

Up-to-the-minute Data Policy Updates for Participatory Studies

Aakash Sharma

Sharing data is the norm

- 147,000 Datasets

- 38.2 Million downloads

- Social Sciences 30,300 (21%)

- Medicine, Health and Life Sciences 7,420 (5.1%)

<https://dataverse.org/metrics>

The screenshot shows a web browser window for 'dataverse.no'. The header includes the DataverseNO logo, a search bar, and a 'Deposit Guide' link. A 'CORE TRUST SEAL' badge is present. The main content area displays search results for datasets. On the left, there are filters for 'Distributor Name', 'Subject', 'Keyword Term', and 'Author Name'. The right side lists three dataset entries with preview icons and detailed descriptions.

Distributor Name	Replication data for: Perinatal exposure to a human relevant mixture of persistent organic pollutants: Effects on mammary gland development, ovarian folliculogenesis and liver in CD-1 mice.
UiT Open Research Data (676)	May 28, 2021 - NMBU Open Research Data Johanson, Silje Modahl, 2021, "Replication data for: Perinatal exposure to a human relevant mixture of persistent organic pollutants: Effects on mammary gland development, ovarian folliculogenesis and liver in CD-1 mice.", https://doi.org/10.18710/MSTY0R , DataverseNO, V1
NORD Open Research Data (132)	Complete raw data for the results presented in the manuscript Perinatal exposure to a human relevant mixture of persistent organic pollutants: Effects on mammary gland development, ovarian folliculogenesis and liver in CD-1 mice. The ability of persistent organic pollutants (POPs...)
UiB Open Research Data (86)	Replication data for: Devonian–Carboniferous collapse and segmentation of the Billefjorden Trough, and Eurekan inversion–overprint and strain partitioning and decoupling along inherited WNW–ESE-striking faults
The Tromsø Repository of Language and Linguistics (55)	May 28, 2021 - UiT Open Research Data Koehl, Jean-Baptiste P., 2020, "Replication data for: Devonian–Carboniferous collapse and segmentation of the Billefjorden Trough, and Eurekan inversion–overprint and strain partitioning and decoupling along inherited WNW–ESE-striking faults", https://doi.org/10.18710/V2
NTNU Open Research Data (22)	High-resolution versions of the figures and supplements of the Koehl et al. (2020) manuscript "Devonian–Carboniferous collapse and segmentation of the Billefjorden Trough, and Eurekan inversion–overprint and strain partitioning and decoupling along inherited WNW–ESE-striking faults"
More...	Replication data for "The Billefjorden Fault Zone north of Spitsbergen: A major terrane boundary?"
Subject	May 28, 2021 Koehl, Jean-Baptiste, 2021, "Replication data for "The Billefjorden Fault Zone north of Spitsbergen: A major terrane boundary?", https://doi.org/10.18710/LP3YMQ , DataverseNO, V2
Earth and Environmental Sciences (684)	High-resolution of the figures in Koehl & Allaart (2021; "The Billefjorden Fault Zone north of Spitsbergen: A major terrane boundary?").
Physics (501)	
Medicine, Health and Life Sciences (207)	
Arts and Humanities (106)	
Social Sciences (47)	
More...	
Keyword Term	
ionosphere (405)	
aurora (404)	
SUN-EARTH INTERACTIONS (403)	
ionodata (403)	
ionogram (403)	
More...	
Author Name	
Tromsø Geophysical Observatory (403)	

Sensitive datasets cannot be public

- Trust issues [Bongartz et al. 2017]
- Easy to identify individuals [Salerno et al. 2017, Goodman and Meslin 2014]
- Consent revocations (GDPR, GPDPR)
- Tromsø study - ‘taushetsplikt’

SPØRRESKJEMA

T T

Dato for utfylling
Dag mnd år

Kjente sykdommer:

1. Har du hatt angina eller hjerteinfarkt..... Ja Nei Usikker
2. Har du hatt hjerneslag eller drypp.....
3. Har du diabetes mellitus (sukkersyke).....

Hvis ja: nåværende behandling..... Kun diett Tablet Insulin

Diagnoseår: Antall år på insulin:

Røyking:
Har du noen gang vært daglig røyker..... Ja Nei
Har du noen gang vært daglig sigarillorøyker.....
Har du noen gang vært daglig piperøyker.....
 Jeg har aldri vært daglig røyket.....

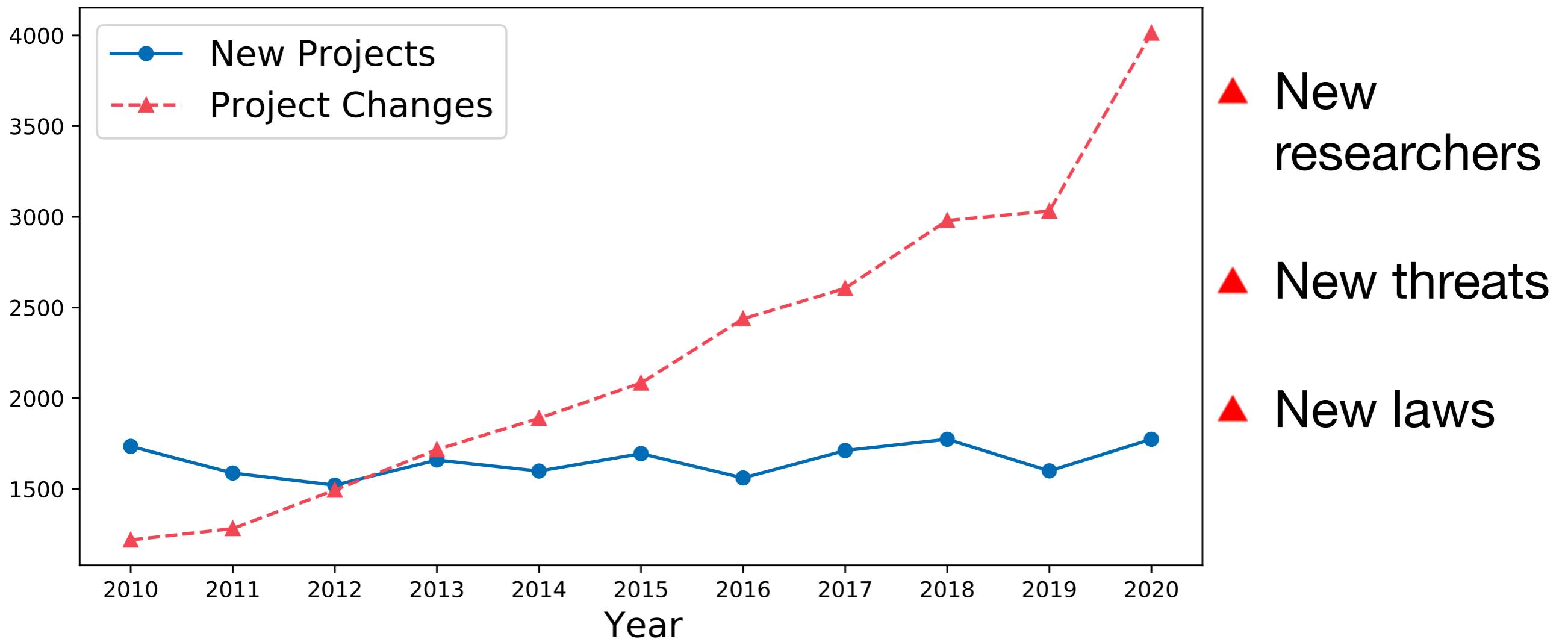
Hvis du er eller har vært daglig røyker:
Alder ved røykestart: Alder i år
Alder ved siste røyk: Alder i år
Antall år med røyking: År
Antall sigarettter (eller tilsvarende) per dag: Antall
1 cigar = 4 sigarettter, 1 sigarillo = 2 sigarettter, 1 pipe = 2,5 sigarettter

Familieanamnese:

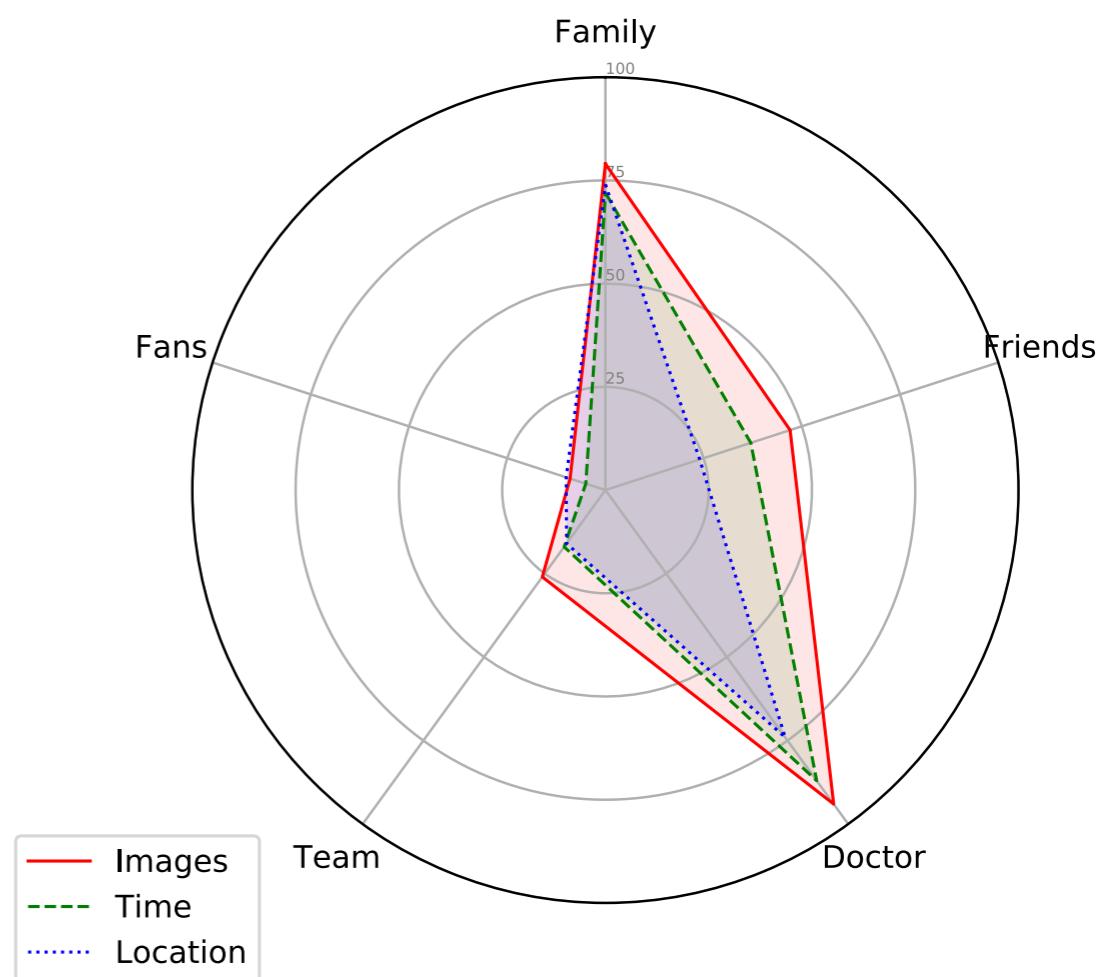
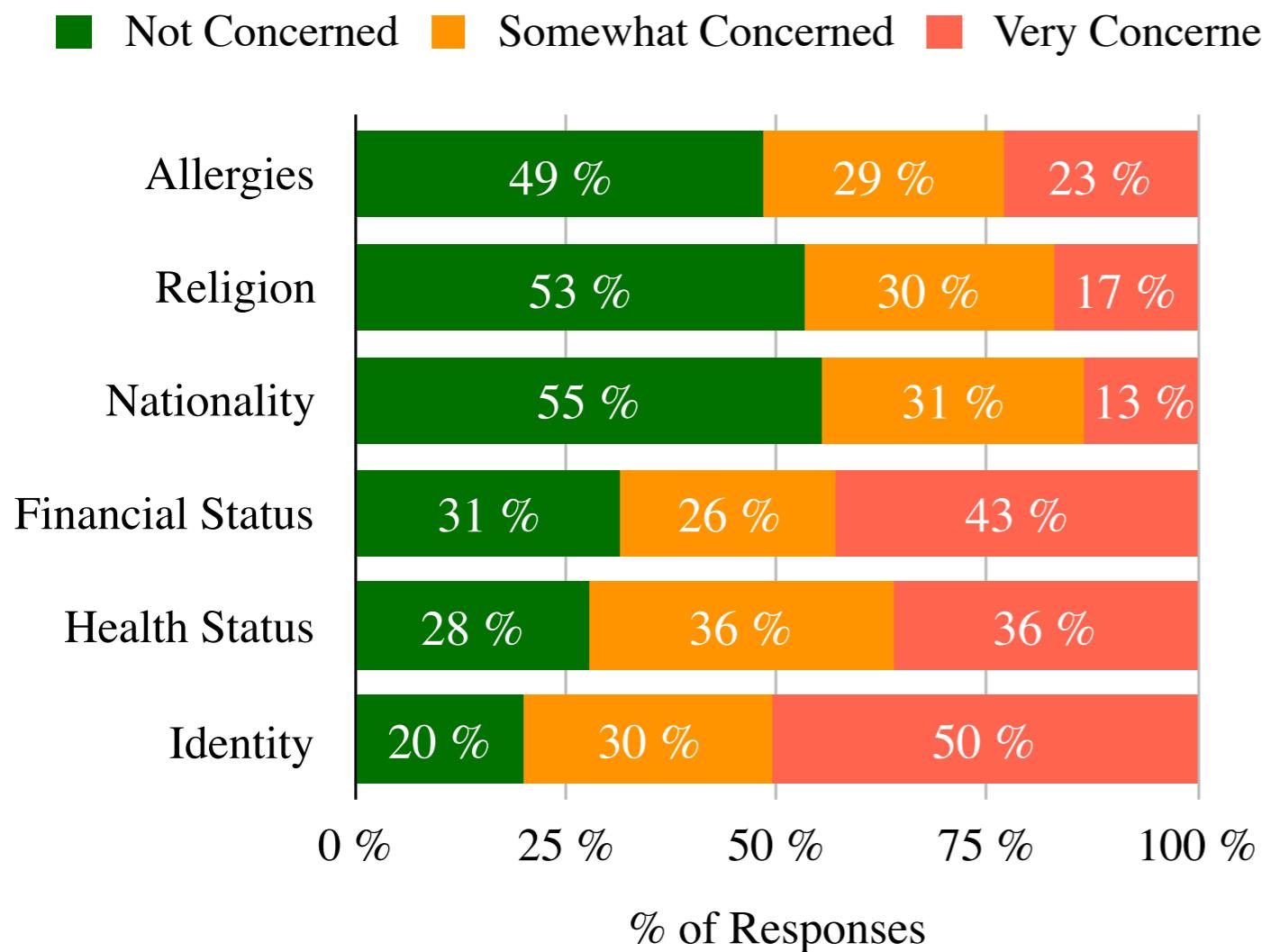
Antall søsken (inkludert deg selv):
Antall barn:
Antall førstegradsslektinger med tykk- eller endetarmskreft:
Antall førstegradsslektinger med tykk- eller endetarmskreft for føltre 60 år:

CONFIDENTIAL

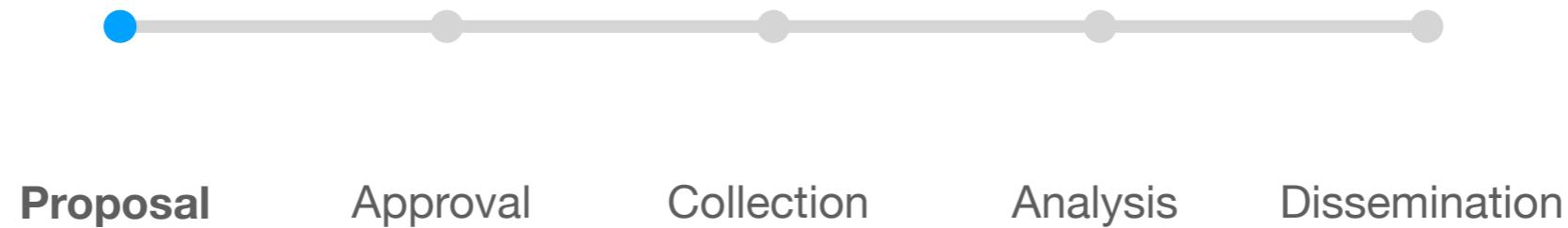
Projects are not static



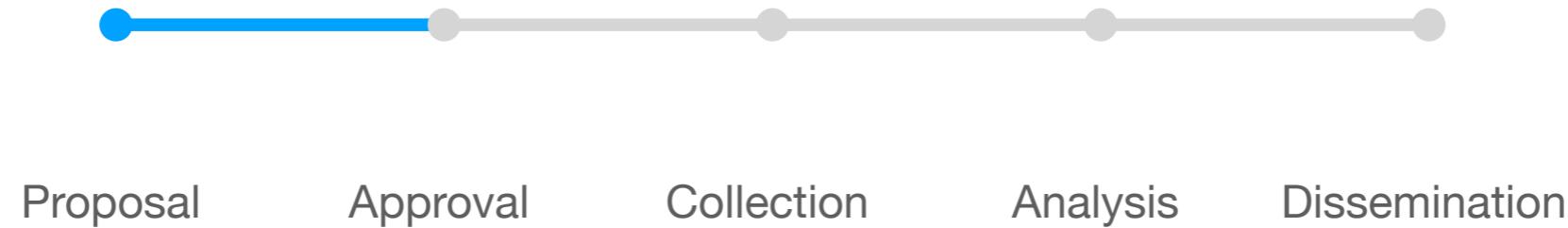
Opinions change



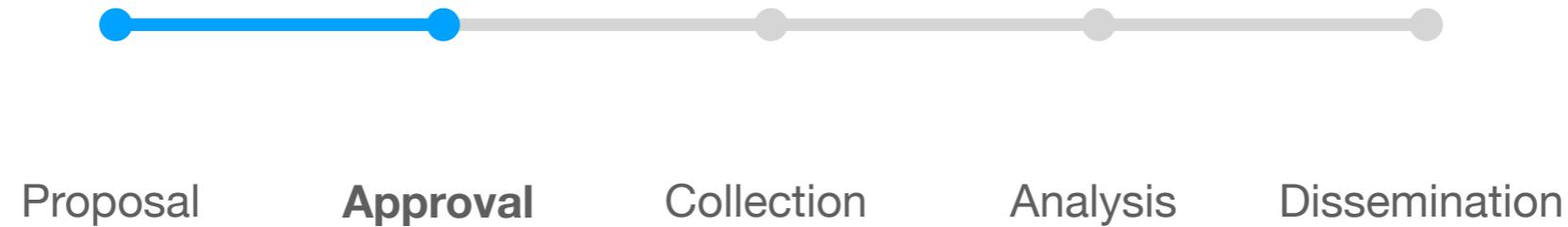
A project's lifecycle



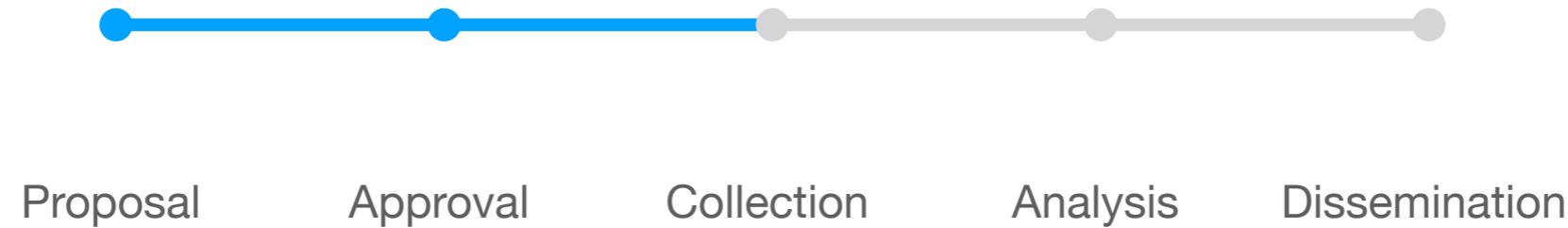
A project's lifecycle



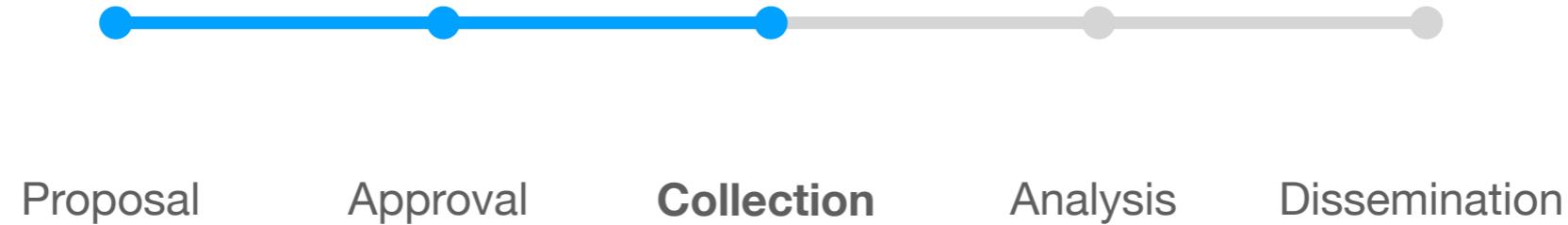
A project's lifecycle



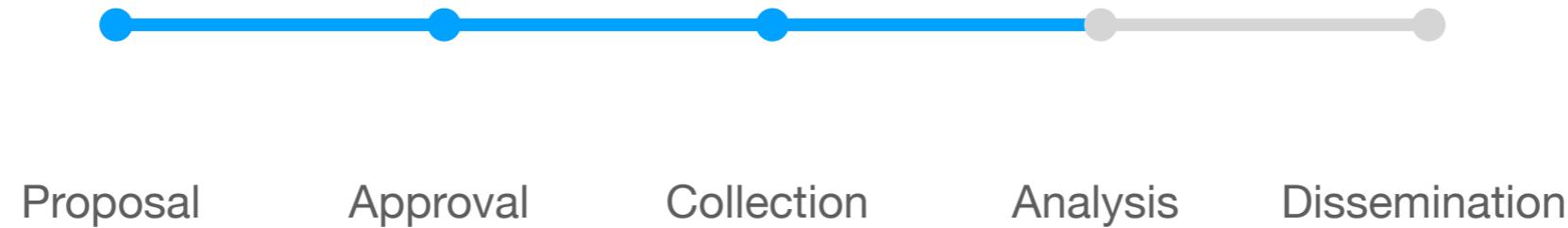
A project's lifecycle



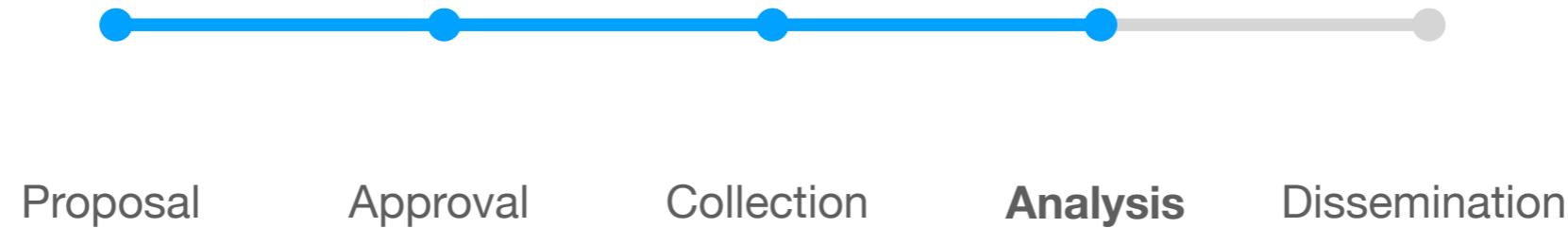
A project's lifecycle



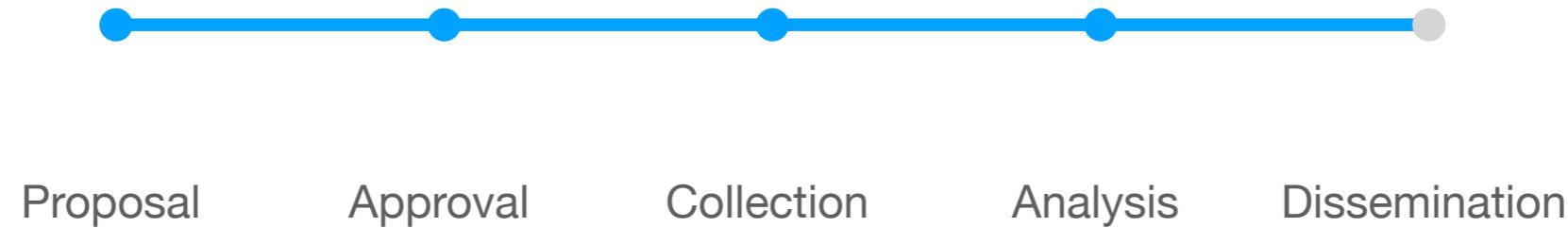
A project's lifecycle



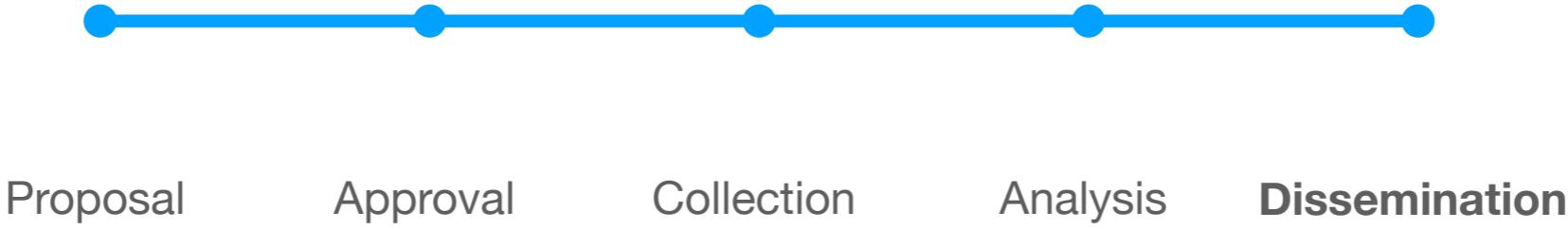
A project's lifecycle



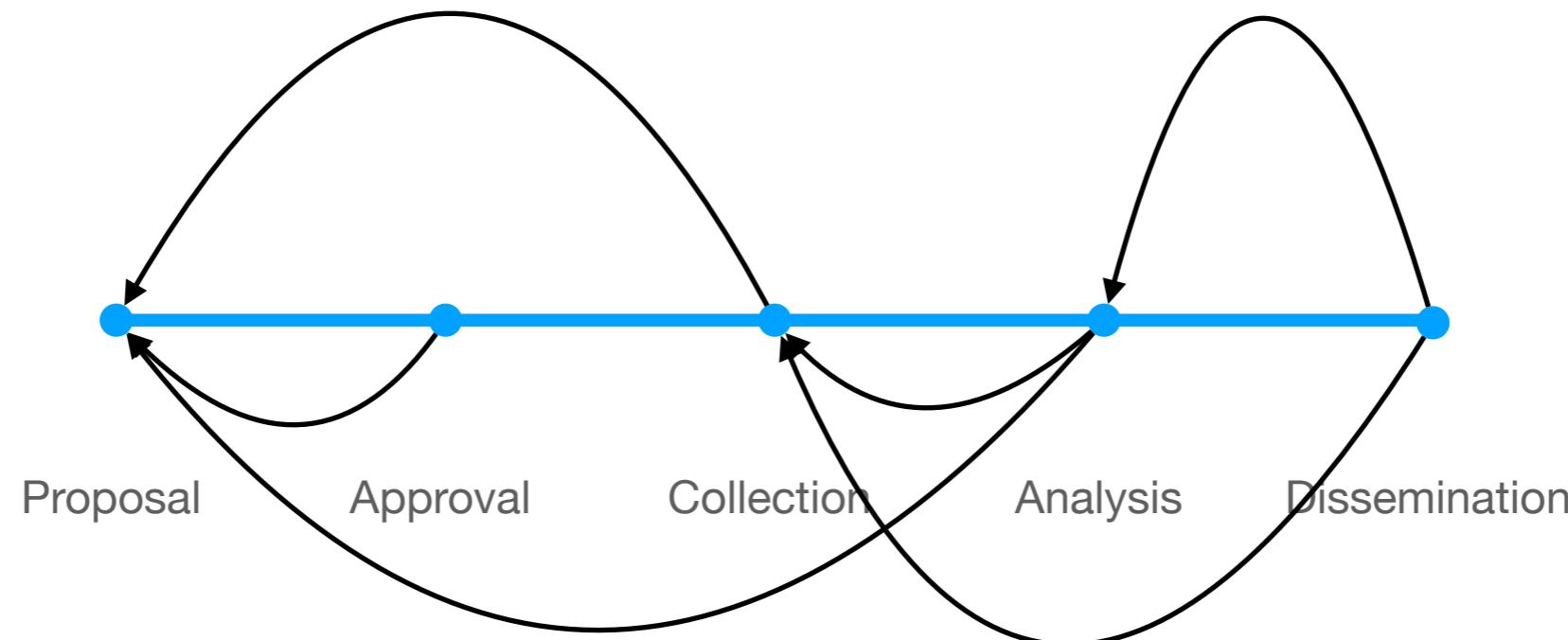
A project's lifecycle



A project's lifecycle



A project's lifecycle is non-trivial



We need system support for dynamic security policies

Sensitive data leads to silos

Tension between

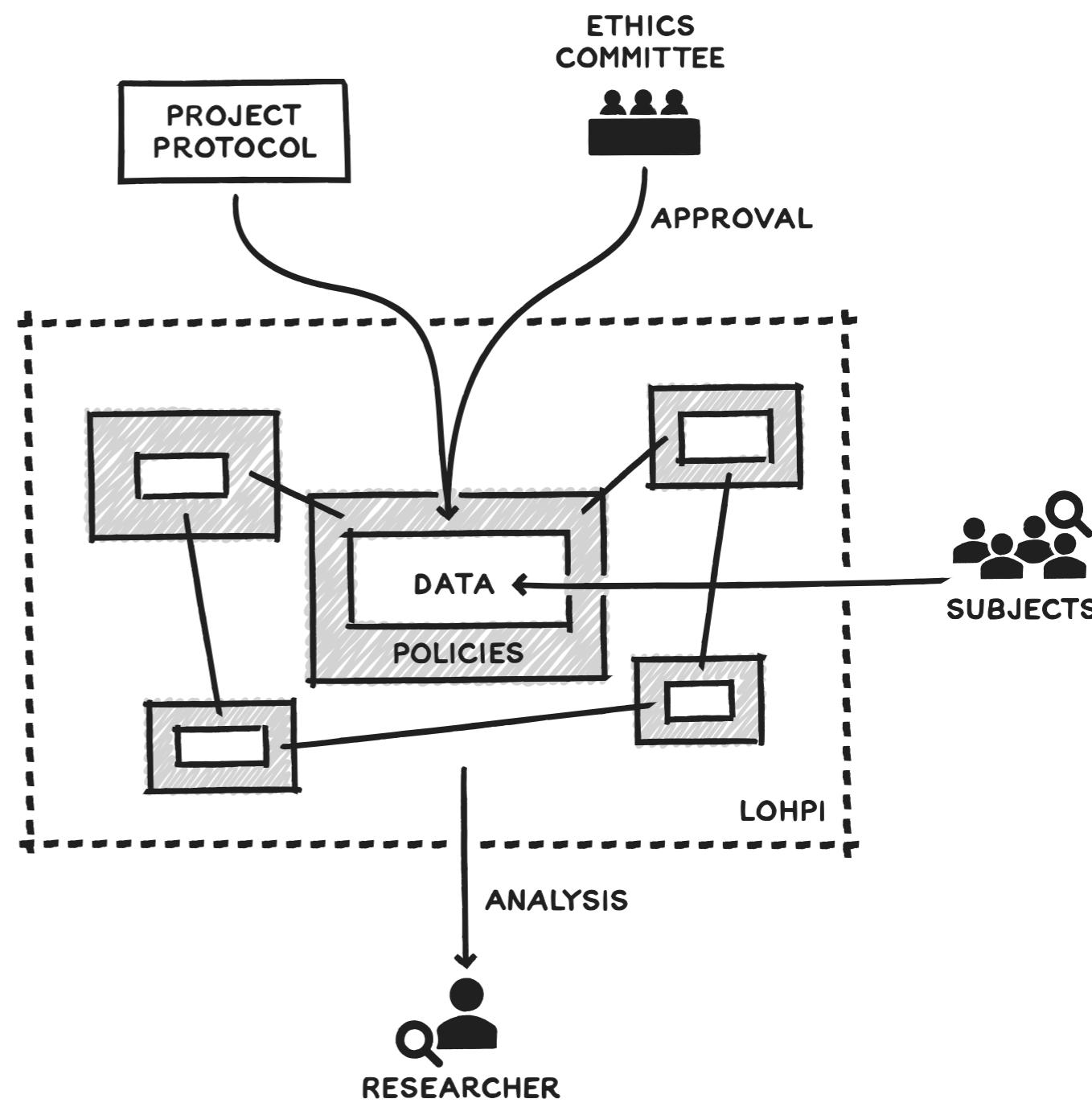
- Open research (public data)
 - ▶ Decentralized research model
 - ▶ Operate on institutional infrastructure
- Privacy risks
 - ▶ Projects operate in silos
 - ▶ Shared computational infrastructure (trusted by all parties, TSD)

Compliant sharing with Lohpi

A distributed **metadata layer** that enables compliant data sharing.

A dataset's security policy can be **updated in near real-time**.

LOHPI[®]



Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
```

```
Login link 🌎:
```

```
https://login.microsoftonline.com/c628abe3-0fe1-4381-bcc0-4721d4387973/oauth2/v2.0/authorize?&state=de=query&response_type=code&scope=https%3A%2F%2Fgraph.microsoft.com%2FUser.read%20&redirect_uri=http%3A%2F%2Flocalhost%3D8080%2Fauth%2Fcallback&code_challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain
```

Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
```

Login link :

https://127.0.0.1:8080/auth?client_id=31-bcc0-4721d4387973/oauth2/v2.0/authorize&state=de=query&uiop&code_challenge_method=plain



Pick an account



Thomas Bye Nilsen
tni068@uit.no
Signed in

:



Use another account

Back

Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
```

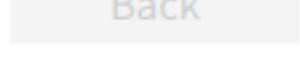
Login link :

https://login.microsoftonline.com/c628abe3-0fe1-4381-bcc0-4721d4387973/oauth2/v2.0/authorize?&state=de=query&response_type=code&scope=https%3A%2F%2Fgraph.microsoft.com%2FUser.read%20&redirect_uri=http://challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain

 Pick an account

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
```

```
{  
  "Identifiers": [  
    "Data.bin",  
    "Fish_Dataset.zip",  
    "notMNIST.zip"  
  ]  
}
```

 Back

Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
```

Login link :

https://login.microsoftonline.com/c628abe3-0fe1-4381-bcc0-4721d4387973/oauth2/v2.0/authorize?&state=de=query&response_type=code&scope=https%3A%2F%2Fgraph.microsoft.com%2FUser.read%20&redirect_uri=https://challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain

Pick an account



```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
```

	Thomas Bye Nilsen tni058@uit.no Signed in
	"Identifiers": ["Data.bin", "Fish_Dataset.zip", "notMNIST.zip"]

Use another account

Back

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh checkout notMNIST.zip
```

Downloading "notMNIST.zip"...

Success! Location: notMNIST.zip

Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
```

Login link :

https://login.microsoftonline.com/c628abe3-0fe1-4381-bcc0-4721d4387973/oauth2/v2.0/authorize?&state=de=query&response_type=code&scope=https%3A%2F%2Fgraph.microsoft.com%2FUser.read%20&redirect_uri=https://challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain

Pick an account



```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
Thomas Bye Nilsen
tni058@uit.no
Signed in
{
  "Identifiers": [
    "Data.bin",
    "Fish_Dataset.zip",
    "notMNIST.zip"
  ]
}
```

Back

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh checkout notMNIST.zip
```

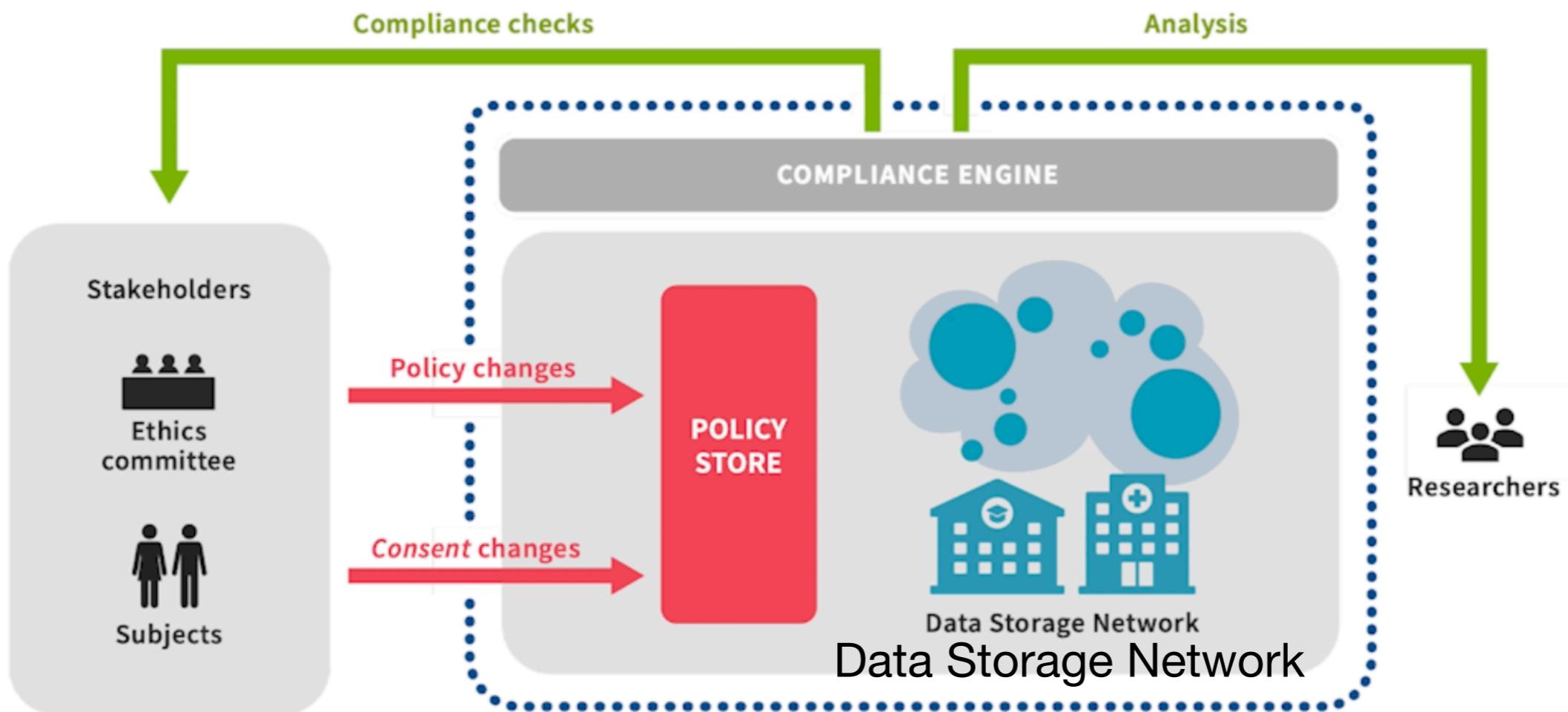
Downloading "notMNIST.zip"...

Success! Location: notMNIST.zip

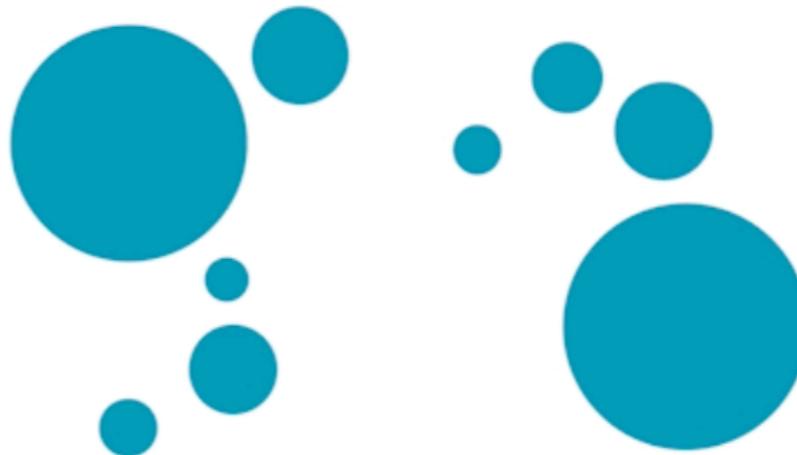
```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh comply notMNIST.zip
```

Dataset checkout is not compliant

Lohpi architecture



Secure dissemination with gossips



Jenkins, Kate, Ken Hopkinson, and Ken Birman. "A gossip protocol for subgroup multicast." Proceedings 21st International Conference on Distributed Computing Systems Workshops. IEEE, 2001.

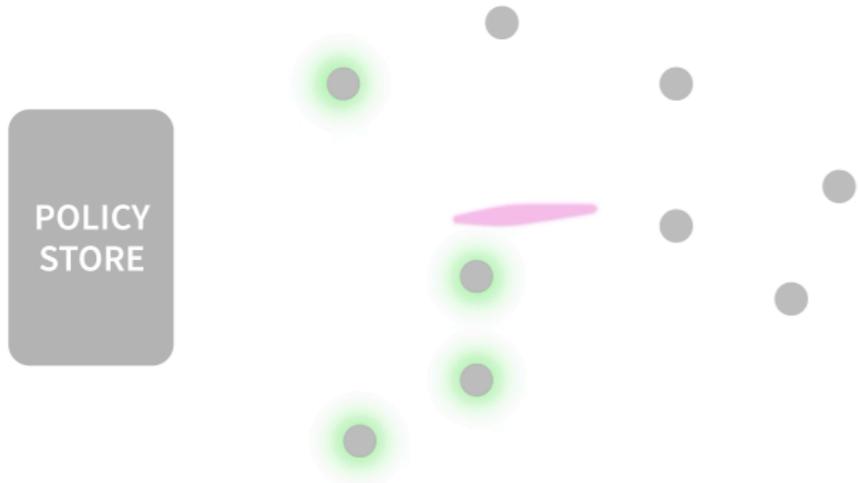
Johansen, H. D., Renesse, R. V., Vigfusson, Y., & Johansen, D. (2015). Fireflies: A secure and scalable membership and gossip service. *ACM Transactions on Computer Systems (TOCS)*, 33(2), 1-32.

Remove bottleneck

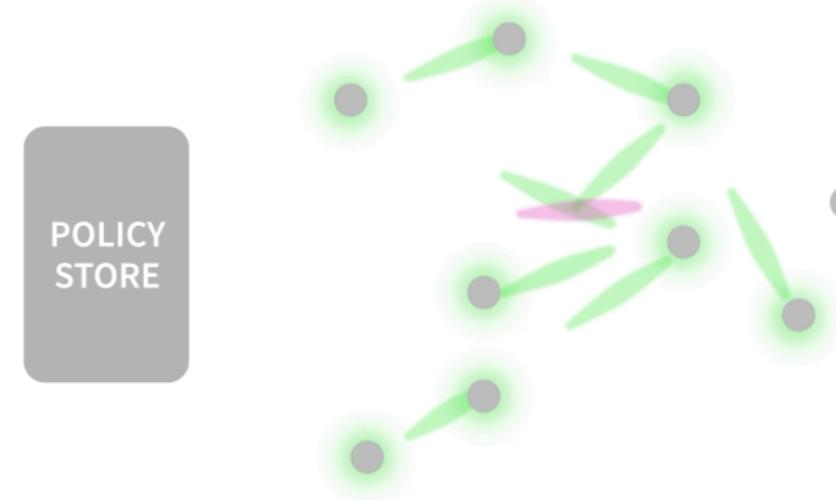


Scalability

DIRECT MESSAGING



GOSSIPING



What's in a gossip?

Info

Index

Updates

MessageID ad528370

Signature Algorithm SHA-256 ECDSA

Signature 71 bytes: 30 45 02 20 7E 3C 48 DA B9 A5
B0 2F 5A 31 EC C4 25 6B 04 ...

Index version 1.0

Policy ID#1 (0f5fbb8f, 2.23)

Policy ID#2 (c5bdaf2f, 2.5)

Policy ID#3 (bd9a29dd, 1.3)

Policy (0f5fbb8f,

{

Location code: EU-NO,
Authority: REK-NOR-8392,
Intents: research-only...

)...

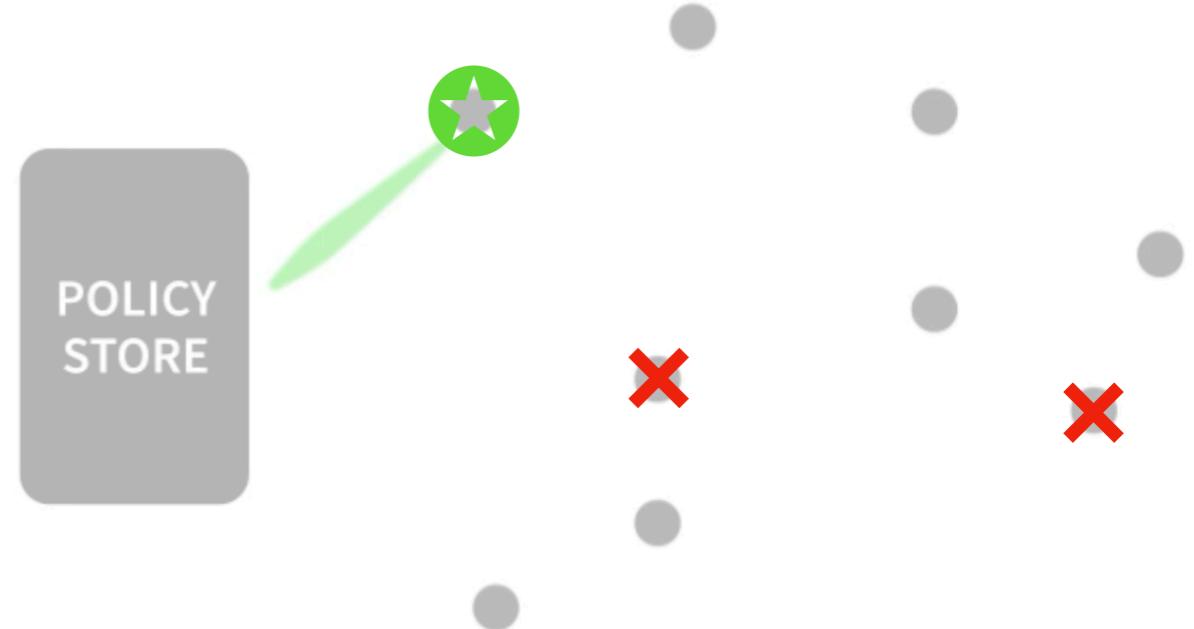
Experimental Setup

N Number of nodes

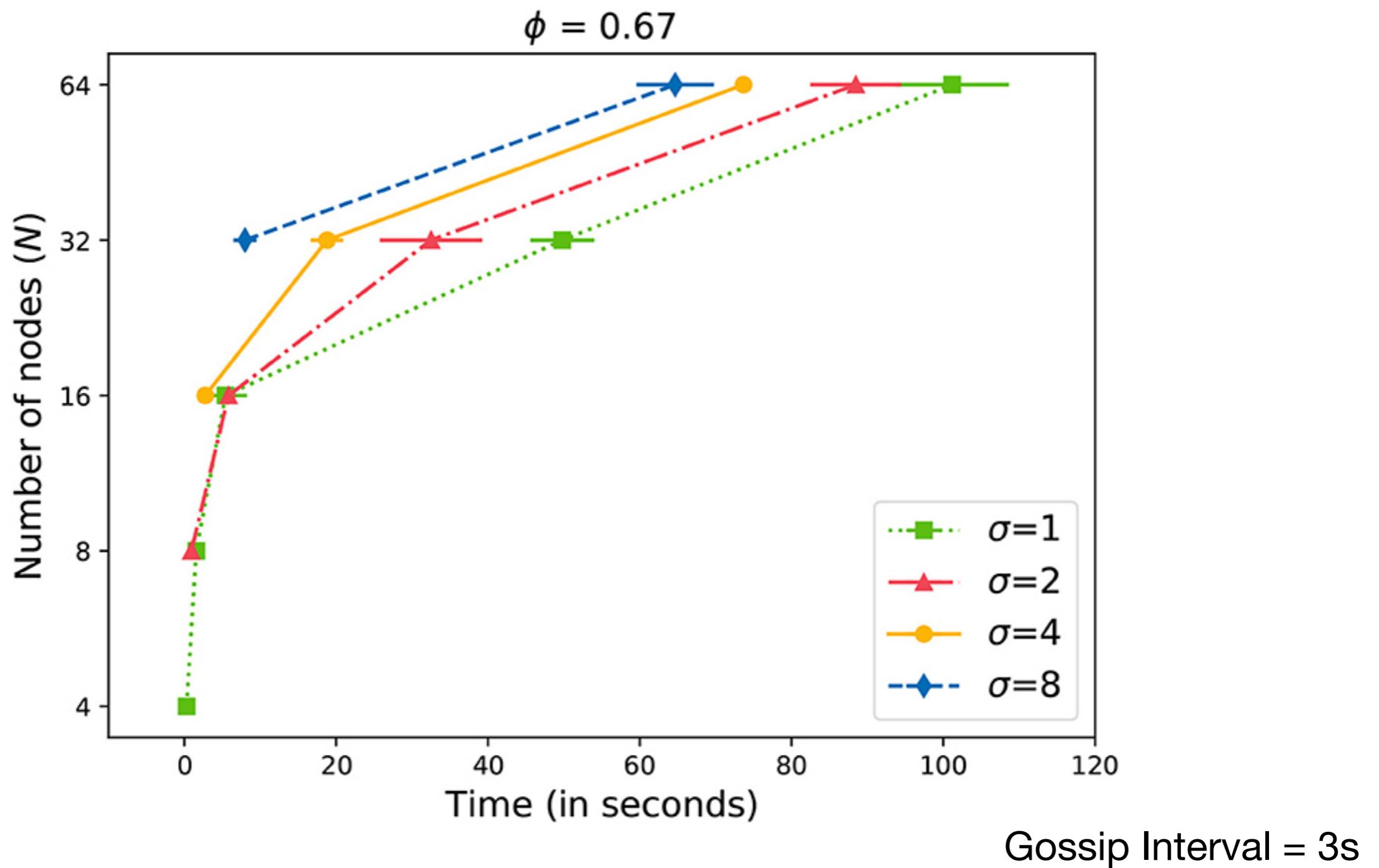
ϕ Percent of nodes that must receive a gossip

σ Multicast node

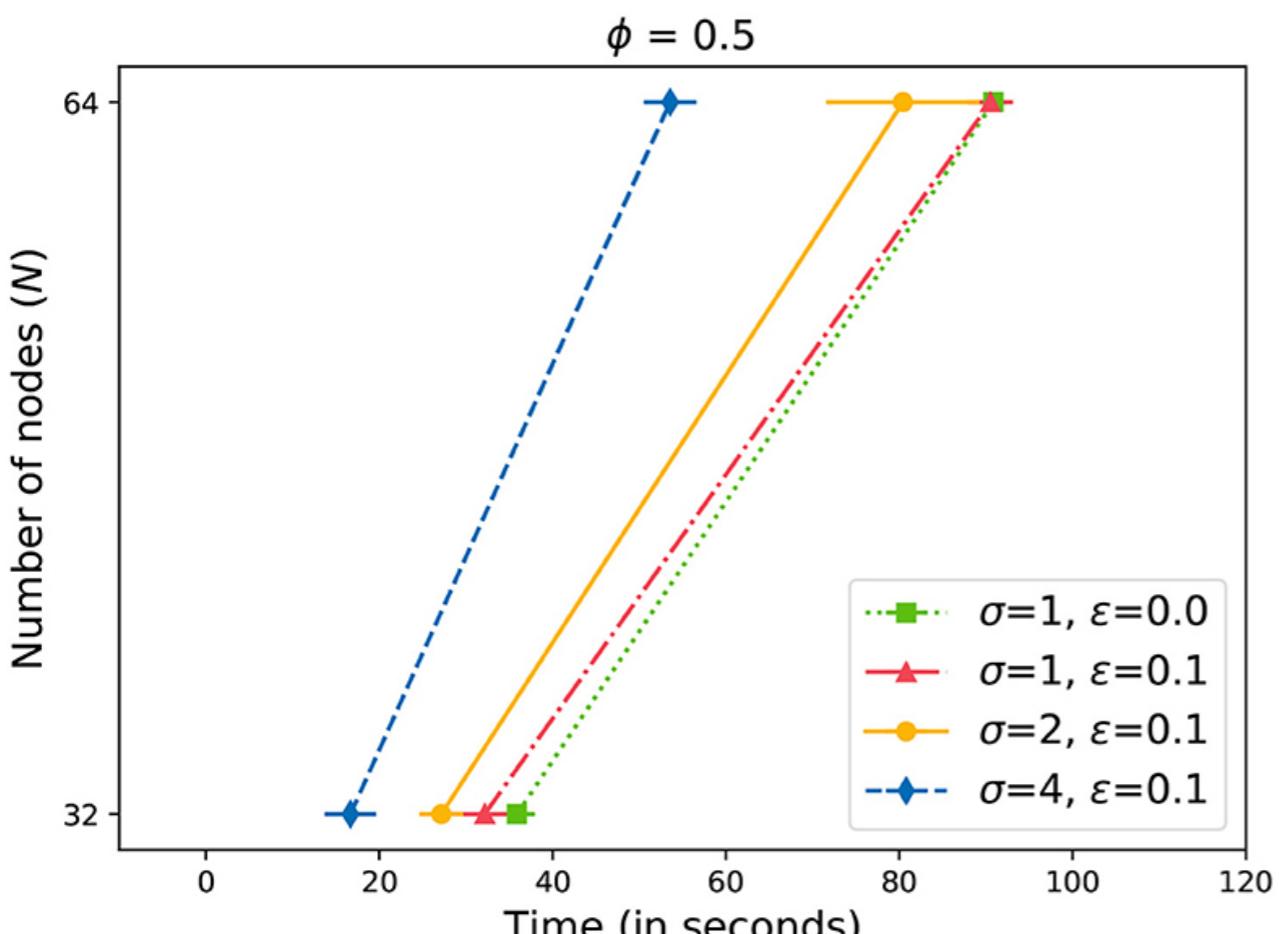
ϵ Failures



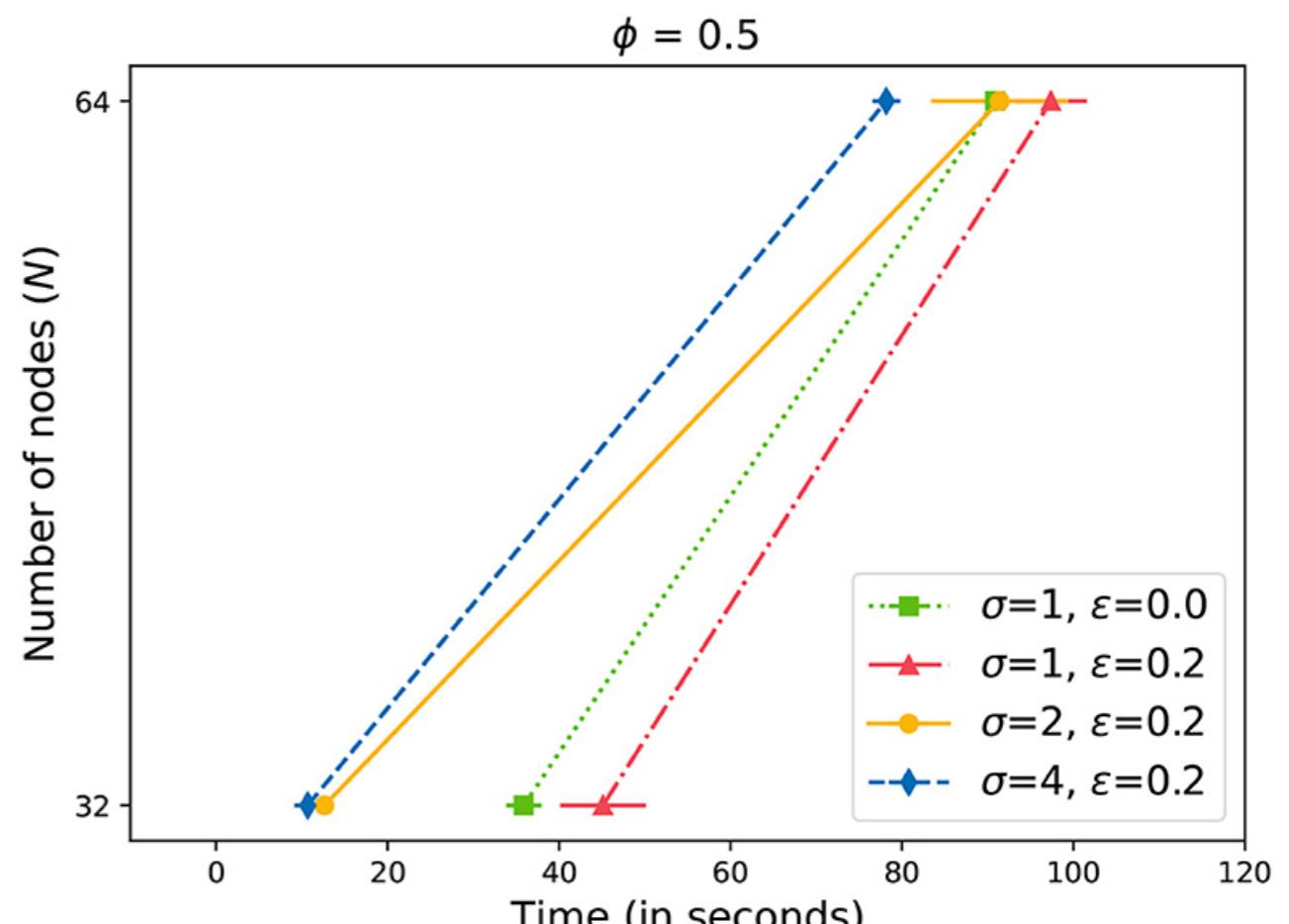
Dissemination



When nodes fail!



10% failure



20% failure

Gossip Interval = 3s

Ongoing work

- Low-code policy language
- Compliance (formal proofs)
- Policy enforcement (Intel SGX, FUSE)
- Cloud-based service for clients
- Host sensitive datasets (Sports, Fisheries Crime)

Conclusion

Lohpi

- A distributed infrastructure to support compliant data sharing and analytics.
- Scalable across multiple ethics committees.
- Cloud-based or local infrastructure.

References

1. Shneiderman, Ben. "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10.4 (2020): 1-31.
2. Bongartz, H., Rübsamen, N., Raupach-Rosin, H., Akmatov, M. K., & Mikolajczyk, R. T. (2017). Why do people participate in health-related studies?. *International journal of public health*, 62(9), 1059-1062.
3. Salerno, Jennifer, et al. "Ethics, big data and computing in epidemiology and public health." *Annals of Epidemiology* 27.5 (2017): 297-301.
4. Jenkins, Kate, Ken Hopkinson, and Ken Birman. "A gossip protocol for subgroup multicast." *Proceedings 21st International Conference on Distributed Computing Systems Workshops*. IEEE, 2001.
5. Goodman, Kenneth W., and Eric M. Meslin. "Ethics, information technology, and public health: duties and challenges in computational epidemiology." *Public Health Informatics and Information Systems*. Springer, London, 2014. 191-209.
6. Sharma, Aakash, et al. "Privacy Perceptions and Concerns in Image-Based Dietary Assessment Systems: Questionnaire-Based Study." *JMIR Human Factors* 7.4 (2020): e19085.
7. Sharma, Aakash, et al. "Up-to-the-minute Privacy Policies via gossips in Participatory Epidemiological Studies." *Frontiers in big Data* 4 (2021).
8. Johansen, Håvard D., et al. "Fireflies: A secure and scalable membership and gossip service." *ACM Transactions on Computer Systems (TOCS)* 33.2 (2015): 1-32.

LOHPI[®] Team



Thomas
Bye Nilsen



Aakash
Sharma



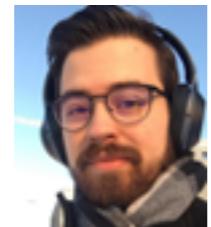
Dag
Johansen



Håvard D.
Johansen



Sivert
Johansen



Joakim A.
Alslie



Øyvind A. M.
Nohr



Lars K.
Kristiansen



Iris H.
Amundsen



Marius J.
Ingebrigsten



Sindre S.
Bjørvik



UiT The Arctic
University of Norway

Lab → NN

SpareBank
NORD-NORGE 1