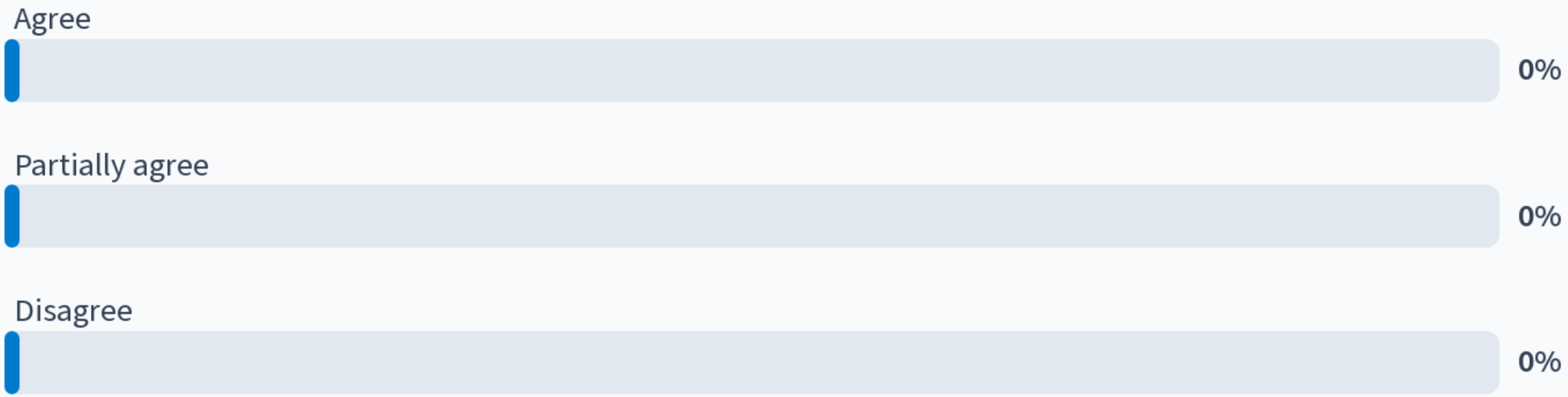
@Different roles should use a living data management plan as a key artefact for periodic audit, review and revision



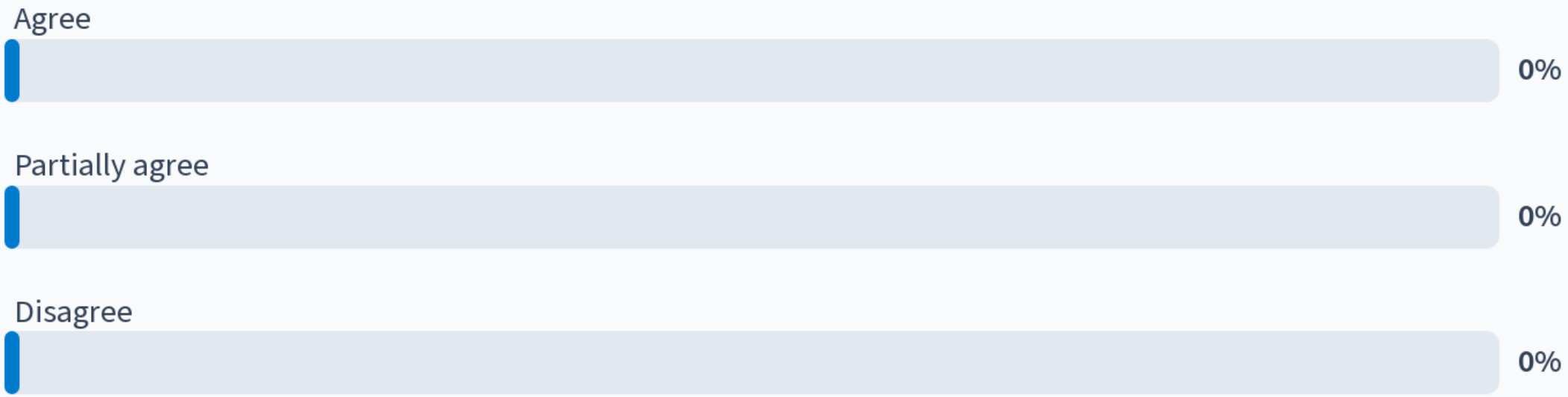
@Policy makers must make the development and implementation of digital preservation explicit in policy applicable to all stakeholders and across the lifecycle. They are accountable for periodic review and revision of policy



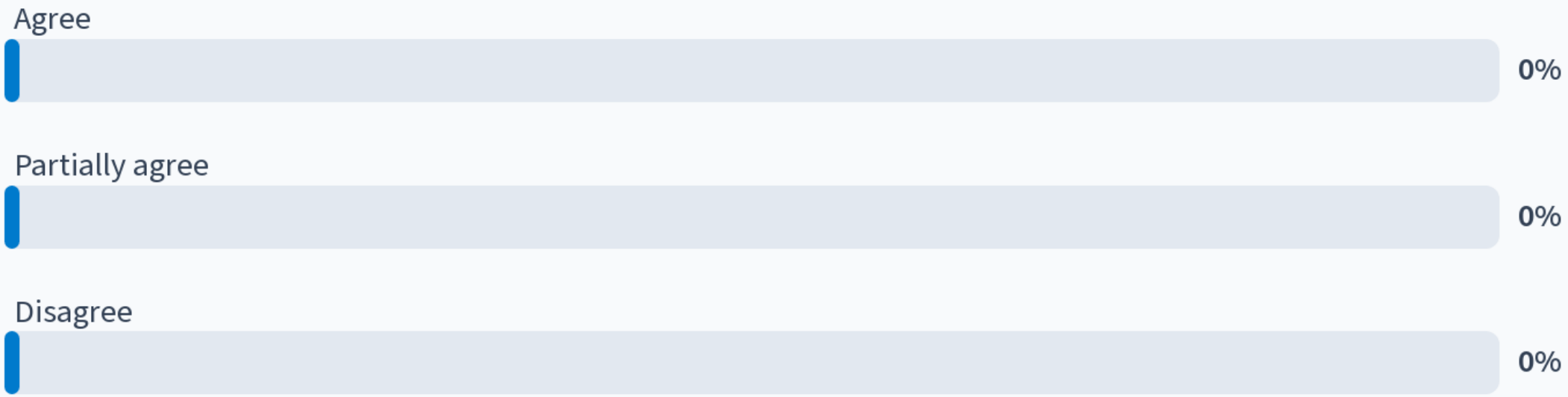
@Executives must adopt preservation policy into their operational and strategic planning



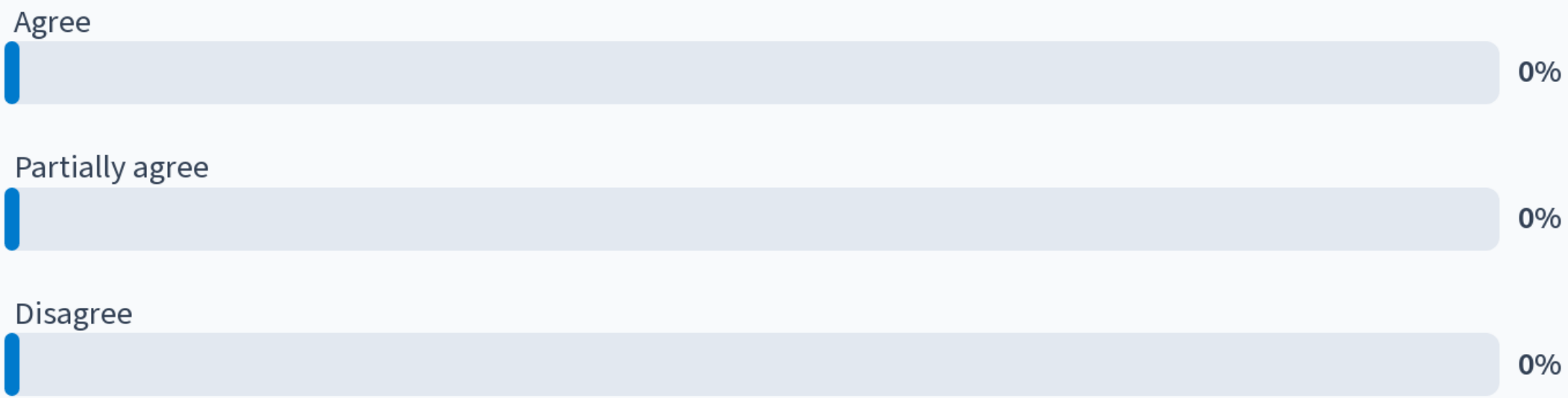
@Managers must integrate preservation planning into operational management including staffing, funding, service development and procurement



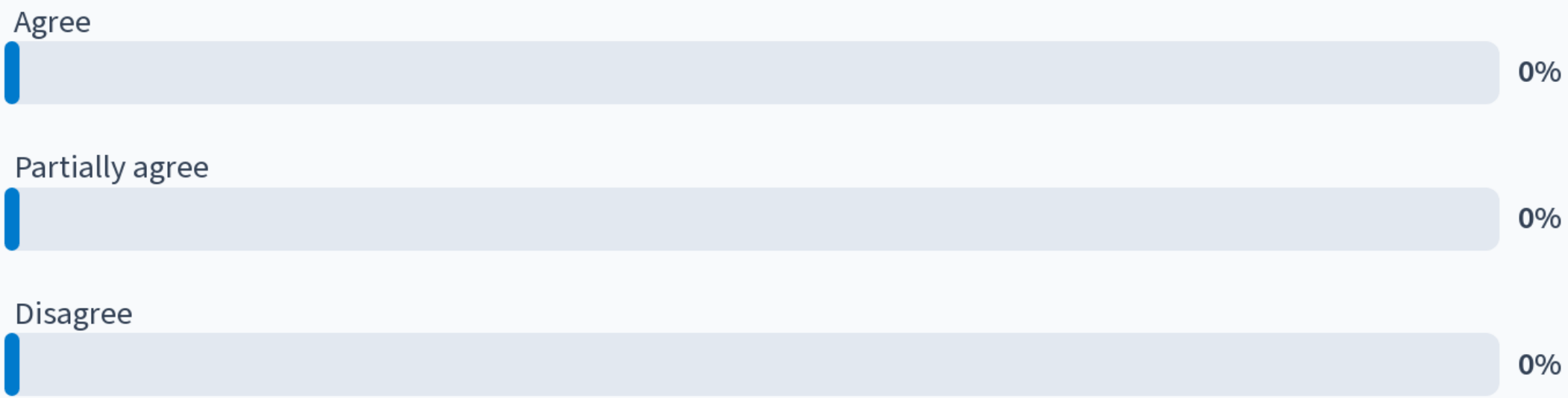
@Practitioners must provide guidance, community and technical monitoring and, where necessary, take preservation actions to ensure optimal preservation outcomes



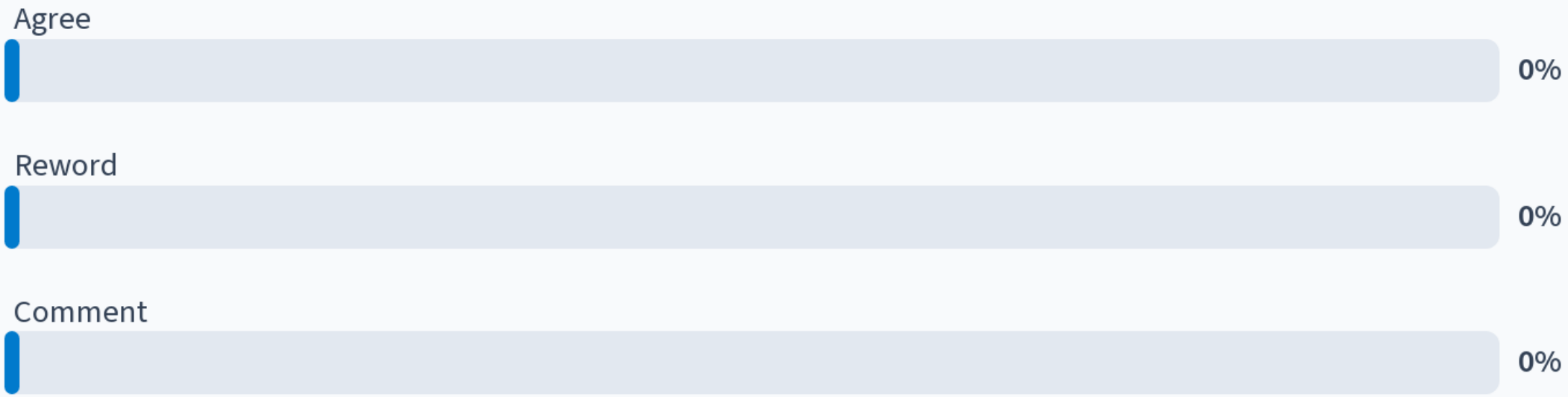
@Data and metadata creators, collectors and reusers, including researchers, should develop the knowledge and skills at a general level and within their own disciplinary and domain area of expertise, so that their actions are preservation-aware



@All of the preservation-specific and supporting research data management roles across the data lifecycle require sustained training based on a rich knowledge base of preservation information



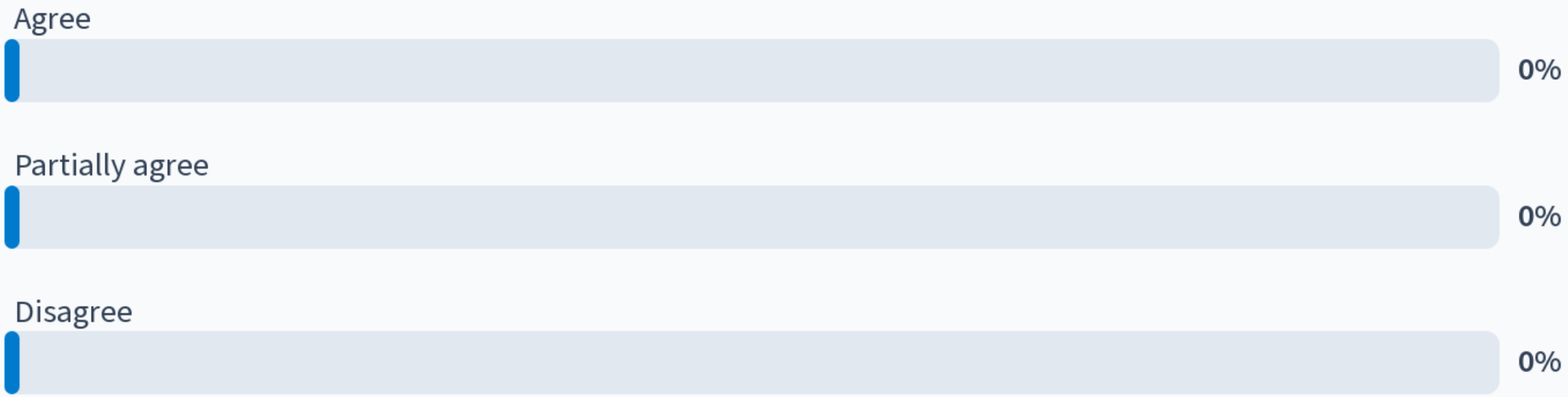
@Clear responsibilities must be in place for developing standards and guidance, for communication and for training



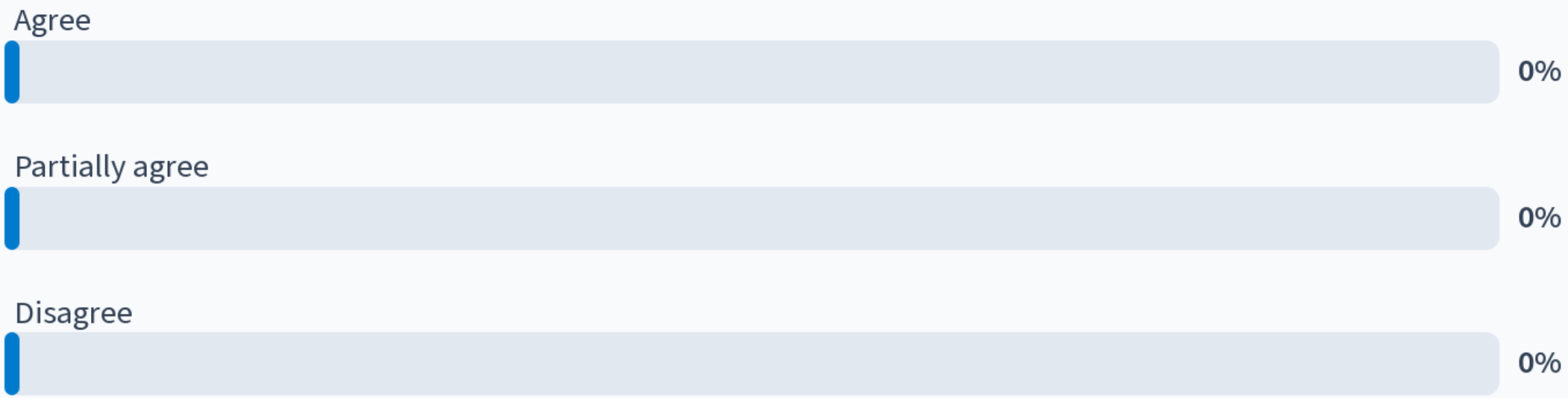
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Finance & Funders

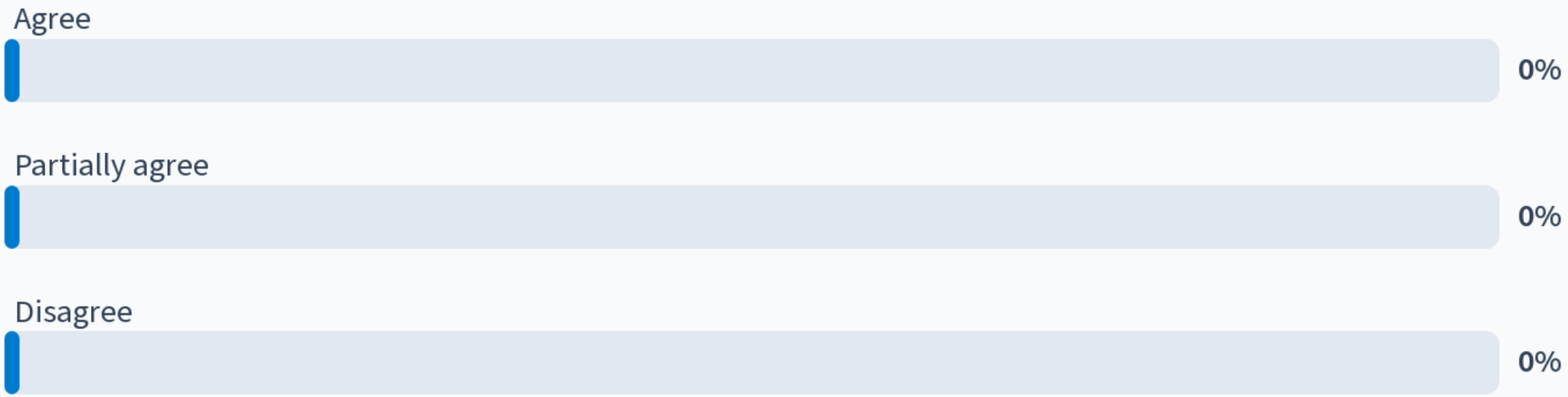
@Further research and analysis is necessary to support business planning based on qualitative and quantitative risks and benefits



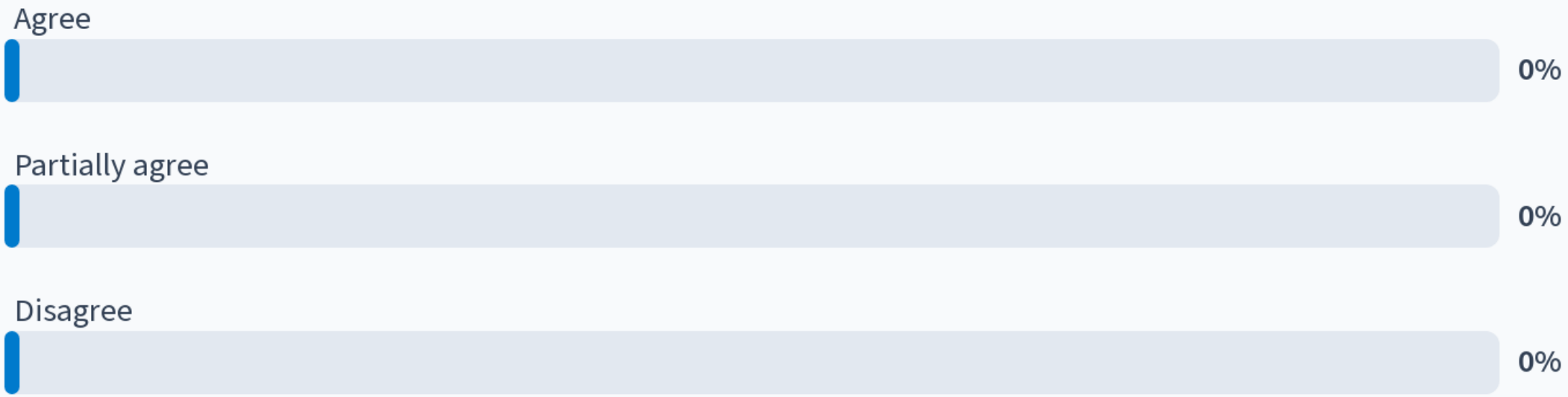
@Funders should clarify to all grant holders that the FAIR Principles and the potential need for preservation are full lifecycle issues



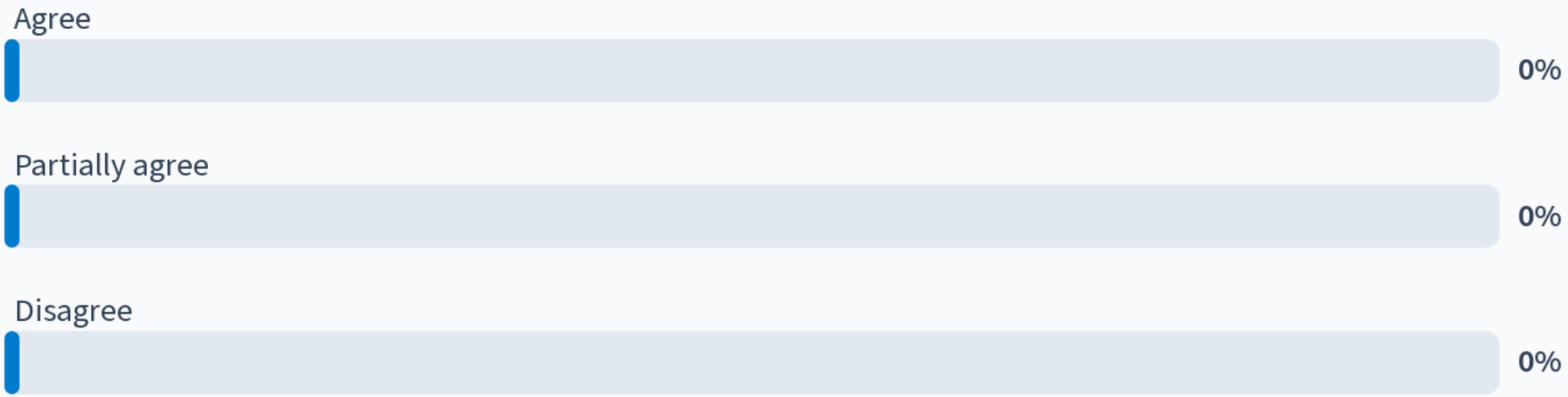
@Funders should integrate into their calls the costs required to meet the needs for compliance with the FAIR principles and long term preservation



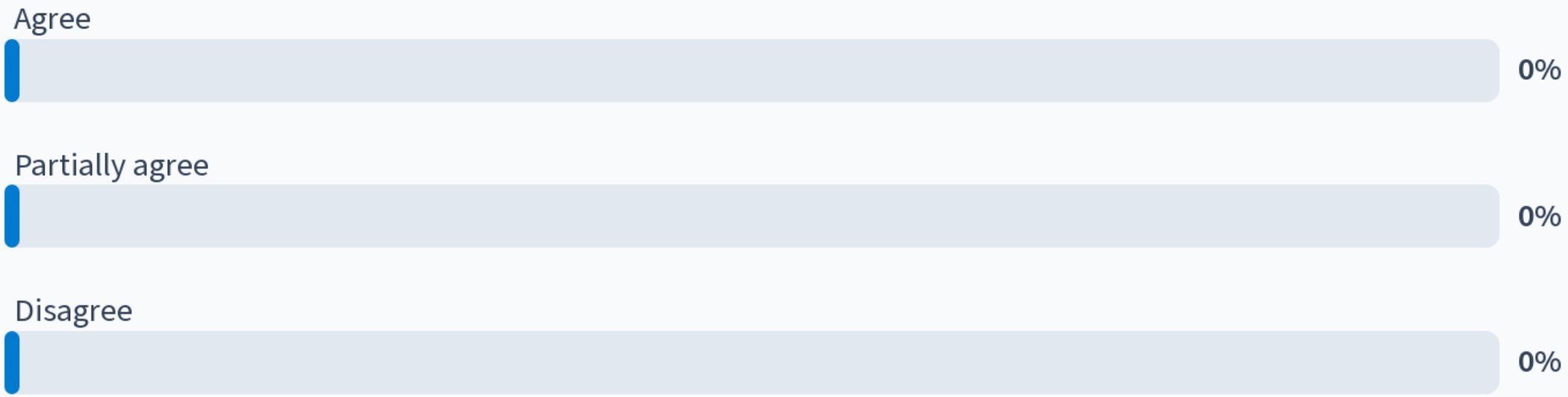
@Audit pathways are essential for all research outputs



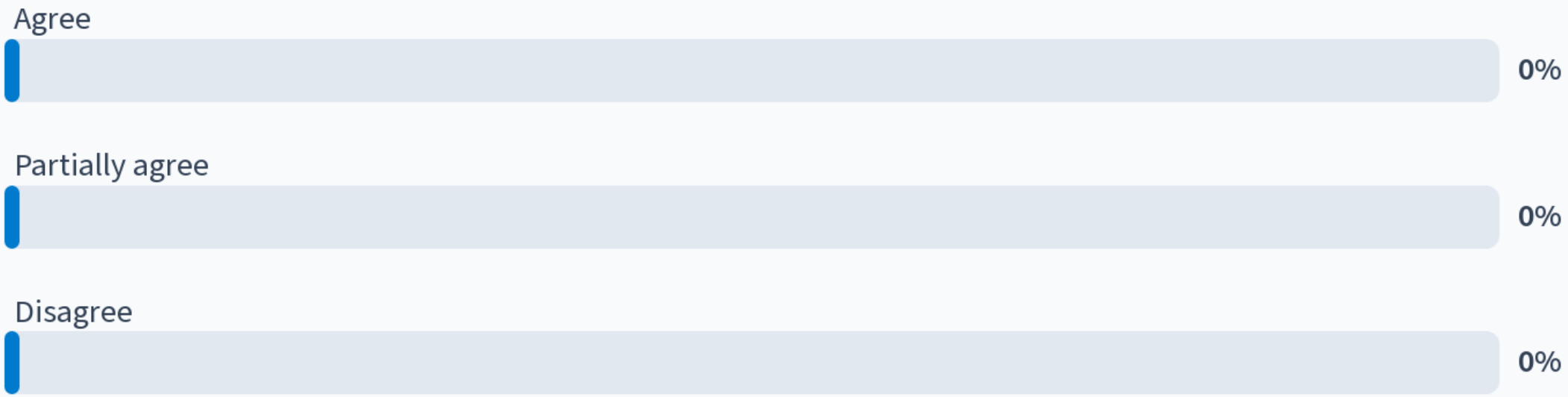
@Data management plans should define obligations and support accountability



@Once identified, critically endangered content that retains value requires investment

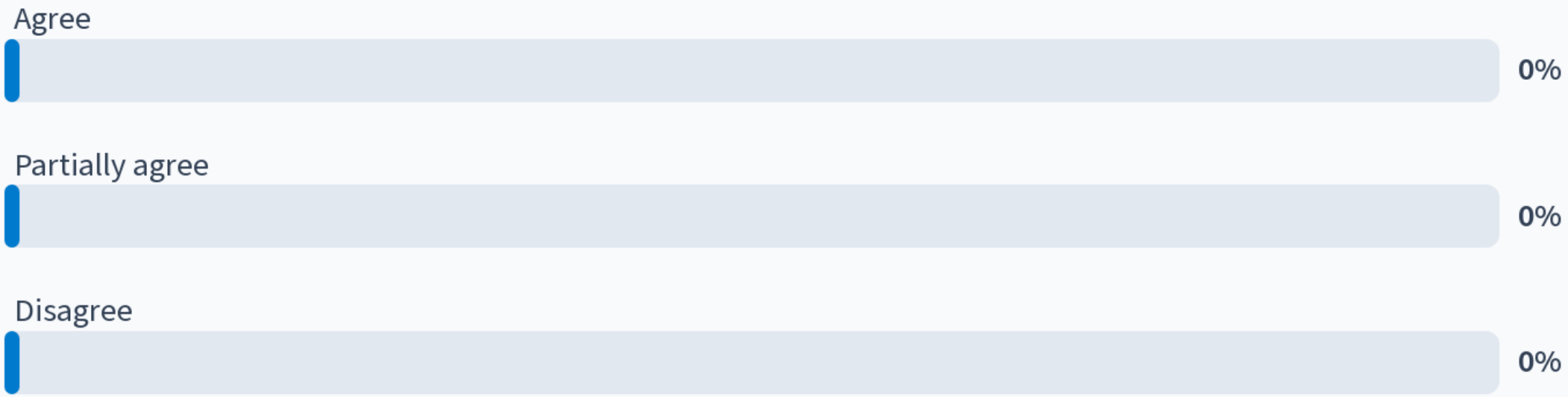


@Full lifecycle costs should be assessed, including the unique costs of preservation using methodologies that include the potential costs of inaction



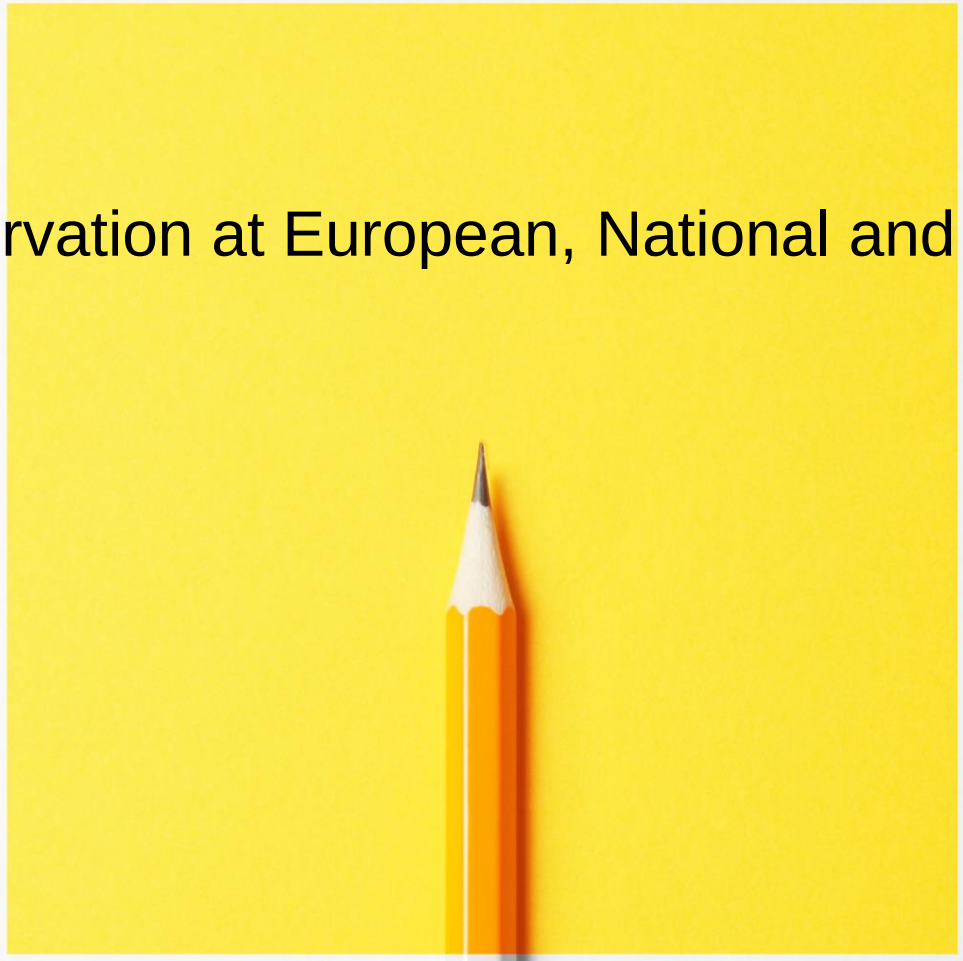
Back

@Identify and support staffing costs for preservation specific roles and responsibilities

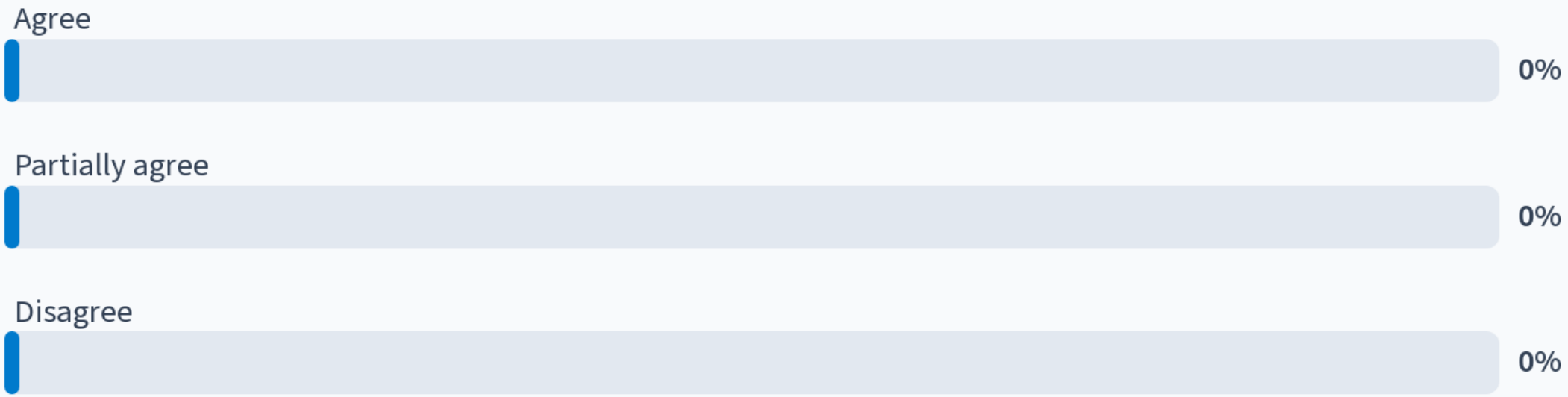


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Addressing Preservation at European, National and Institutional Level

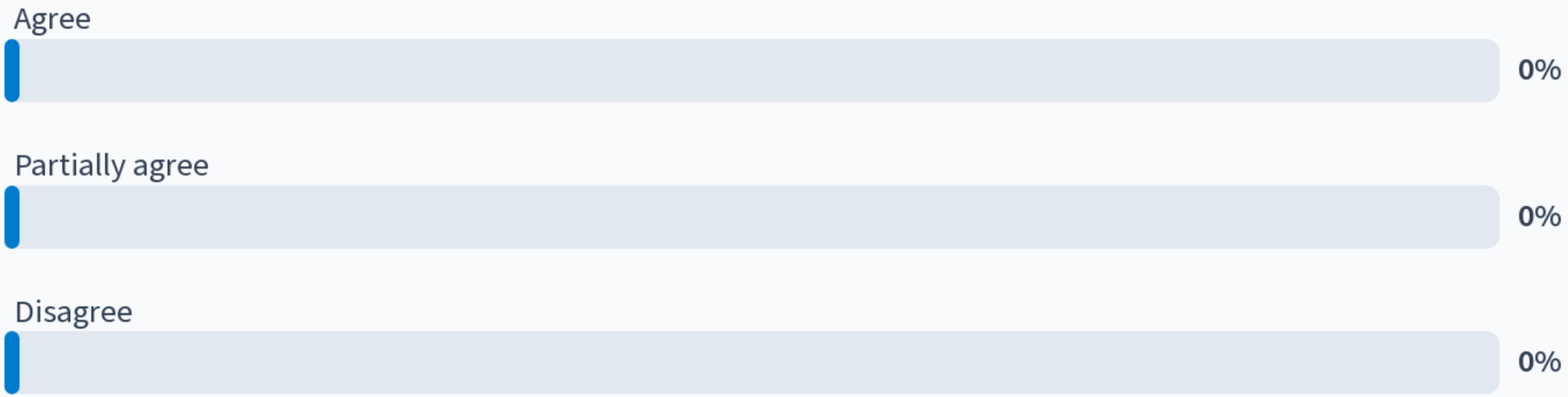


@Greater clarity is needed in the EOSC vision and for the application of preservation activity across the lifecycle and at EU, National and institutional level



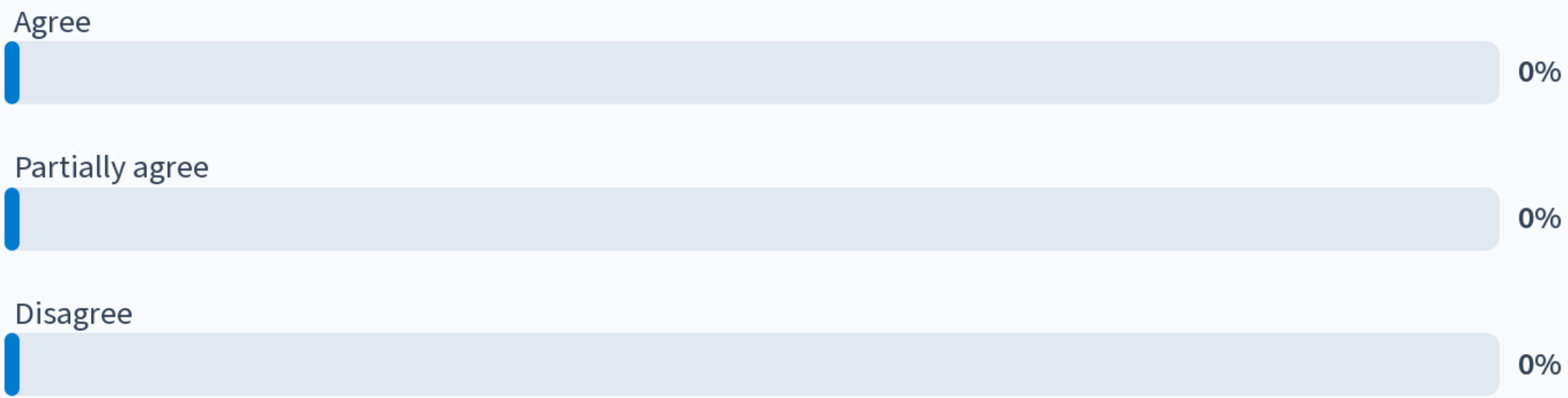
Recommendations at European Level

@Digital preservation risks and opportunities must be made explicit in the EOSC vision and monitored and addressed as the EOSC evolves

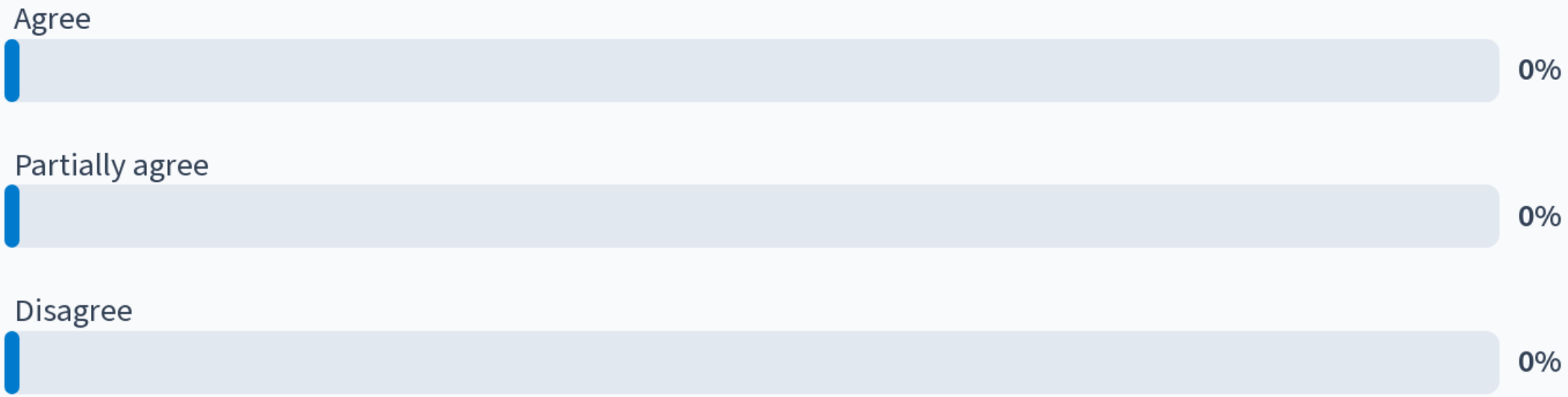


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

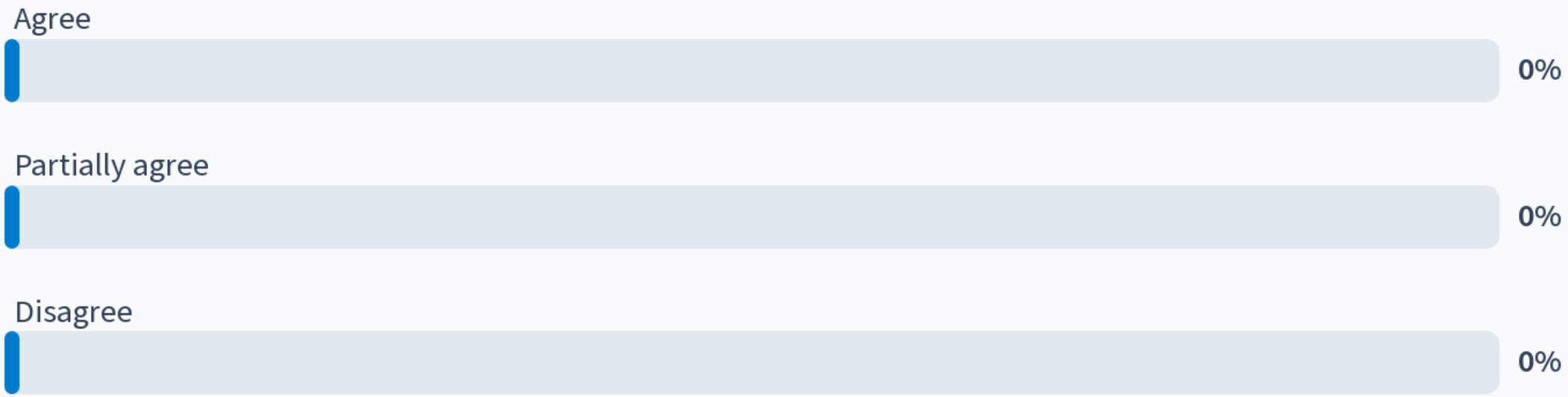
@EOSC cannot rely on a single generic canon of 'digital preservation practice'. Instead, workflows should leverage large scale infrastructures while remaining faithful to discipline-specific requirements.



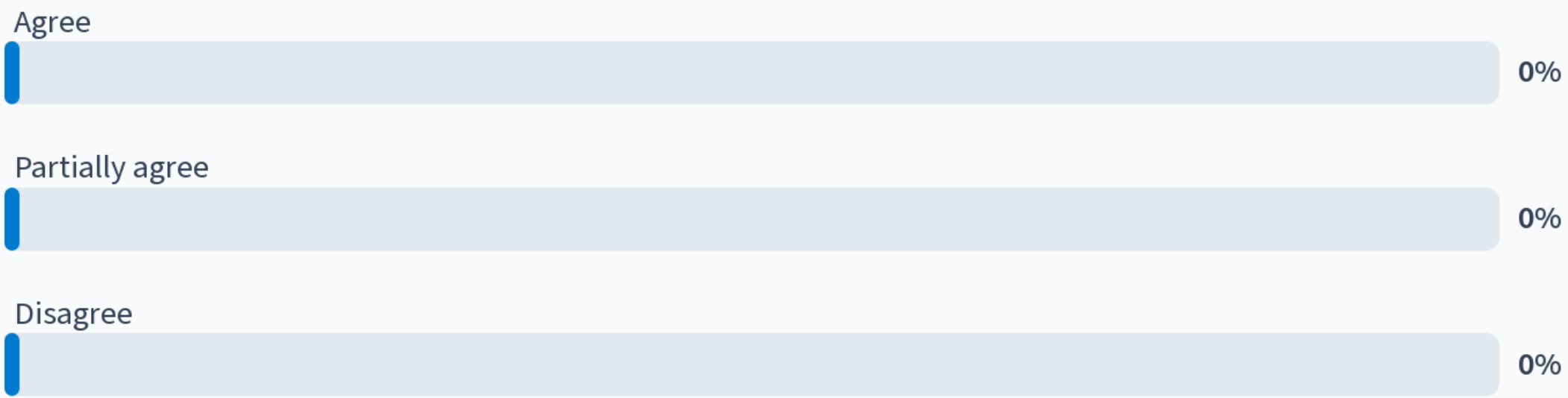
@Roles, responsibilities, and accountabilities for preservation in EOSC should be clarified



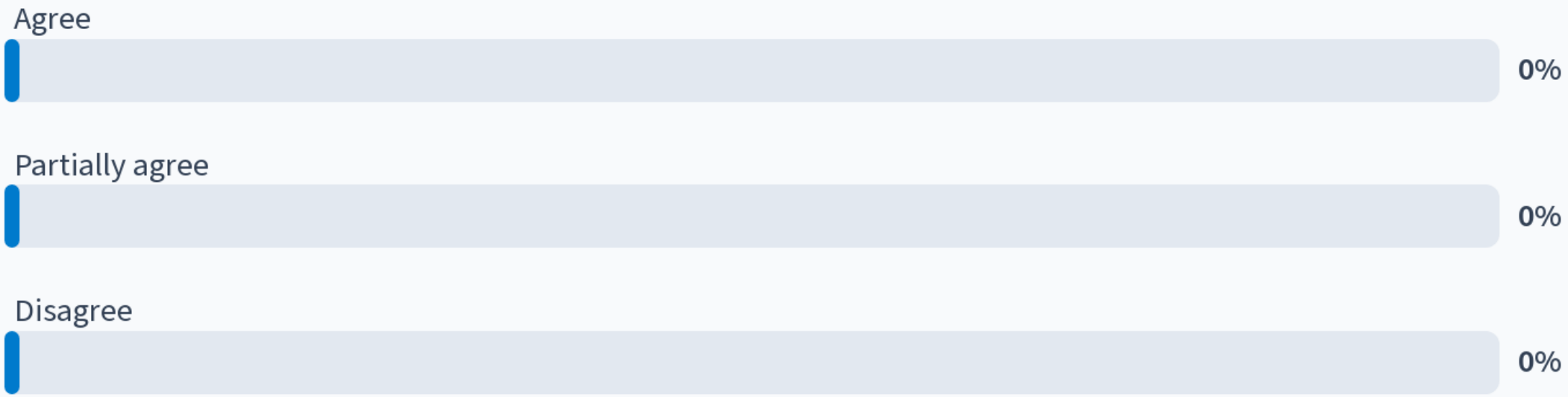
@EOSC should establish a workplan for policy development and implementation within EOSC services and partners ...



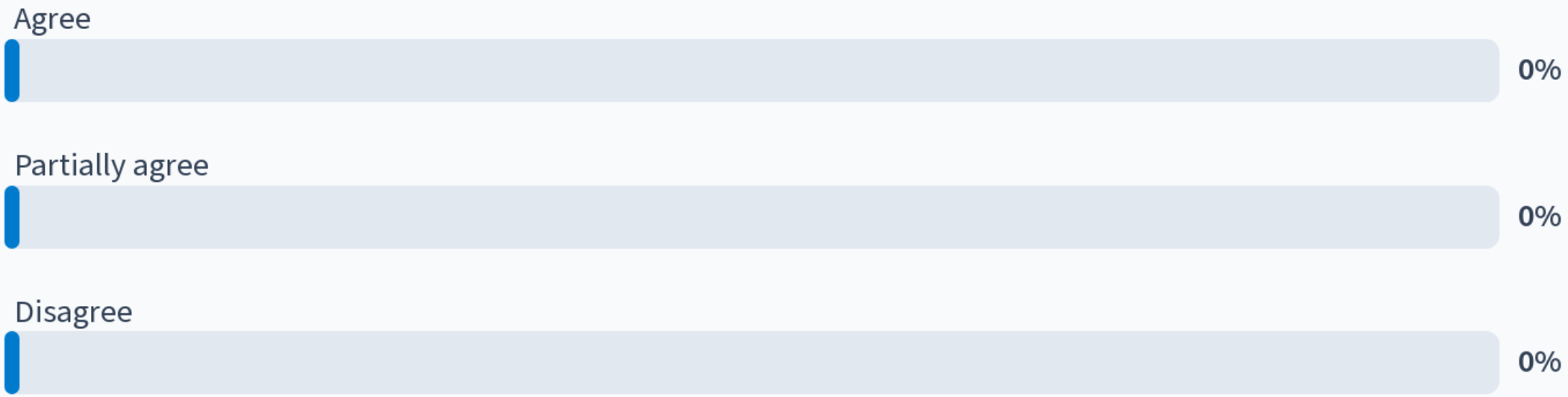
@Support the alignment between the interpretations of FAIR for digital objects and the criteria for data services, including repositories to be FAIR enabling at whatever level of care they provide



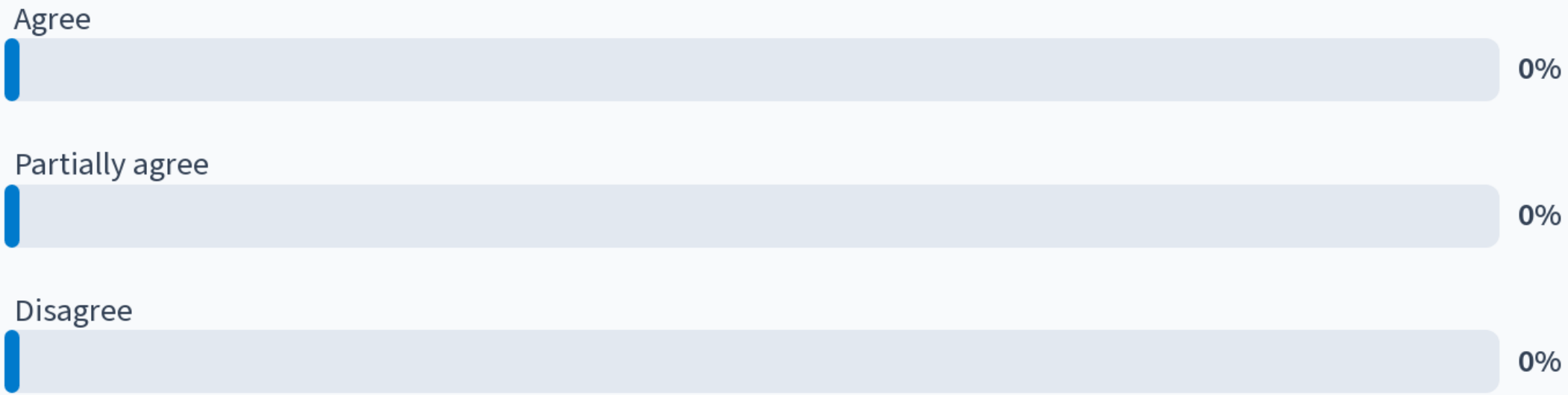
@Establish pathways for continuous quality improvement ...



@Support the development of verified business models for data services, including preservation services

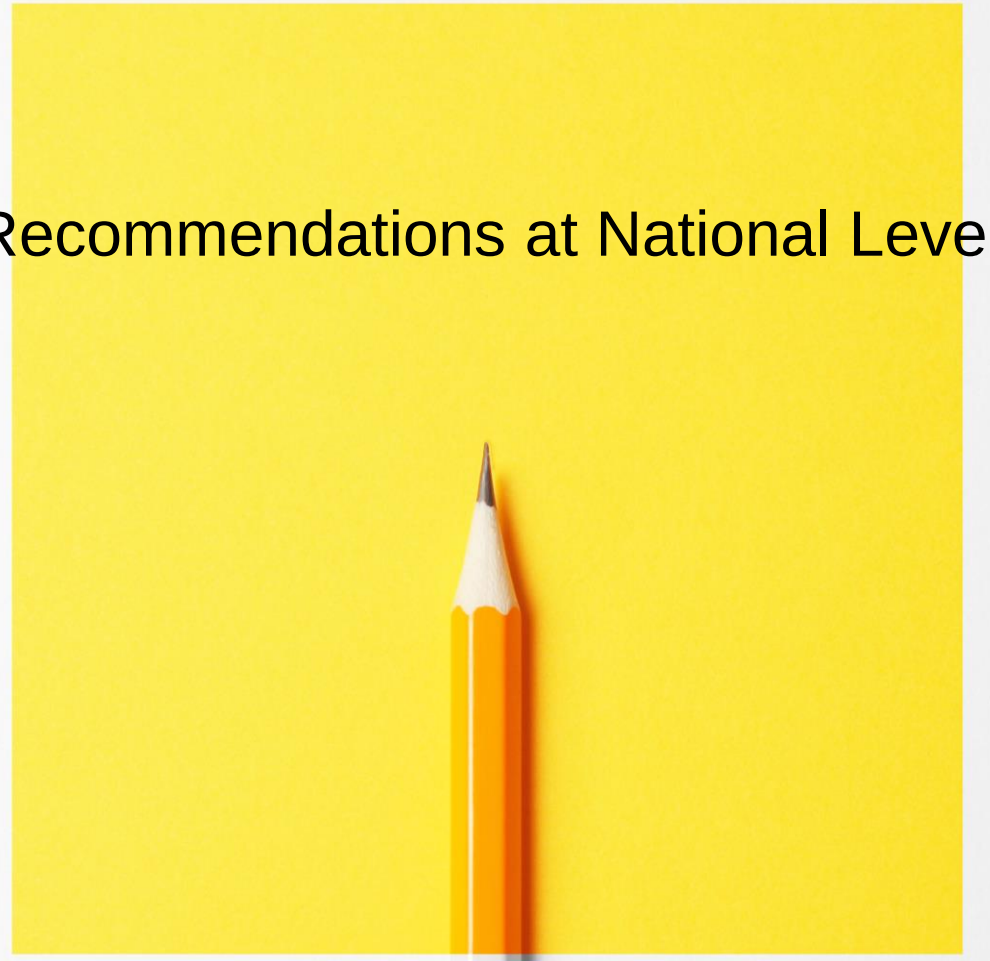


@Support the development of digital rights management standards and mechanisms that support the transition of digital objects and data services' licences and responsibilities transition to explicit, machine-actionable, scalable interoperability ...

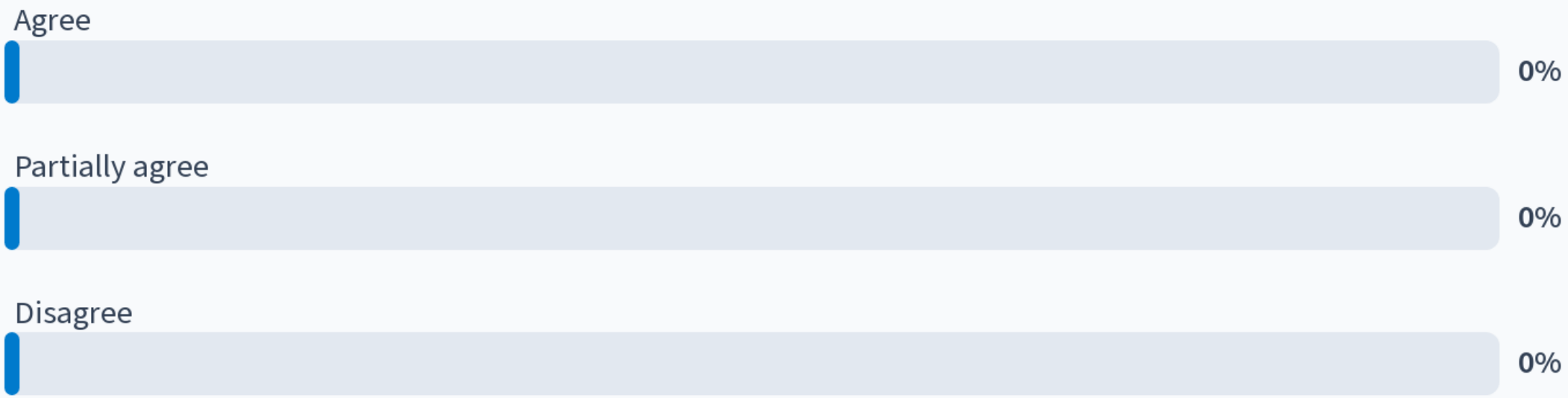


Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

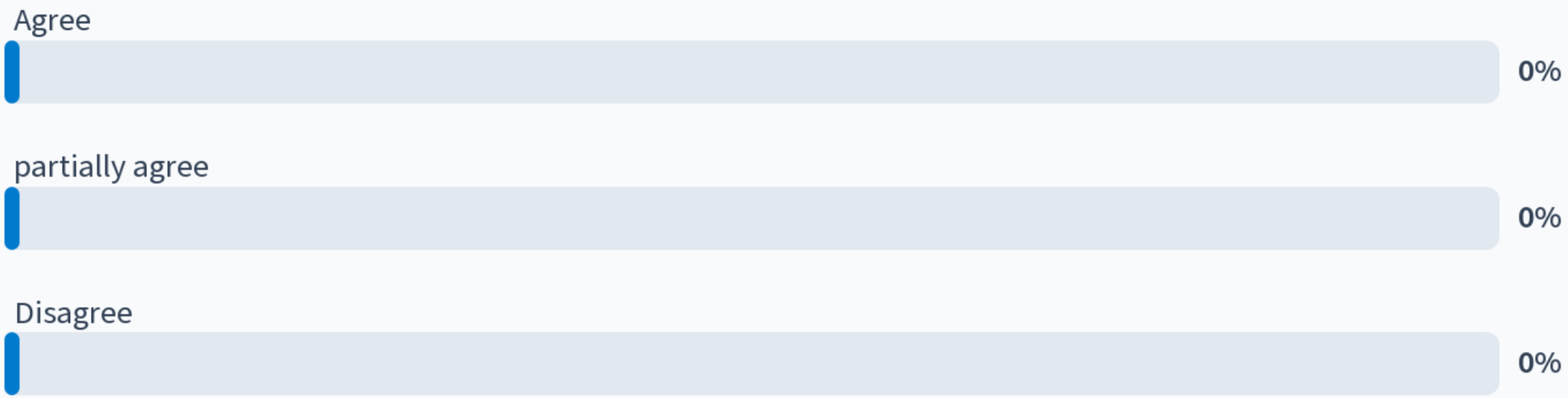
Recommendations at National Level



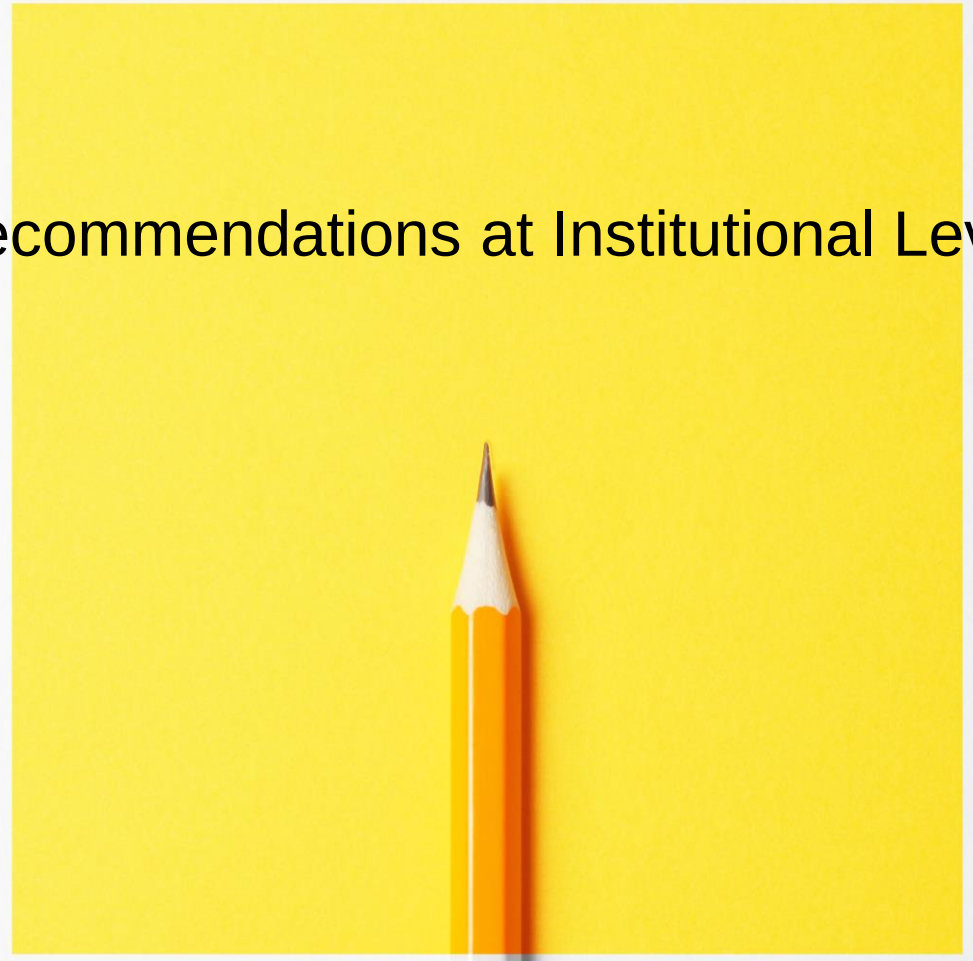
@Actions taken at European level should be adopted, adapted for national needs and support provided for implementation



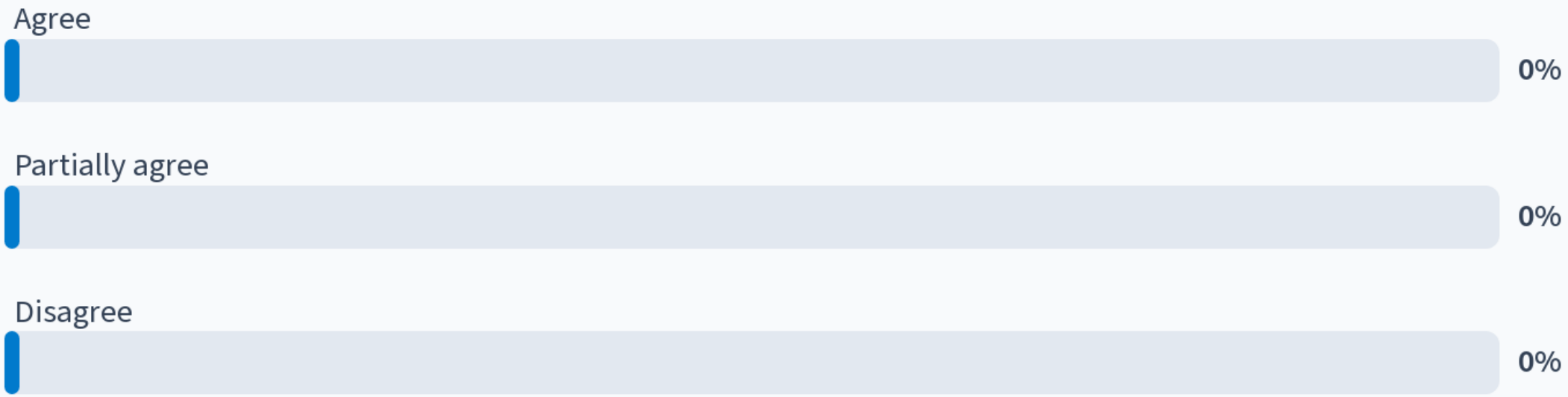
@When action cannot be taken at European level they should be developed and implemented at national level



Recommendations at Institutional Level

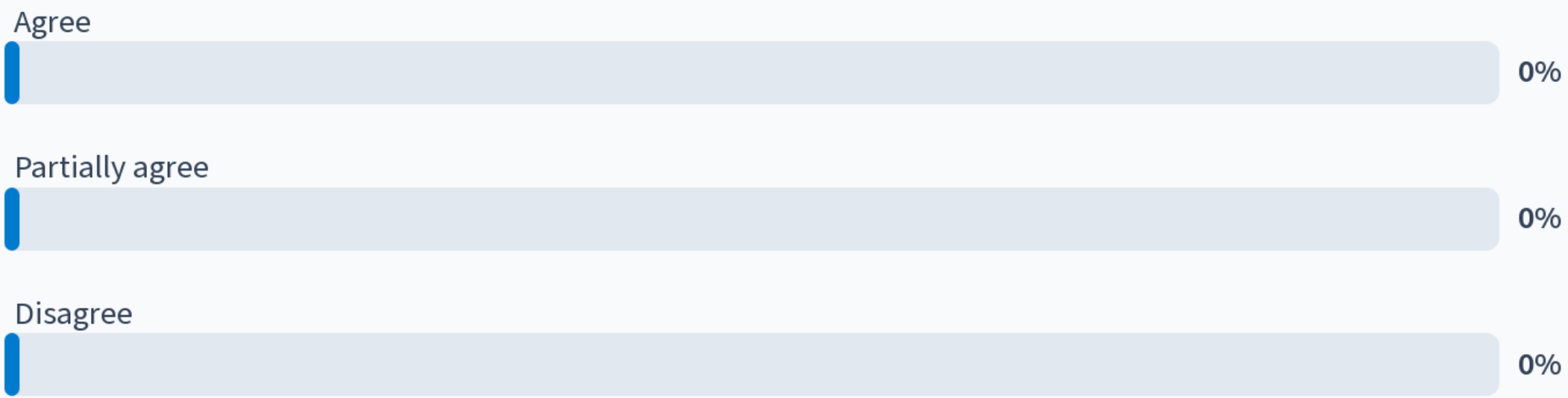


@Support researchers in digital object design, creation, storage, curation and preservation whether through locally provided repositories and data services or via third parties



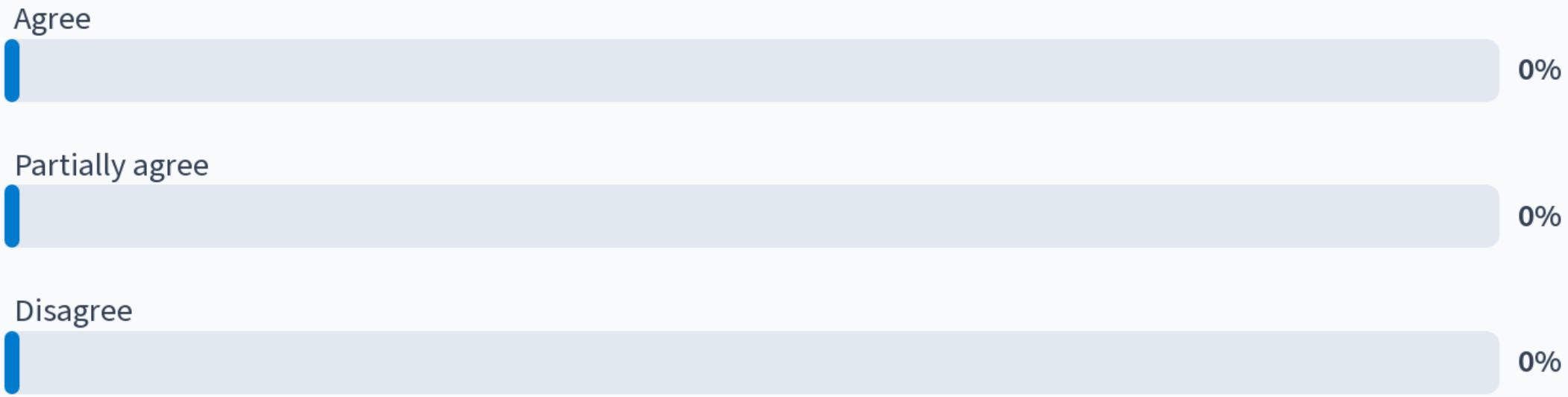
Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

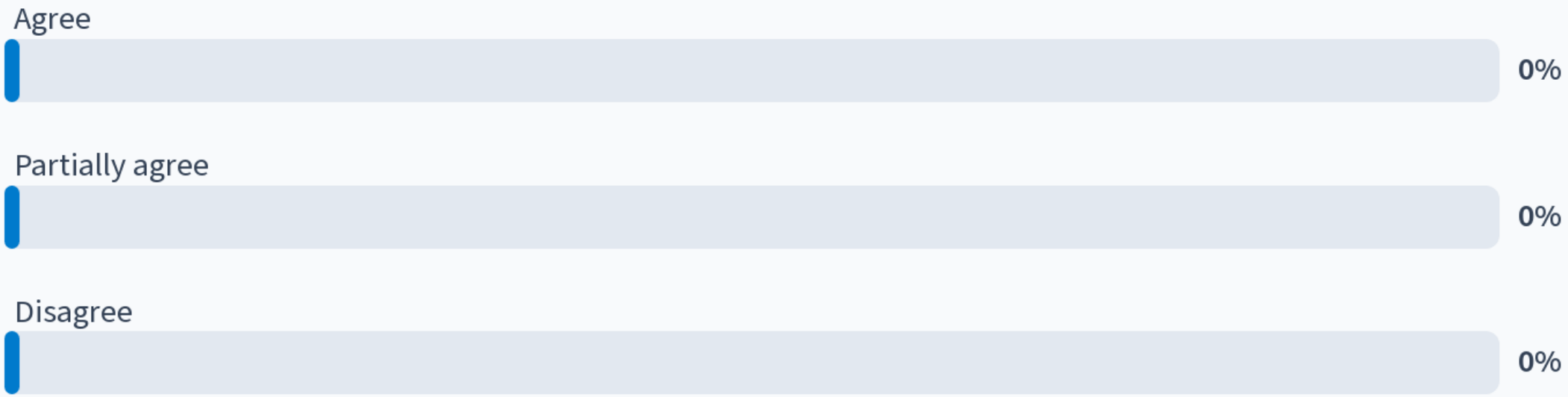
@Take appropriate responsibility for relevant phases of the research data lifecycle including assurances that digital objects are made FAIR and that services are FAIR enabling and trustworthy



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

@Support ongoing review and audit of practice across the lifecycle



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

@Funders should commission repositories to conduct audits, and repositories should undertake these audits