Facilitating Data on EOSC

sli.do: https://app.sli.do/event/nGUR465MsYAjQJxGKTxkgL

#EOSCPillar2710

Theresa Kernecker (University of Vienna)

Rebecca Reichenbach (Fraunhofer IWM)

Inge van Nieuwerburgh (Ghent University)

Facilitating data on EOSC

* A main EOSC goal is to **support the initiatives for national coordination of data infrastructures** and services and **provide stakeholders with an environment to publish, find and reuse data, tools and services for research**

* To achieve this goal, EOSC thus needs reliable and easily accessible data on infrastructures on stakeholders

Facilitating data on EOSC

* **Question:** How do we collect and facilitate quality data on EOSC infrastructures and stakeholders in a coordinated and sustainable way?

* A main way for EOSC to collect data has been via web surveys

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

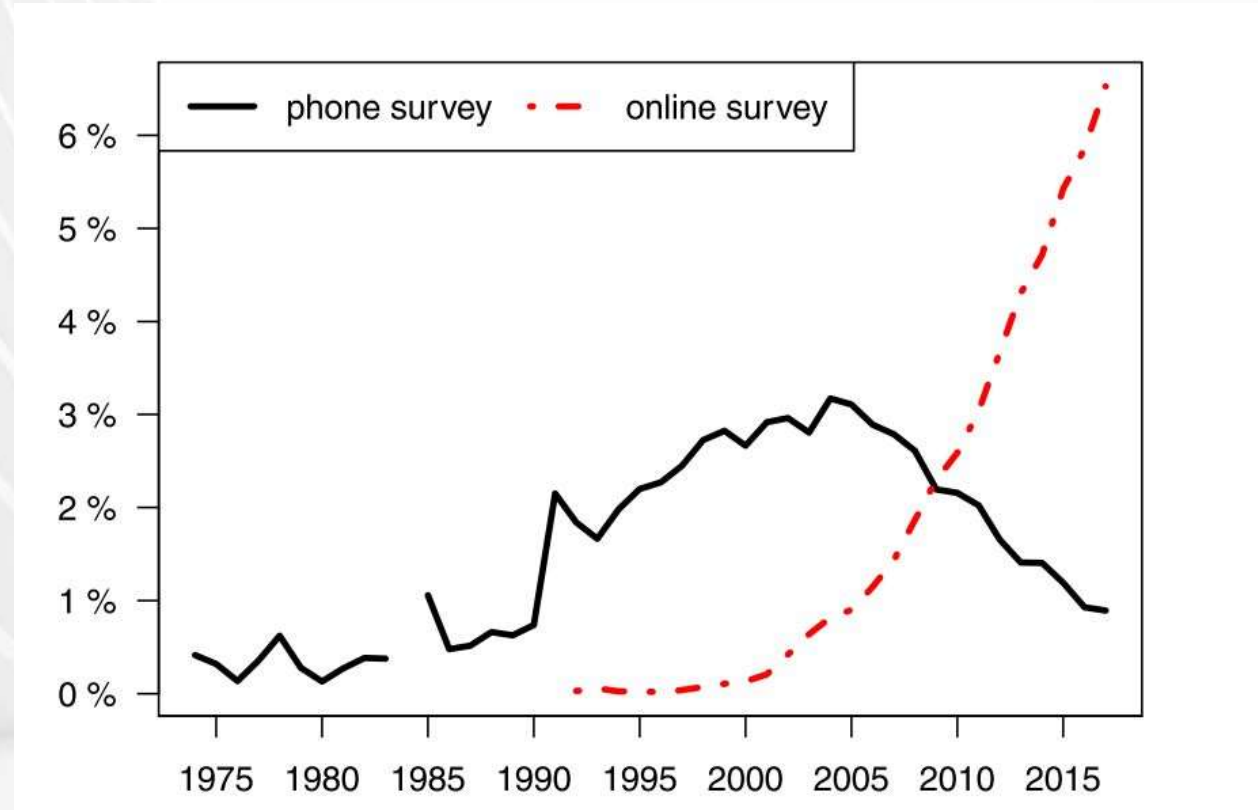# Questions for Audience

* In the past two years, how often have you received survey invitations regarding EOSC, open science, or FAIR principles?
    * very often, often, sometimes, rarely, never

* How often have you completed the surveys you were invited to?
    * very often, often, sometimes, rarely, never

@EOSCPillar /company/eosc-pillar www.eosc-pillar.eu

Facilitating data on EOSC

Phone and online survey publications as percentage of all survey publications in the Social Sciences Citation Index (Lehdonvirta et al., 2021, p. 136)

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

The limitations of web surveys I: Survey Quality

* **Population parameters:** Findings are not generalizable if the population to which surveys are distributed **cannot be identified, understood or described** (Sammut et al, 2021, Lehdonvirta et al 2021, Wu et al 2022)

* **Probability versus nonprobability samples:** nonprobability surveys fall short in representativeness (Lehdonvirta et al., 2021) and lead to coverage bias: **under/overcoverage, self-selection**

The limitations of web surveys II: Survey quality

* **Lower response rates** with online surveys compared to conventional methods (Wu et al, 2022; Sammut et. al.,2021; Daikeler et. Al, 2020; Manfreda et al. 2008; Couper, 2007), similar to item nonresponse (Cehovin et al., 2022)

* **COVID-19** has led to an **increase in surveys**, also a three-fold increase in survey dissemination by social media -> increase in **survey fatigue leads to nonresponse** (de Koning et al.,2021)

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

Question for Audience

* What motivates you to answer online surveys about EOSC?
    * open answer
* What annoys you about online surveys?
    * open answer

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

Example: National Initiatives Survey

* Focus on mapping the research (supporting) infrastructure in Europe
  * Business models, SLAs and user support, access policies, AAI and licensing, data management and FAIRness, user communities and services
  * Targets: E-infrastructures, research infrastructures, universities, funding bodies in AT, BE, DE, IT, FR
  * Method: online survey
  * Extensive research and preparation behind survey design and defining the population
  * 2,204 organizations invited: 688 started survey, 603 completed the survey
  * Response rate: 27%

@EOSCPillar /company/eosc-pillar www.eosc-pillar.eu

Example: Researcher Survey

* Focus on communities and usage of services, research input regarding reuse and services
    * Targets: Researchers from selected universities in AT, BE, DE, IT, FR
    * Reuse behavior, aspects and attitudes of data reuse and perspectives on available services
    * Method: online survey
        * Invitations and reminders were sent by different means– legal issues—variability amongst universities
    * Response rate: Number of researchers contacted unknown. 941 researchers initiated survey and 592 completed it. We cannot calculate response rate.

@EOSCPillar /company/eosc-pillar www.eosc-pillar.eu

Organisations: Services provided

Which service(s) does your organisation provide to the research community?
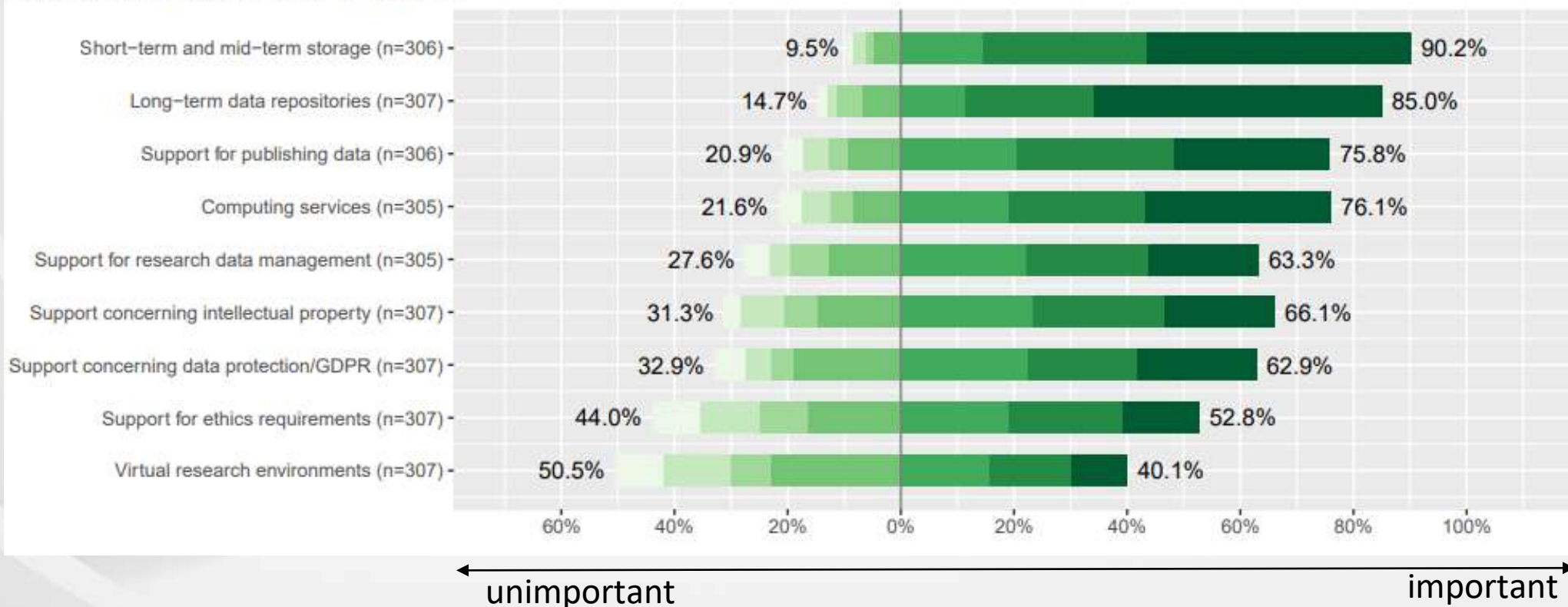
* Data infrastructures for storing and managing data:
253 "yes"-responses

* High-bandwidth networks that transport research data:
31 "yes"-responses

* High-performance computing for processing research data:
69 "yes"-responses

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

Researchers: Importance of different services

Important services for research

Service	unimportant	important
Short-term and mid-term storage (n=306)	9.5%	90.2%
Long-term data repositories (n=307)	14.7%	85.0%
Support for publishing data (n=306)	20.9%	75.8%
Computing services (n=305)	21.6%	76.1%
Support for research data management (n=305)	27.6%	63.3%
Support concerning intellectual property (n=307)	31.3%	66.1%
Support concerning data protection/GDPR (n=307)	32.9%	62.9%
Support for ethics requirements (n=307)	44.0%	52.8%
Virtual research environments (n=307)	50.5%	40.1%

unimportant — important

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

FAIRness: NIS and researcher survey data

* NIS Dataset and documentation

* https://data.aussda.at/dataset.xhtml?persistentId=doi:10.11587/VOSVGK

Bodlos, Anita; Hönegger, Lisa; Kaczmirek, Lars; Beckmann, Volker; Breton, Vincent; Romier, Geneviève; Van Wezel, Jos; Streit, Achim; Stevanovic, Uros; Galeazzi, Fulvio; Tanlongo, Federica; Van Nieuwerburgh, Inge, 2019, "EOSC Pillar "National Initiatives" Survey (SUF edition)", https://doi.org/10.11587/VOSVGK.

* Researcher dataset and documentation

* https://data.aussda.at/dataset.xhtml?persistentId=doi:10.11587/D0UMOH

Bodlos, Anita; Hönegger, Lisa; Kernecker, Theresa; Geistberger, Julia; Di Giorgio, Sara; Mathieu, Gilles; Van Nieuwerburgh, Inge; Romier, Genevieve; Szabo, Dimitri; Tanlongo, Federica; Vernet, Marine; Van Wezel, Jos; Reichenbach, Rebecca; Rouchon, Olivier; Hermans, Emilie; Sanaa, Yosra, 2022, "EOSC Pillar Researcher Survey (SUF edition)", https://doi.org/10.11587/D0UMOH.

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

Questions

* How can we use surveys to produce reliable data on EOSC?

* How can we track infrastructure development and stakeholder perspectives over time, even after projects are finished?

* How can we easily check if any other EOSC projects collected data on similar topics and share efforts?

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu

Looking forward

* Two major ways we can accomplish sustainable data collection and data management in EOSC
    * Improving the quality of surveys across EOSC projects
        * Collect questions from all EOSC surveys
        * Provide templates and methodology on survey design, etc.
        * Designate resources that aim at maximizing reusability
        * Avoid over-researching and encourage data sharing
    * Provide resources to a centrally-coordinated EOSC entity
        * Collect questions from all EOSC surveys and provide question templates to all EOSC projects
        * Collect all surveys by different realms of EOSC and make them easily accessible and reusable

@EOSCPillar   /company/eosc-pillar   www.eosc-pillar.eu