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

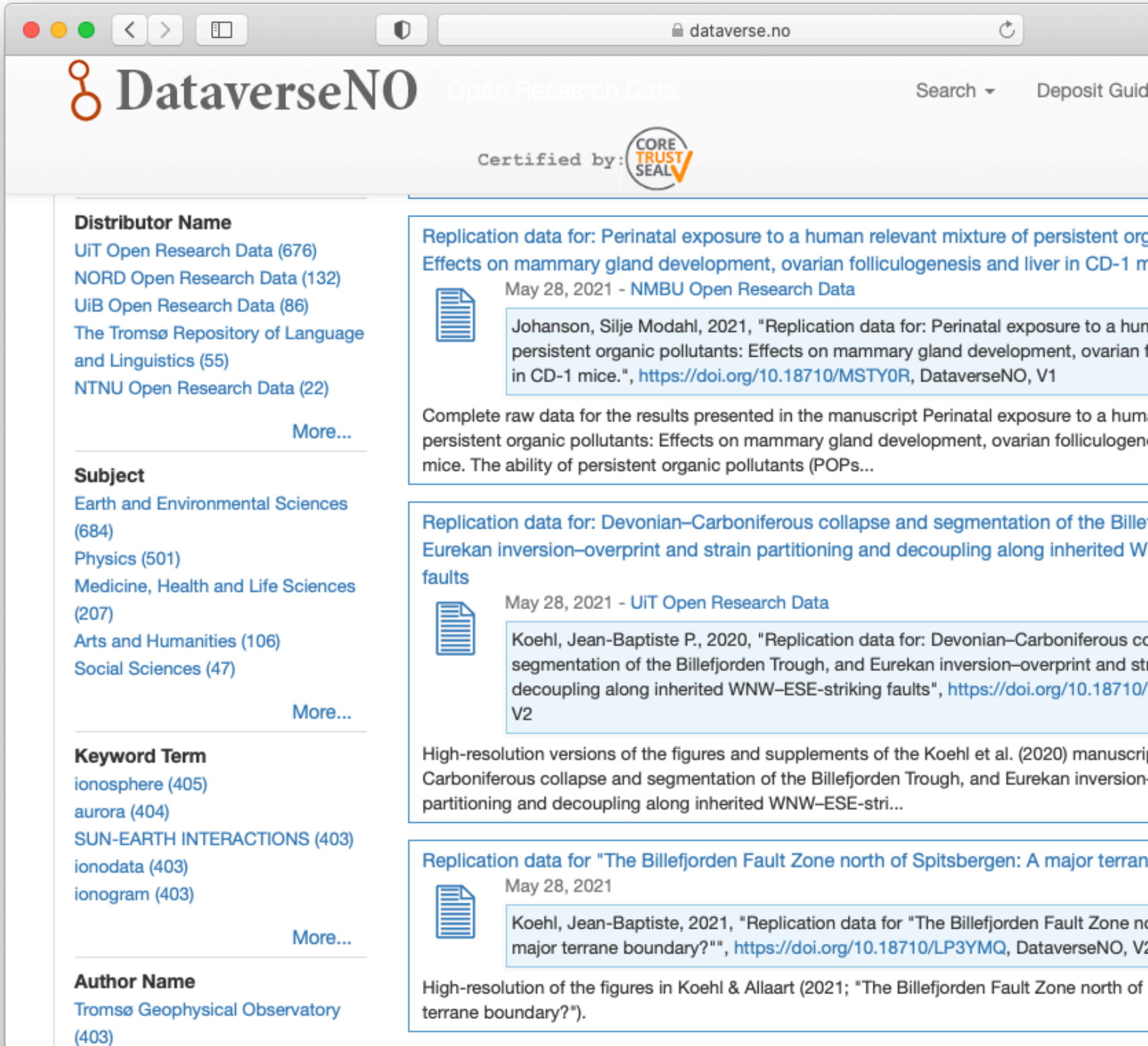
Aakash Sharma

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



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

Dato for utfylling

Dag mnd år

Kjente sykdommer:

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

	Kun diett	Tablett	Insulin
Hvis ja: nåværende behandling……	☐	☐	☐

Diagnoseår: ☐☐ Antall år på insulin: ☐☐

Røyking:

	Ja	Nei
Har du noen gang vært daglig røyker……	☐	☐
Har du noen gang vært daglig sigarillorøyker……	☐	☐
Har du noen gang vært daglig piperøyker……	☐	☐

☐ Jeg har aldri vært daglig røyker

Hvis du er eller har vært daglig røyker:

Alder ved røykestart:…… ☐☐ Alder i år

Alder ved siste røyking:…… ☐☐ Alder i år

Antall år med røyking:…… ☐☐ År

Antall sigaretter (eller tilsvarende) per dag:…… ☐☐ Antall

1 sigar = 4 sigaretter, 1 sigarillo =2 sigaretter, 1 pipe = 2,5 sigaretter

Familieanamnese:

Antall søsken (inkludert deg selv):…… ☐☐

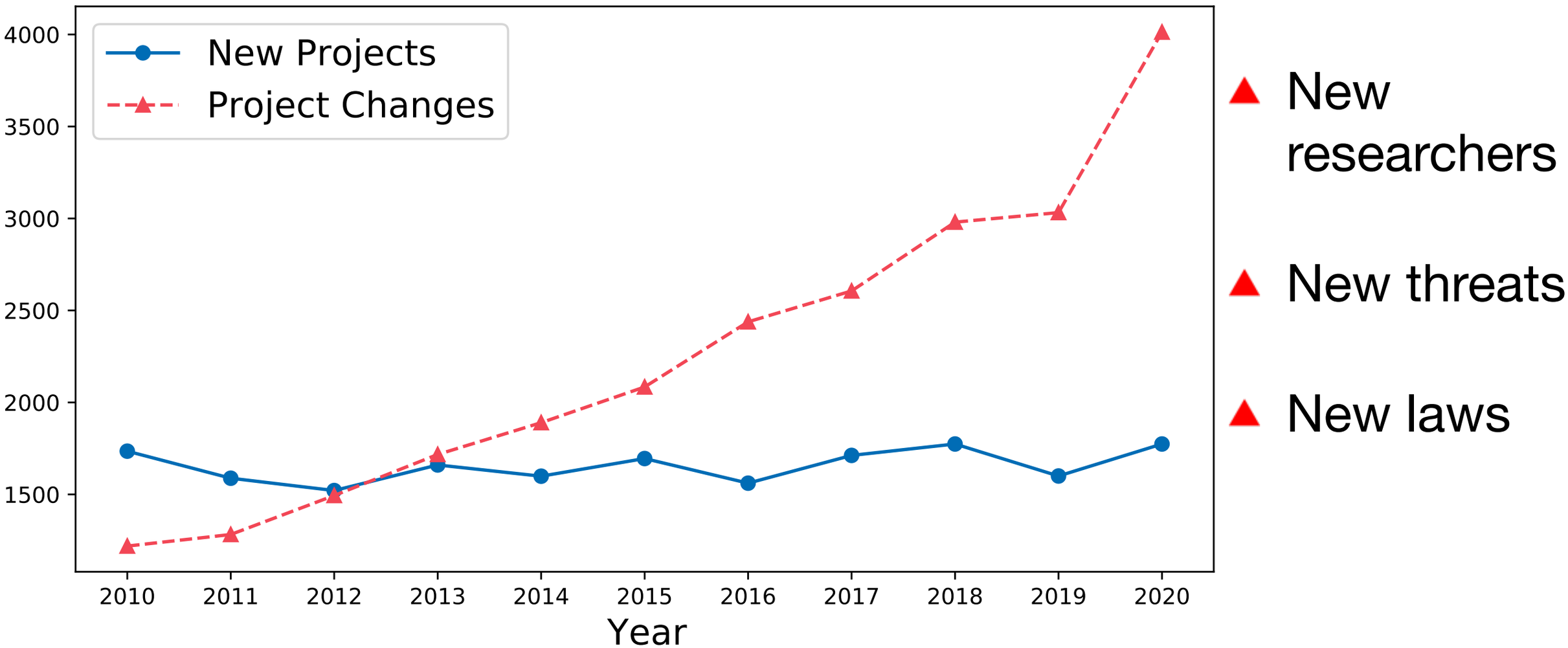
Antall barn:…… ☐☐

Antall førstegradsslektninger med tykk- eller endetarmskreft:…… ☐☐

Antall førstegradsslektninger med tykk- eller endetarmskreft før fylte 60 år: ☐☐

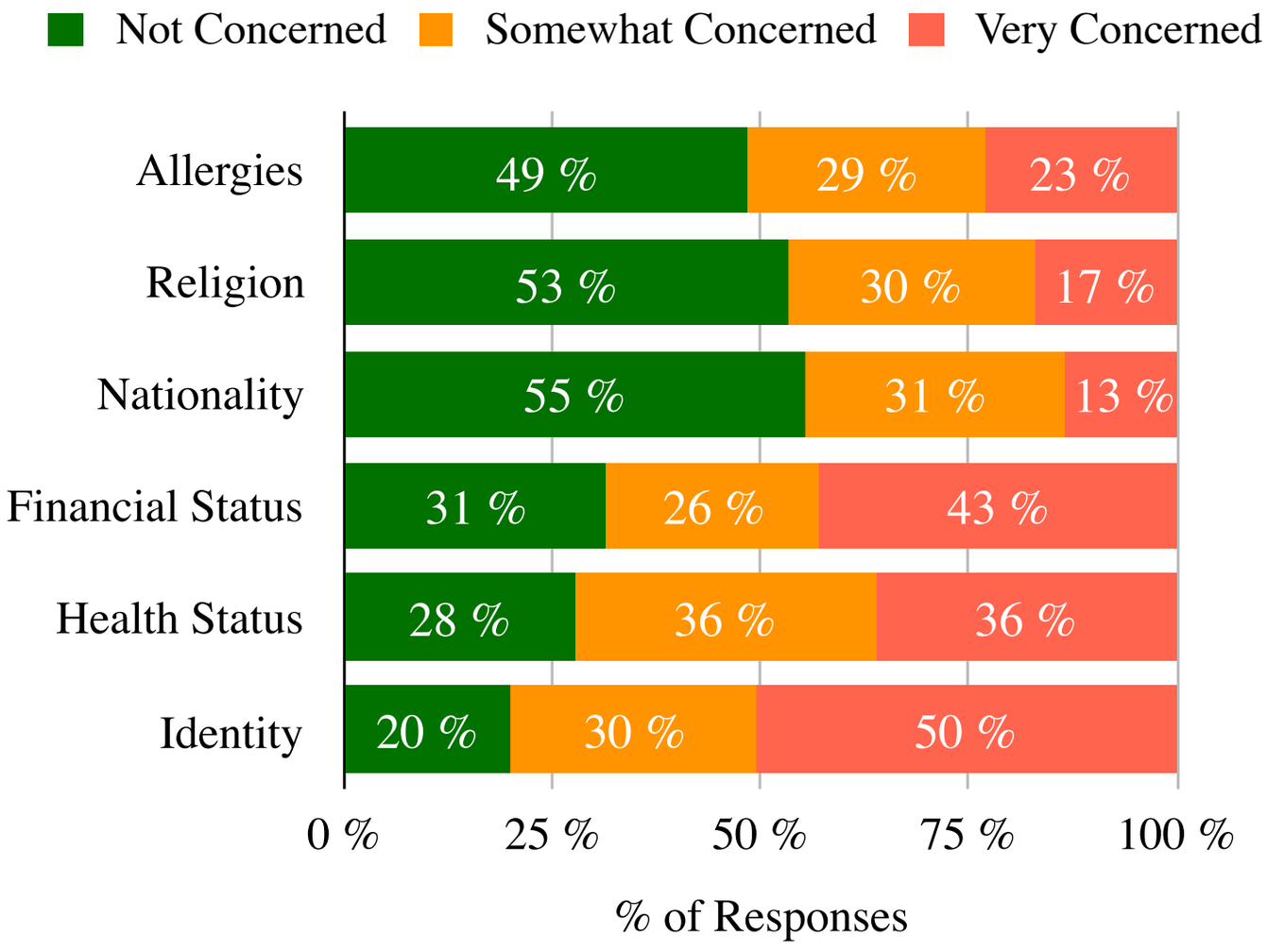
CONFIDENTIAL

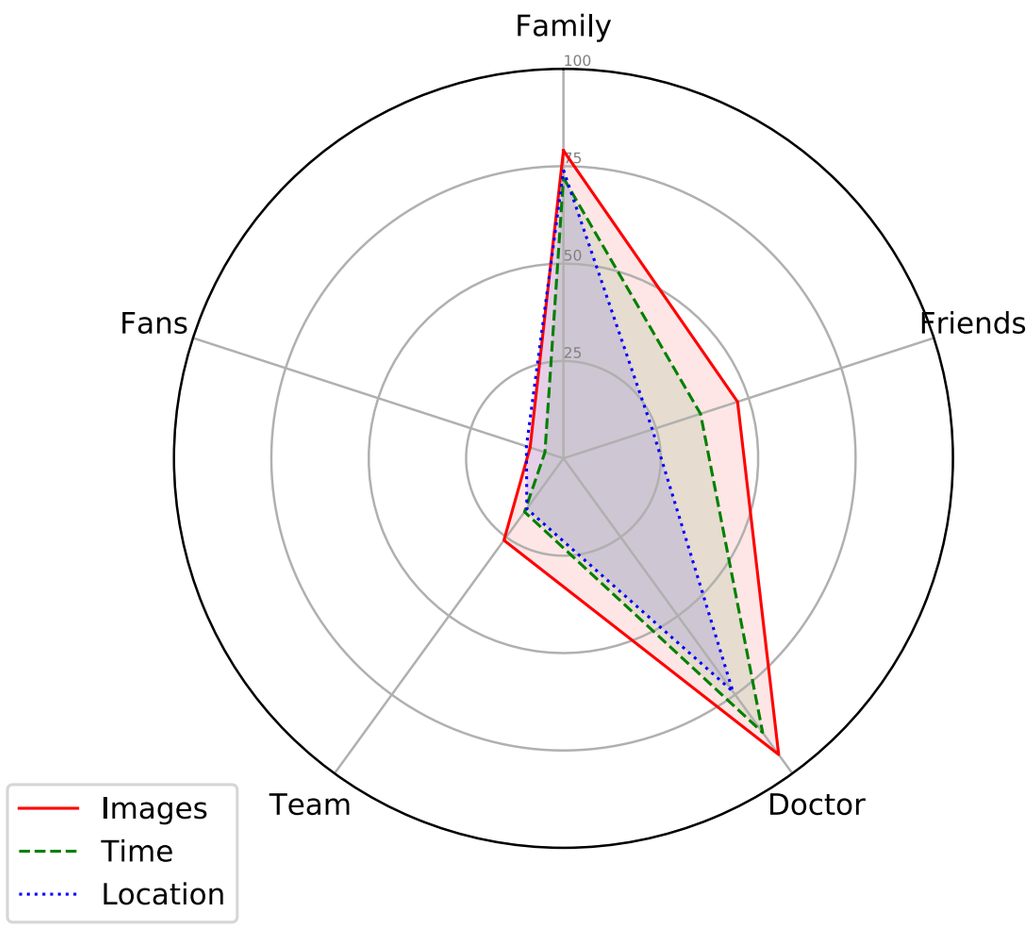
Projects are not static



https://rekportalen.no

Opinions change





Sharma, Aakash, et al. "Privacy Perceptions and Concerns in Image-Based Dietary Assessment Systems: Questionnaire-Based Study." JMIR Human Factors 7.4 (2020): e19085.

A project's lifecycle

Proposal Approval Collection Analysis Dissemination

A project's lifecycle

Proposal Approval Collection Analysis Dissemination

A project's lifecycle

Proposal **Approval** Collection Analysis Dissemination

A project's lifecycle

Proposal Approval Collection Analysis Dissemination

A project's lifecycle

Proposal Approval **Collection** Analysis Dissemination

A project's lifecycle

Proposal Approval Collection Analysis Dissemination

A project's lifecycle

Proposal Approval Collection **Analysis** Dissemination

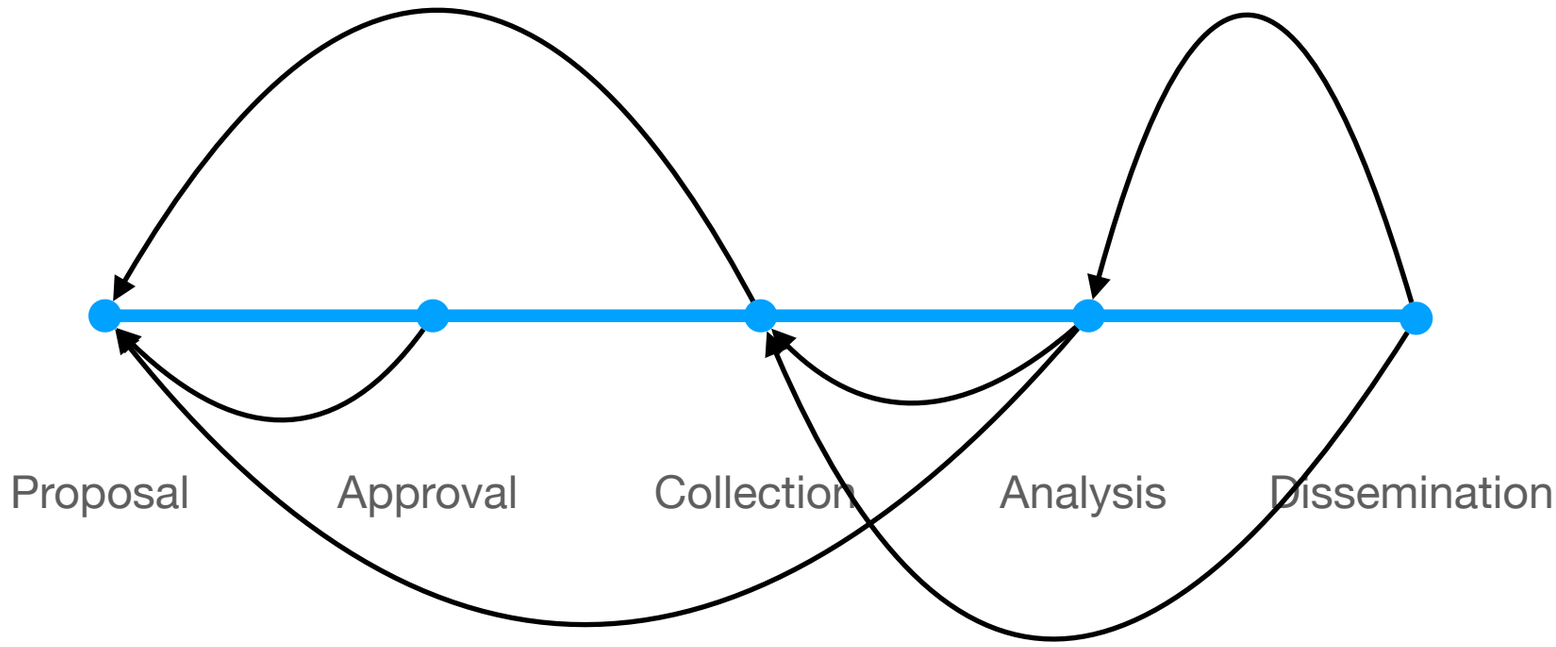
A project's lifecycle

Proposal Approval Collection Analysis Dissemination

A project's lifecycle

Proposal Approval Collection Analysis **Dissemination**

A project's lifecycle is non-trivial



We need system support for dynamic security policies

Sensitive data leads to silos

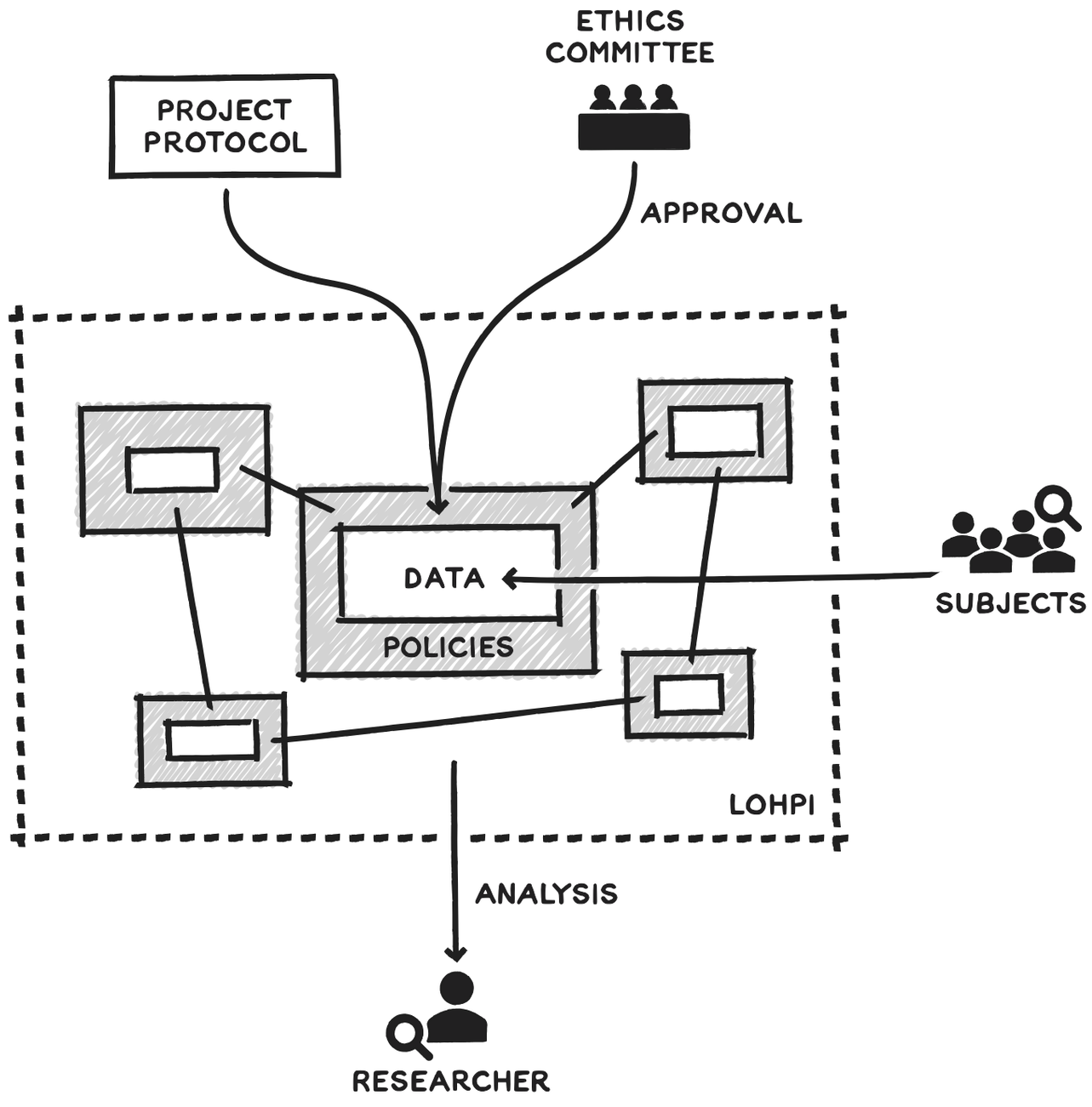
Tension between

- 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**.



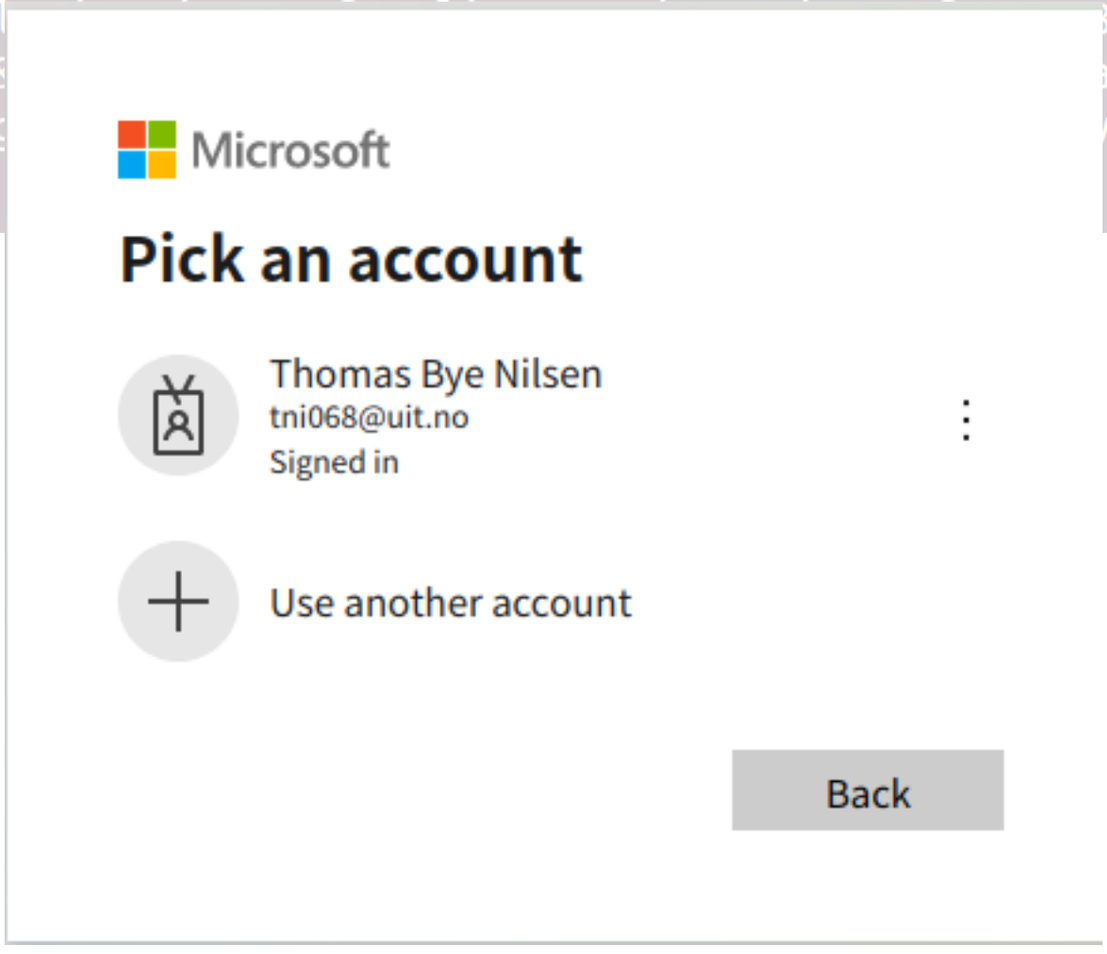
LOHPI

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=htt
_challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain
```

Lohpi workflow

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh login
Login link 🌍:
https://l                                                    81-bcc0-4721d4387973/oauth2/v2.0/authorize?&state
de=query&                                                    aph.microsoft.com%2FUser.read%20&redirect_uri=htt
_challeng                                                    vuiop&code_challenge_method=plain
```



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=htt
_challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain
```

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
{
  "Identifiers": [
    "Data.bin",
    "Fish_Dataset.zip",
    "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=htt
_challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain
```

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
{
    "Identifiers": [
        "Data.bin",
        "Fish_Dataset.zip",
        "notMNIST.zip"
    ]
}
```

```
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=htt
_challenge=1qaz2wsx3edc4rfv5tgb6yhn1234567890qwertyuiop&code_challenge_method=plain
```

Pick an account

Thomas Bye Nilsen

Signed in

Use another account

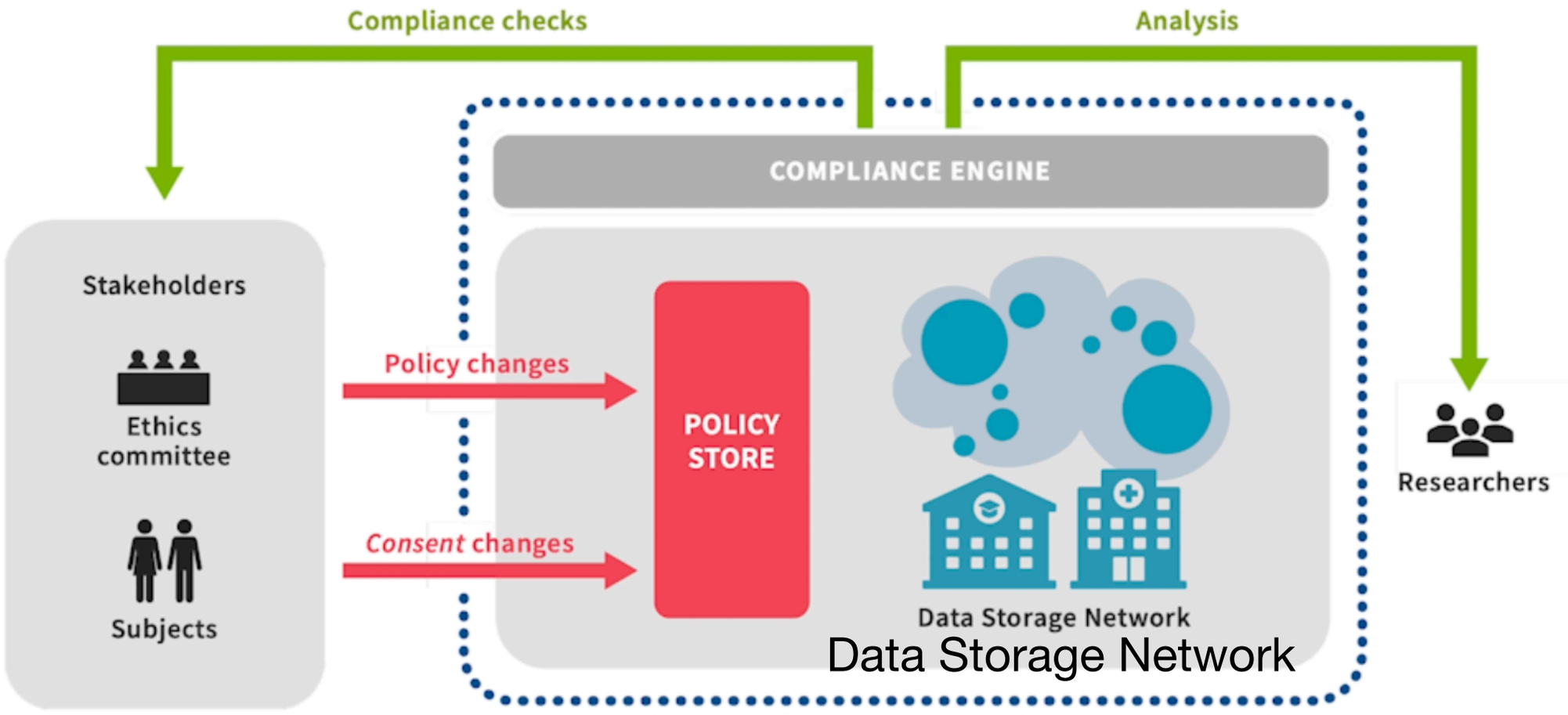
Back

```
thomas@thomas-ThinkCentre-M900 ~/D/r/l/t/debug (main)> ./lh list
{
    "Identifiers": [
        "Data.bin",
        "Fish_Dataset.zip",
        "notMNIST.zip"
    ]
}
```

```
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

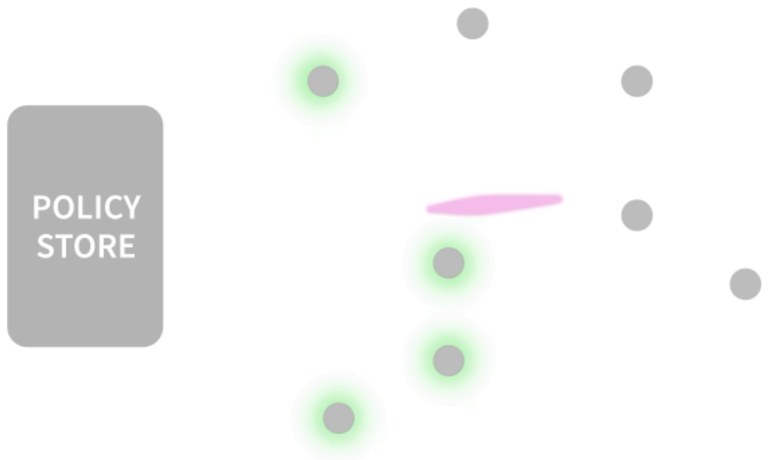
Jenkins, Kate, Ken Hopkinson, and Ken Birman. "A gossip protocol for subgroup multicast." Proceedings 21st International Conference on Distributed Computing Systems Workshops. IEEE, 2001.

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.

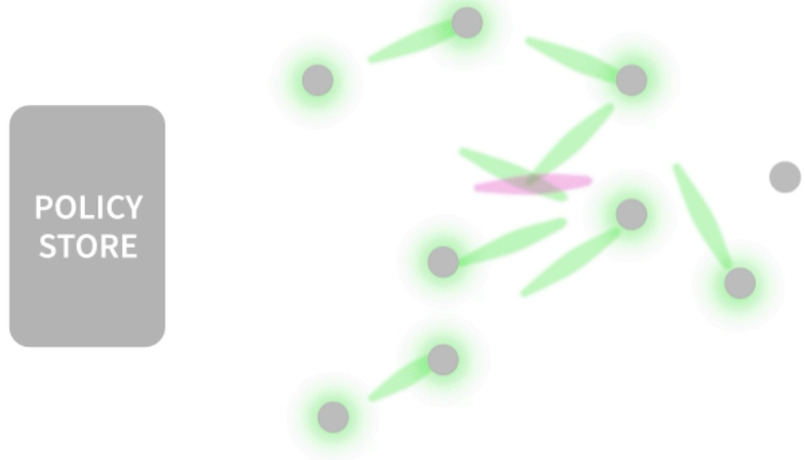
Remove bottleneck

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…
})…

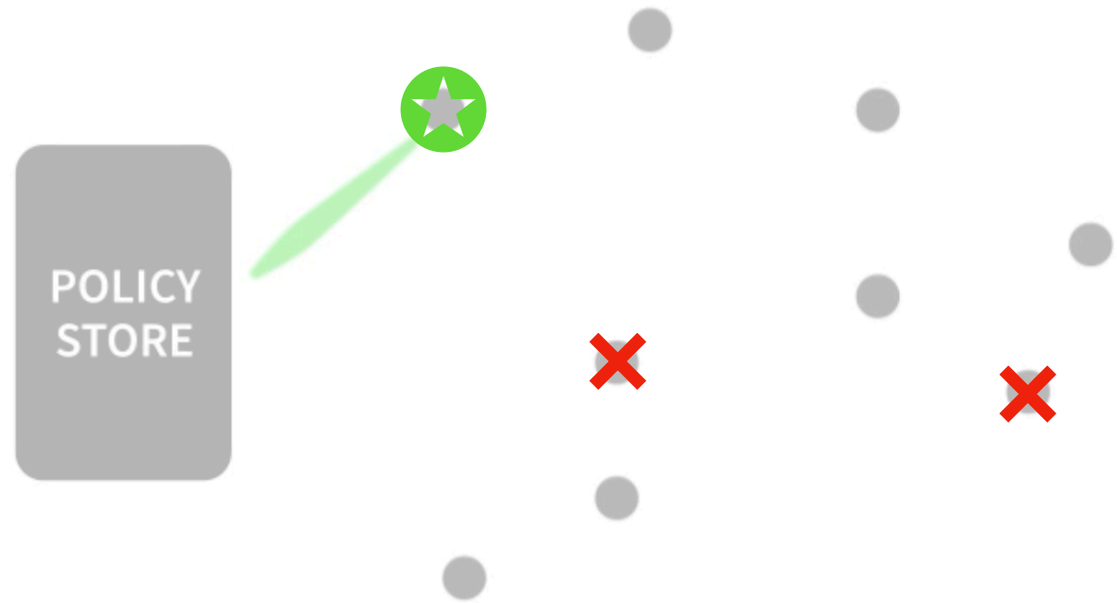
Experimental Setup

N Number of nodes

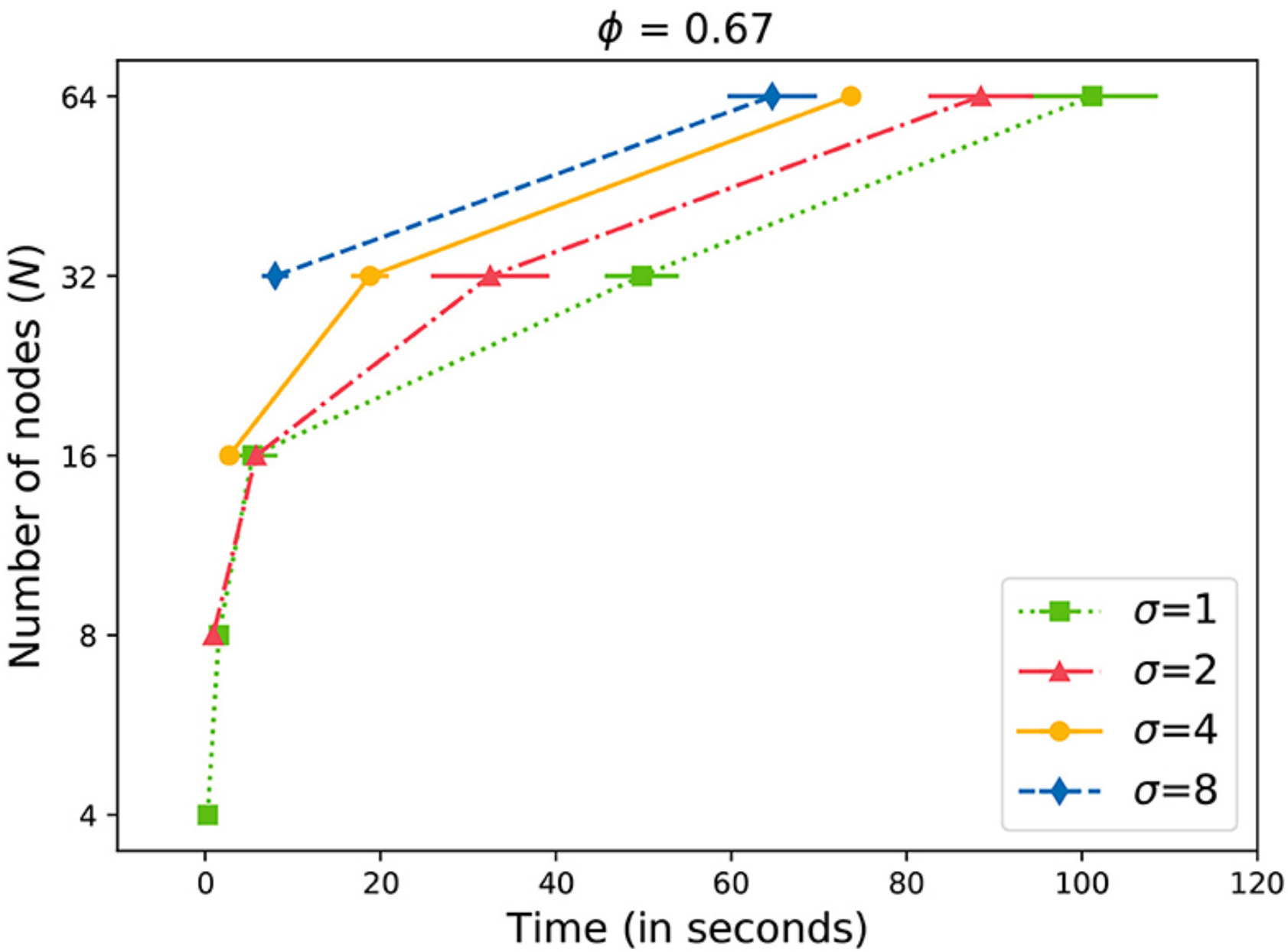
ϕ Percent of nodes that must receive a gossip

σ Multicast node

ϵ Failures

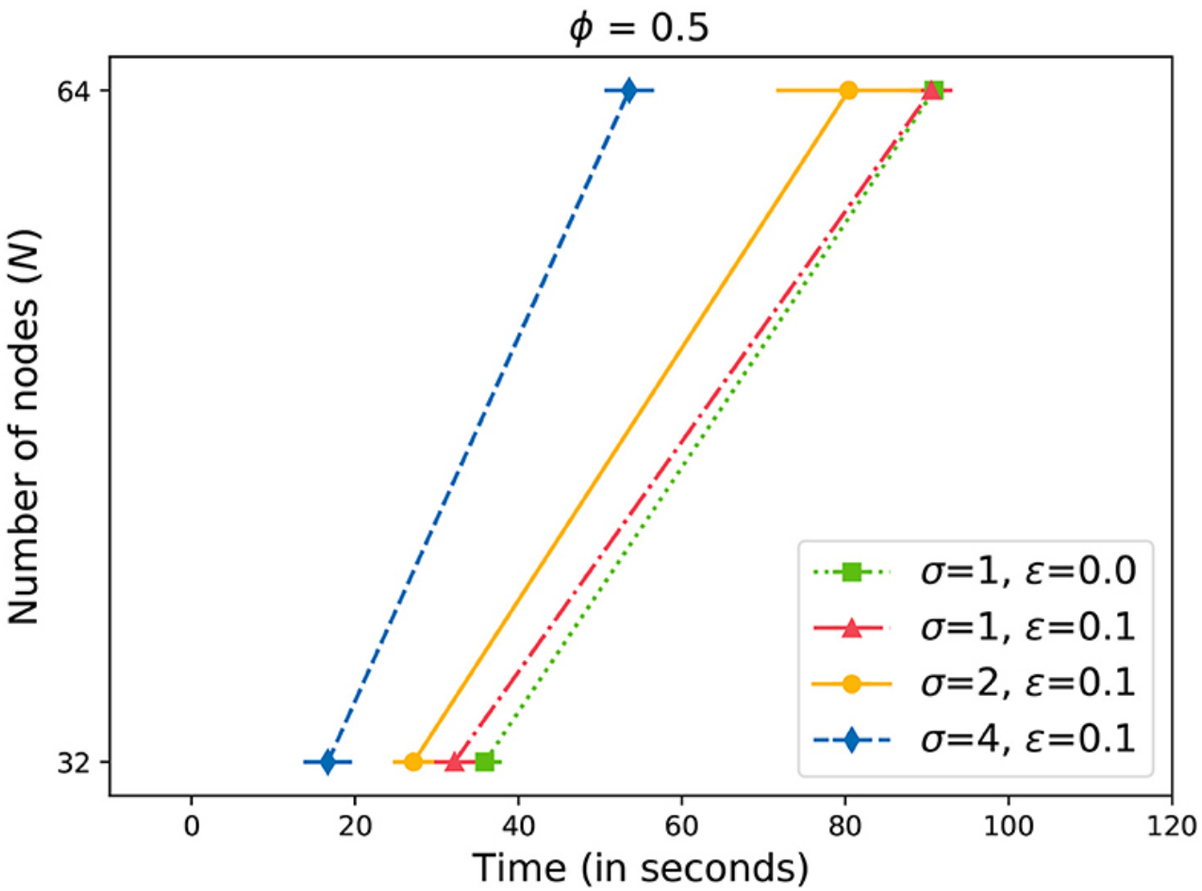


Dissemination

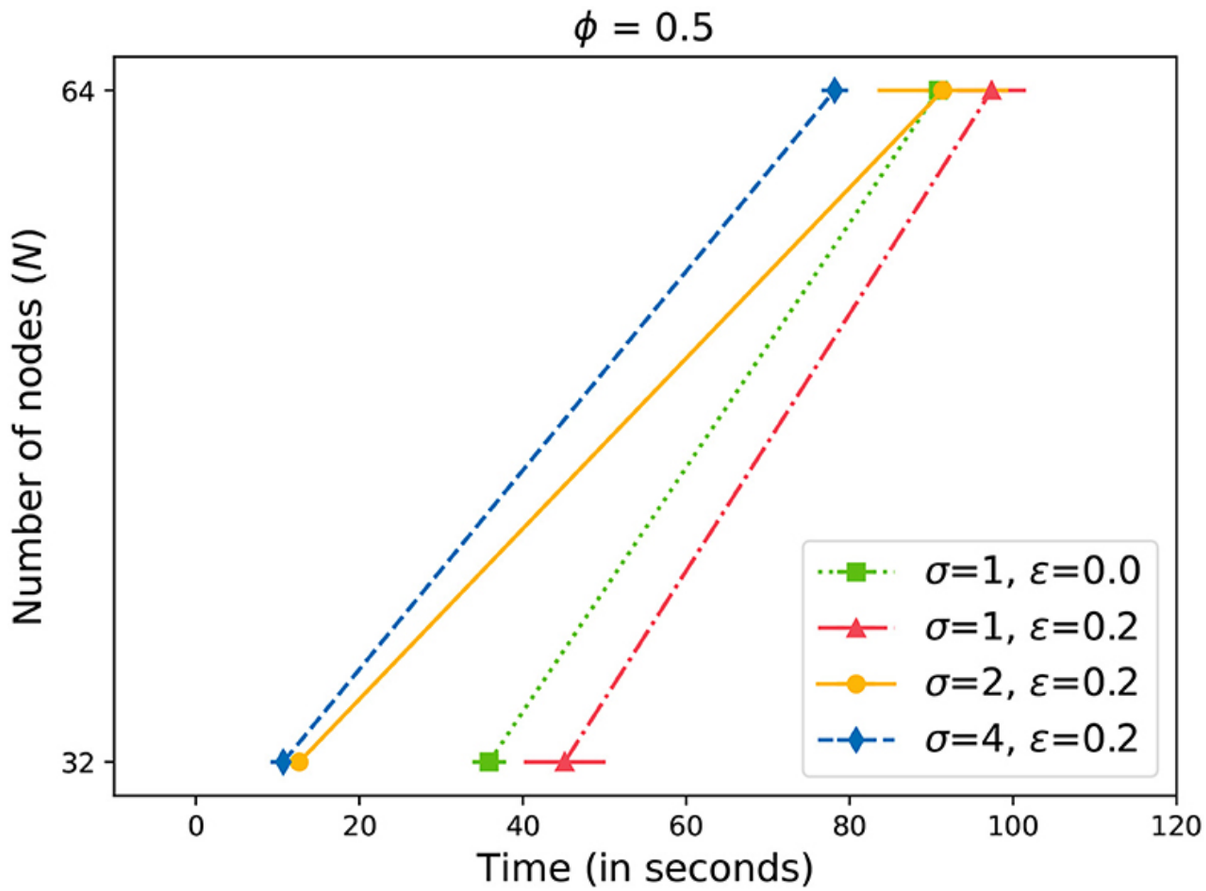


Gossip Interval = 3s

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.

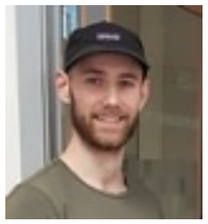
LOHPI Team

Thomas Bye Nilsen

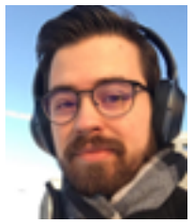
Aakash Sharma

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 1 NORD-NORGE