Compliant Sharing of Sensitive Data with Dataverse and Lohpi

Aakash Sharma, Thomas Bye Nilsen, Håvard D. Johansen
UiT The Arctic University of Norway

Dataverse Community Meeting 2021

LOHPI Team

Thomas Bye Nilsen

Aakash Sharma

Dag Johansen

Håvard D. Johansen

UiT The Arctic University of Norway

Lab→NN

SpareBank 1 NORD-NORGE

Lohpi logo designed by Katja Pauline Czerwinska

Dataverse Community Meeting 2021

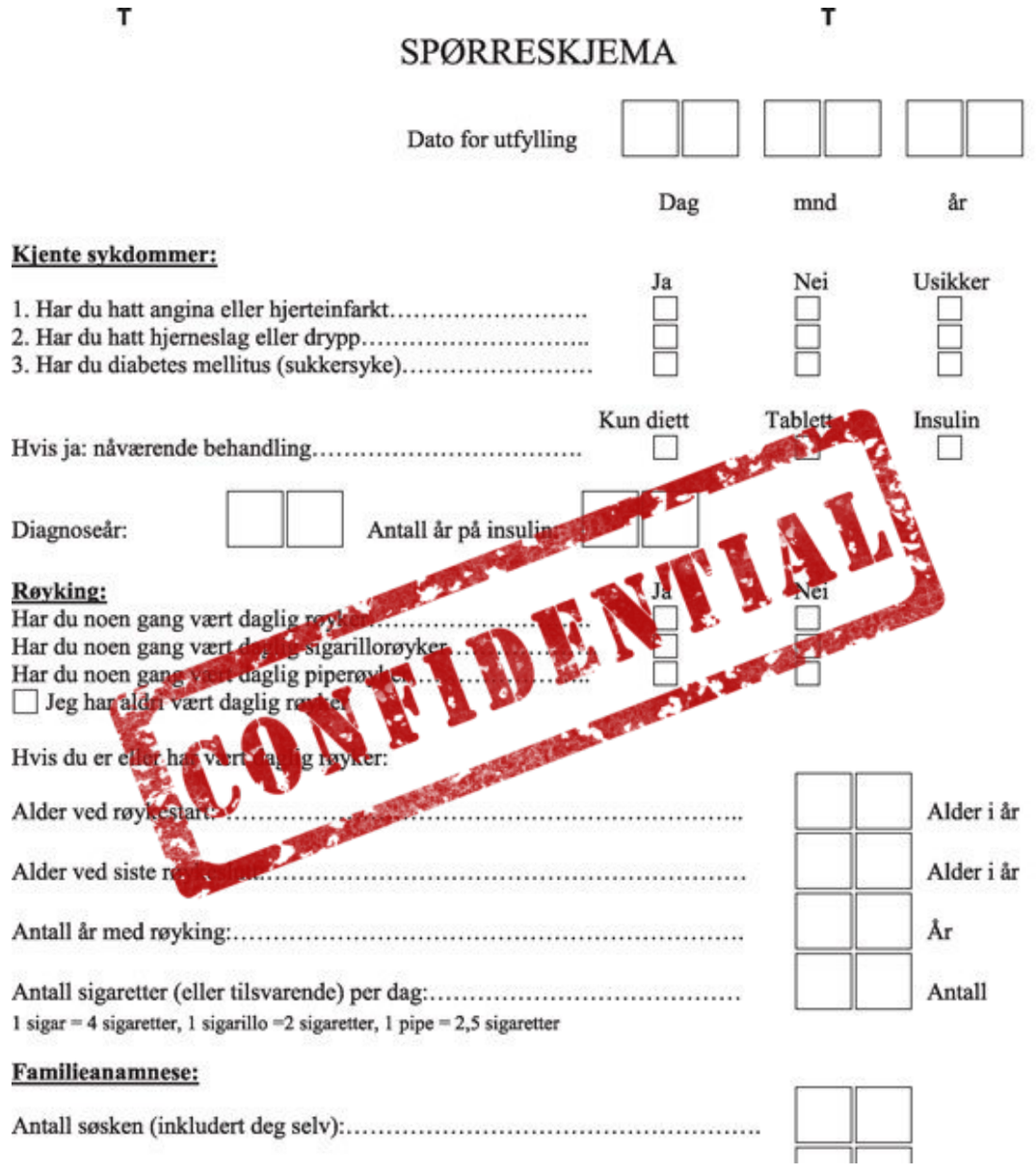
Sharing data is the norm

- 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)


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)……	☐	☐	☐

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

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 røyk start:…….. ☐☐ Alder i år

Alder ved siste r……. ☐☐ Alder i år

Antall år med røyking:…… ☐☐ År

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

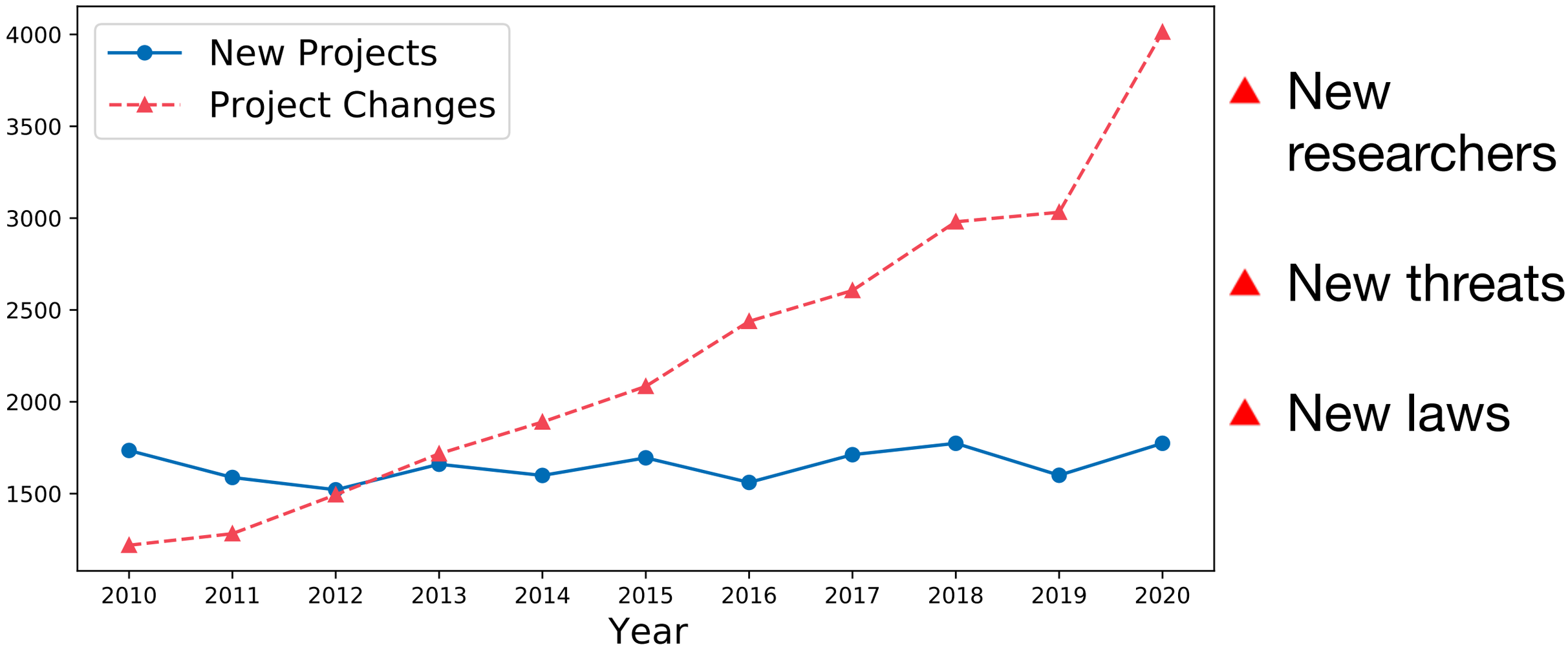
1 sigar = 4 sigaretter, 1 sigarillo =2 sigaretter, 1 pipe = 2,5 sigaretter

Familieanamnese:

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

CONFIDENTIAL

Projects are not static



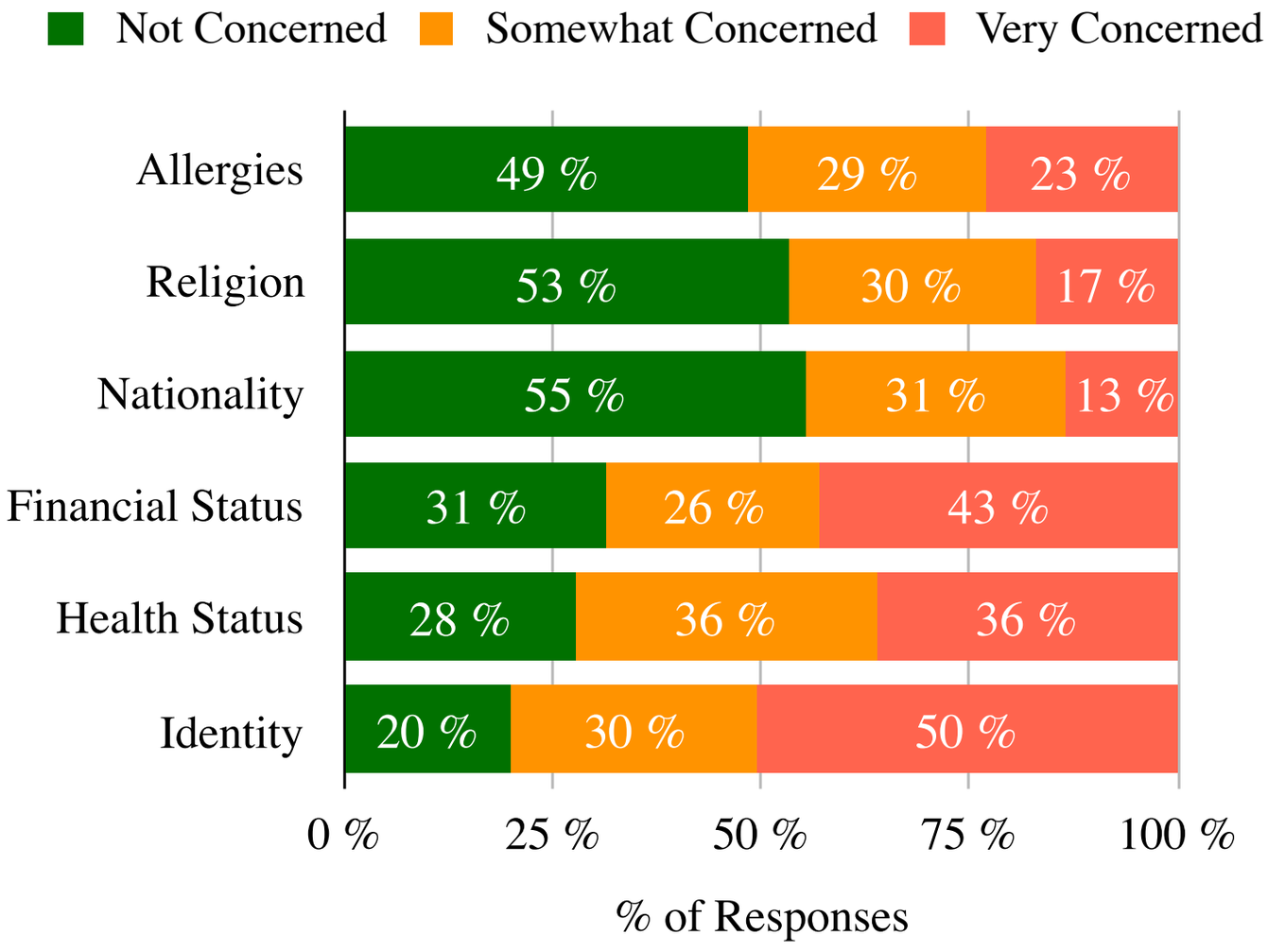
New researchers

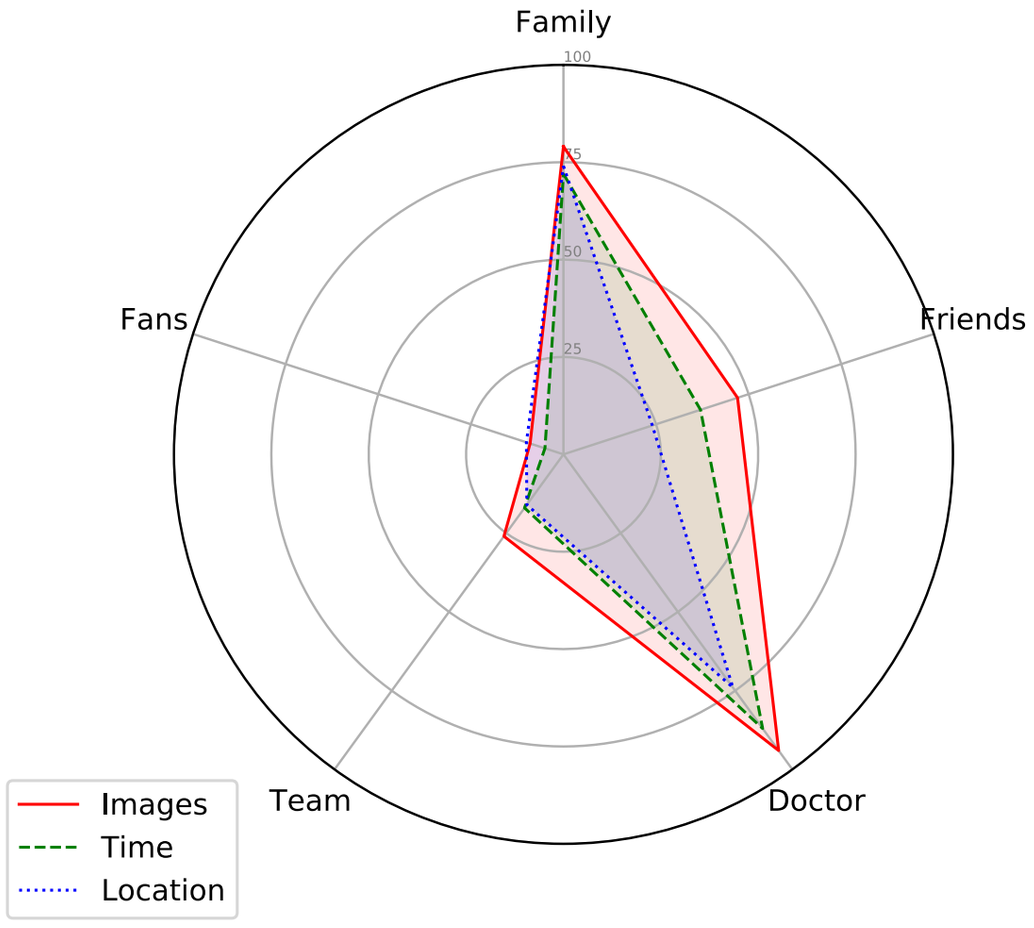
New threats

New laws

https://rekportalen.no

Opinions change



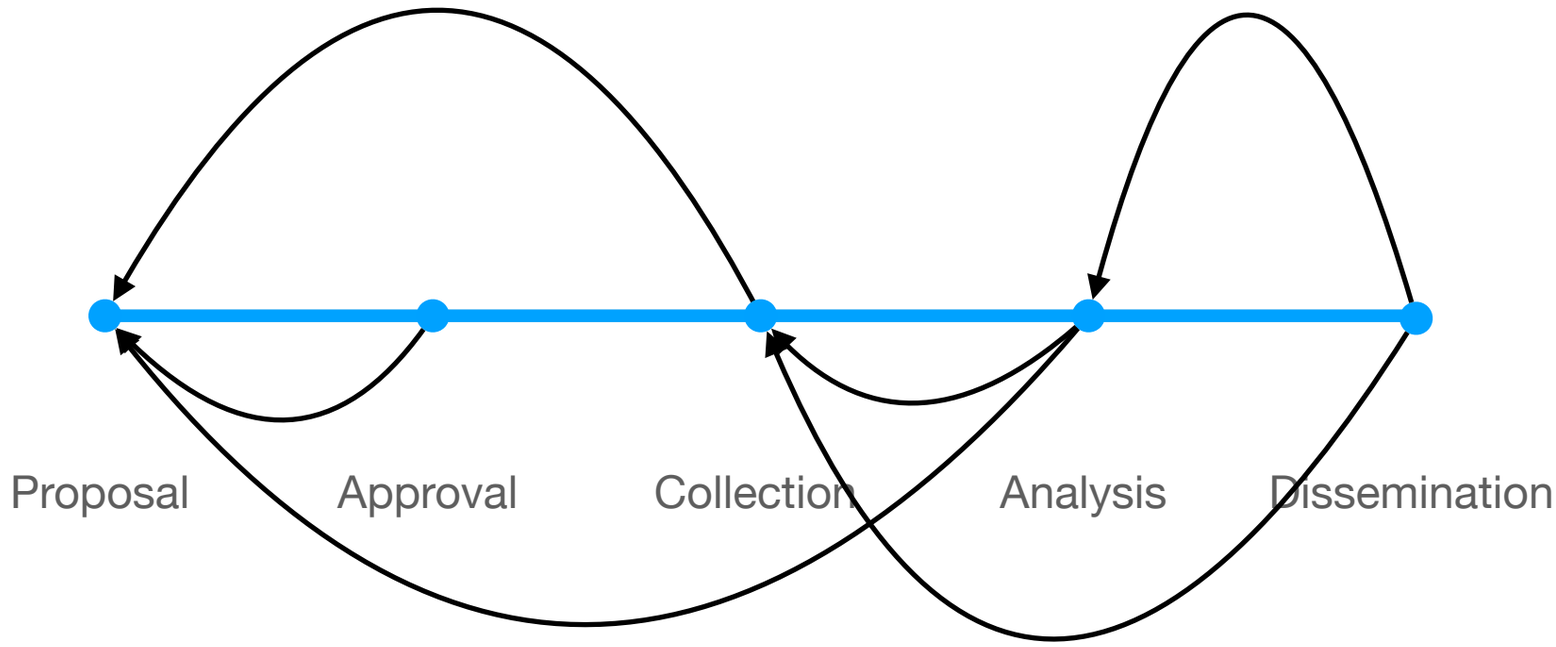


Sharma, Aakash, et al. "Privacy Perceptions and Concerns in Image-Based Dietary Assessment Systems: Questionnaire-Based Study." JMIR Human Factors 7.4 (2020): e19085.

A project's lifecycle

Proposal Approval Collection Analysis **Dissemination**

A project's lifecycle is non-trivial



We need system support for dynamic security policies

Sensitive data leads to silos

Tension between

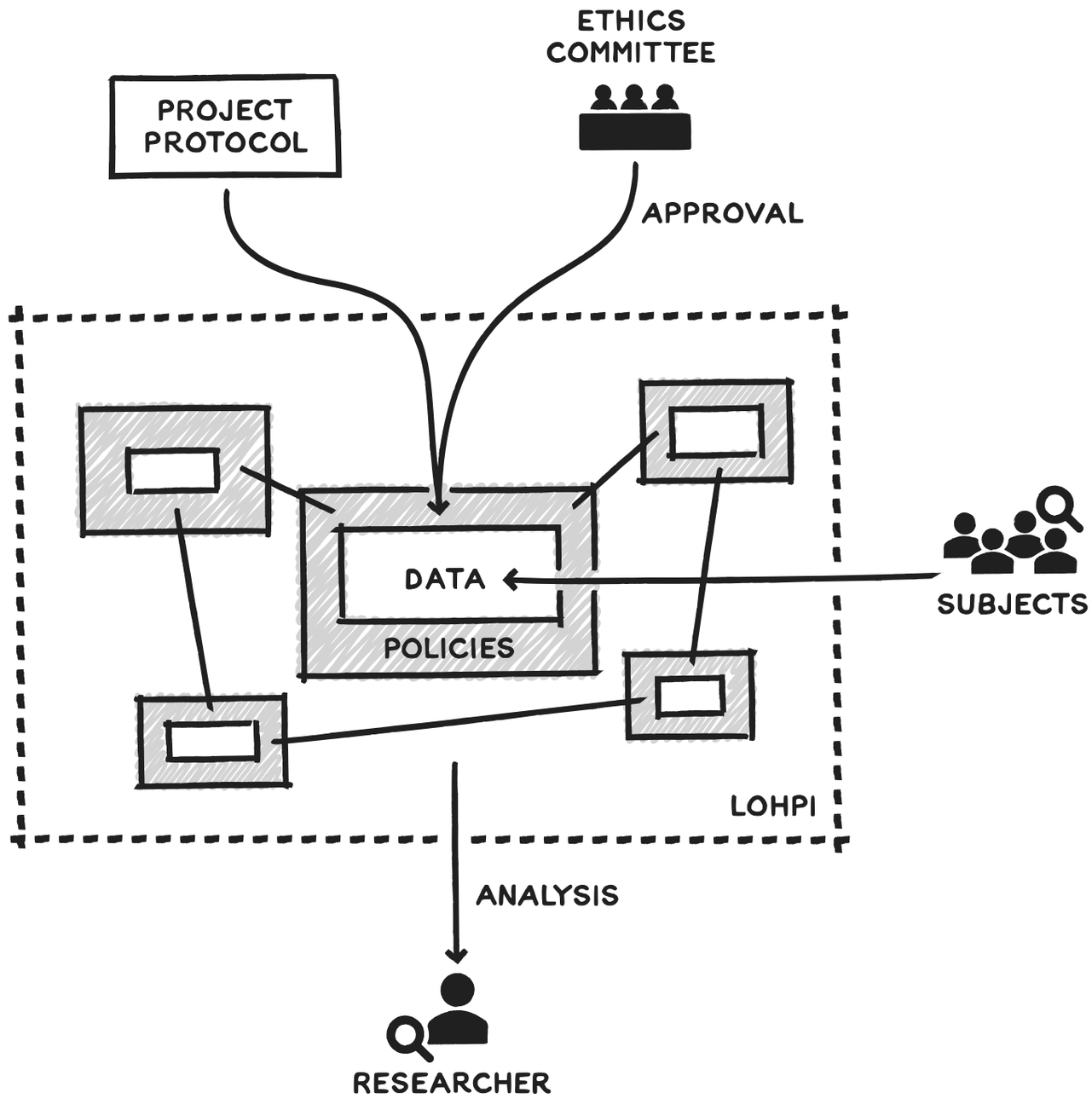
- Open research (Dataverse)
  - Decentralized research model
  - Operate on institutional infrastructure
- Privacy risks
  - Projects operate in silos
  - Shared computational infrastructure (trusted by all parties, TSD)

https://www.uio.no/english/services/it/research/sensitive-data/

Compliant sharing with Lohpi

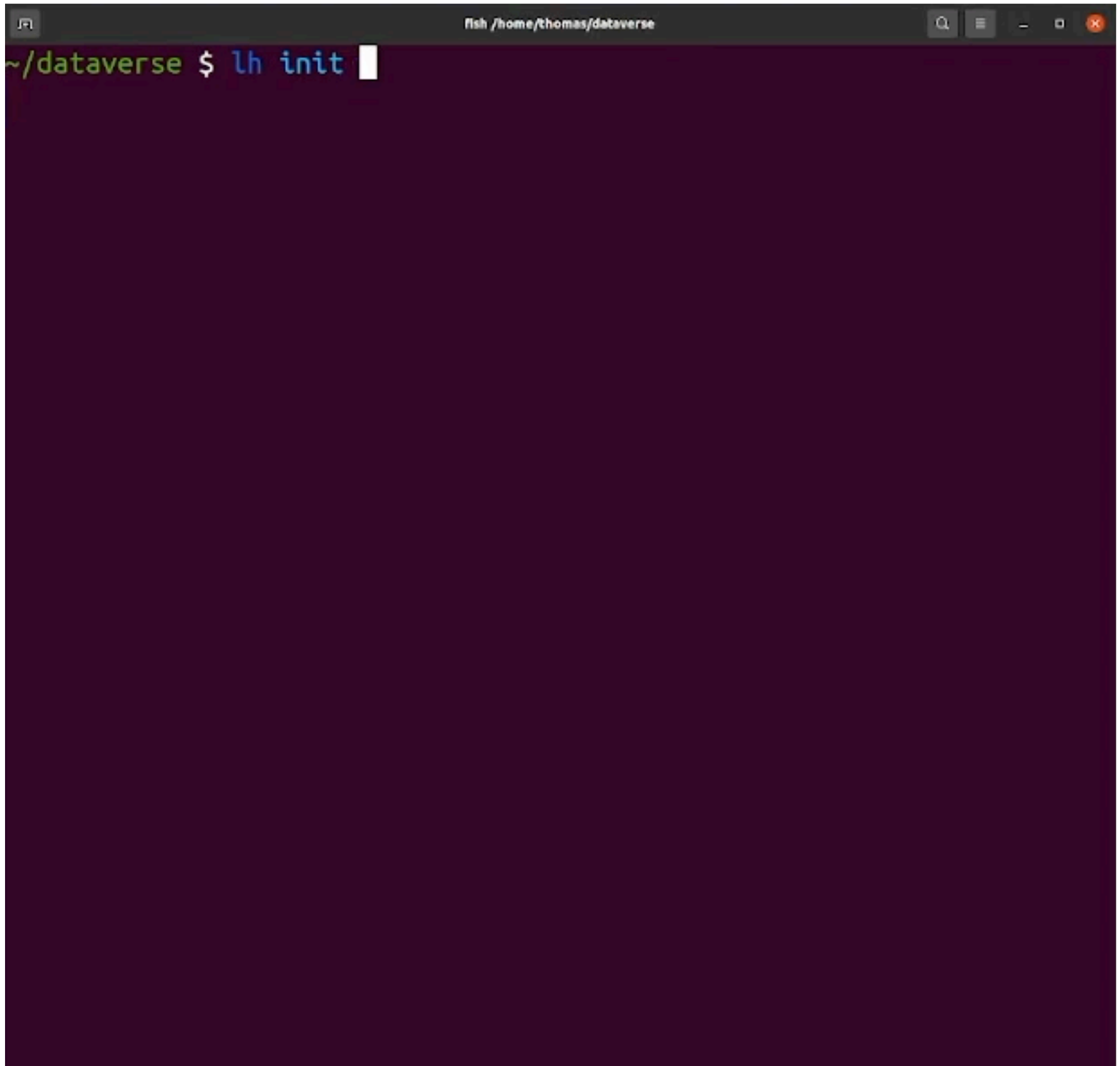
A distributed **metadata layer** that enables compliant data sharing.

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



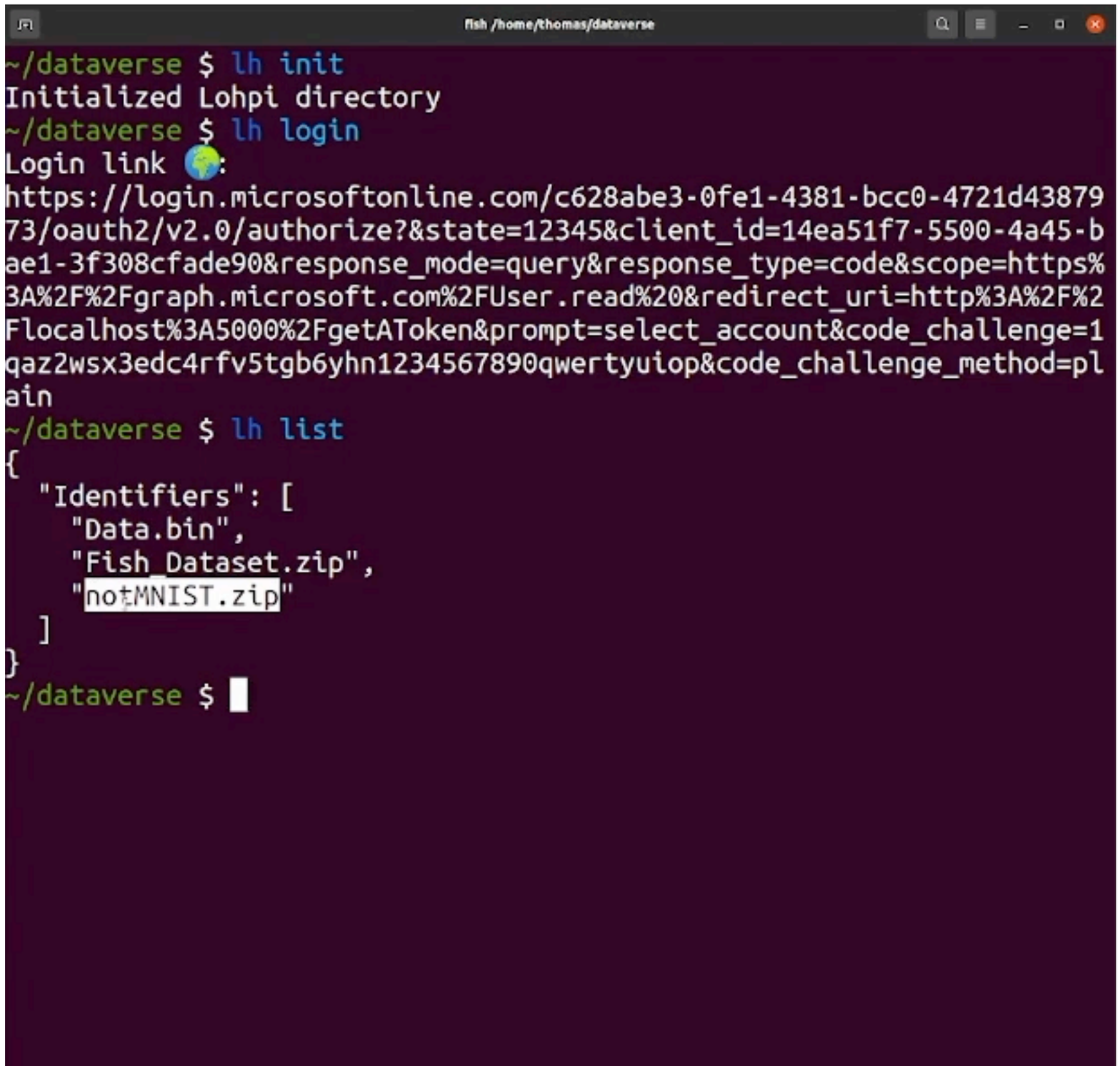
LOHPI

Lohpi workflow



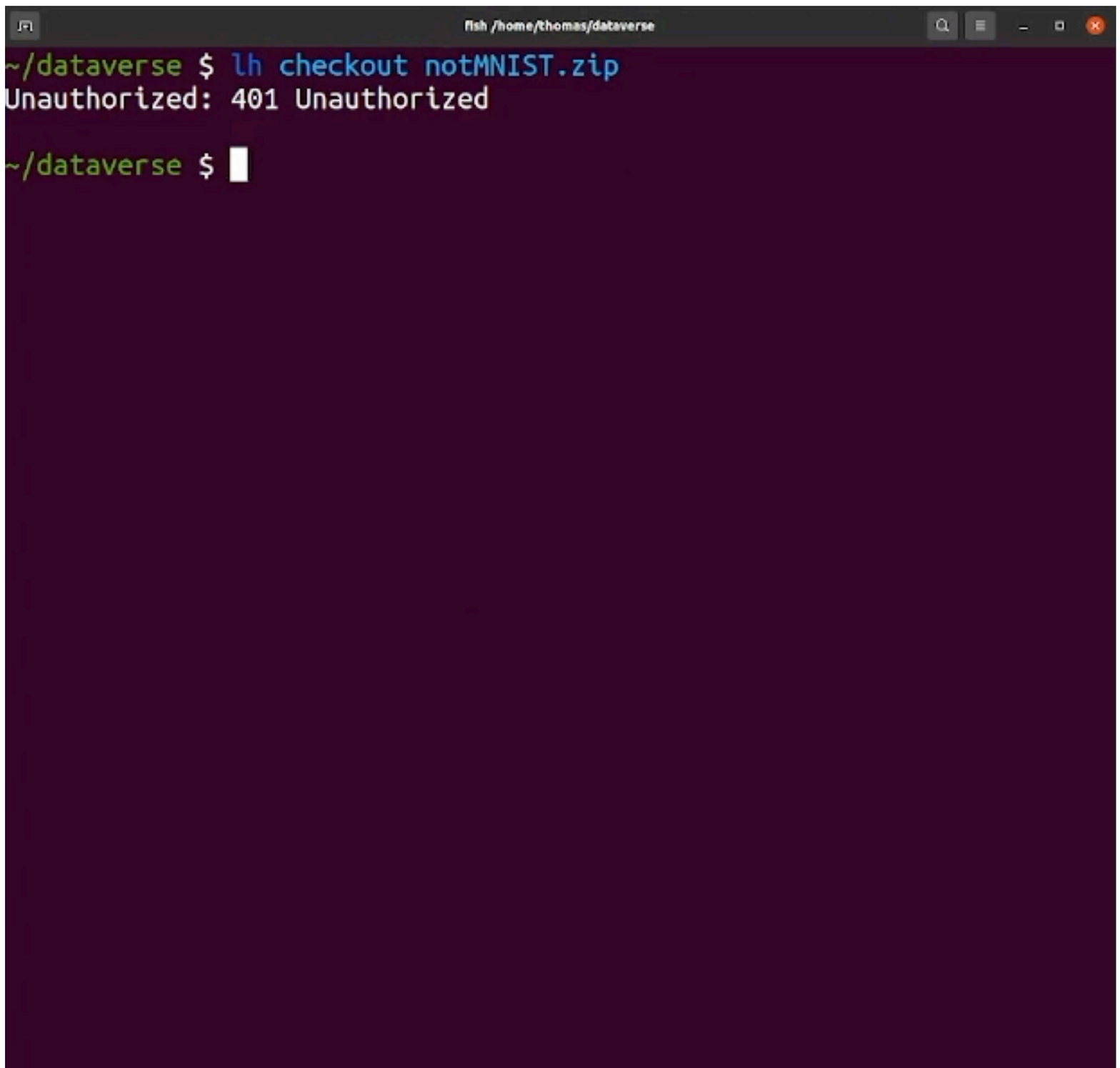
- Integrates with existing authentication services

Lohpi workflow



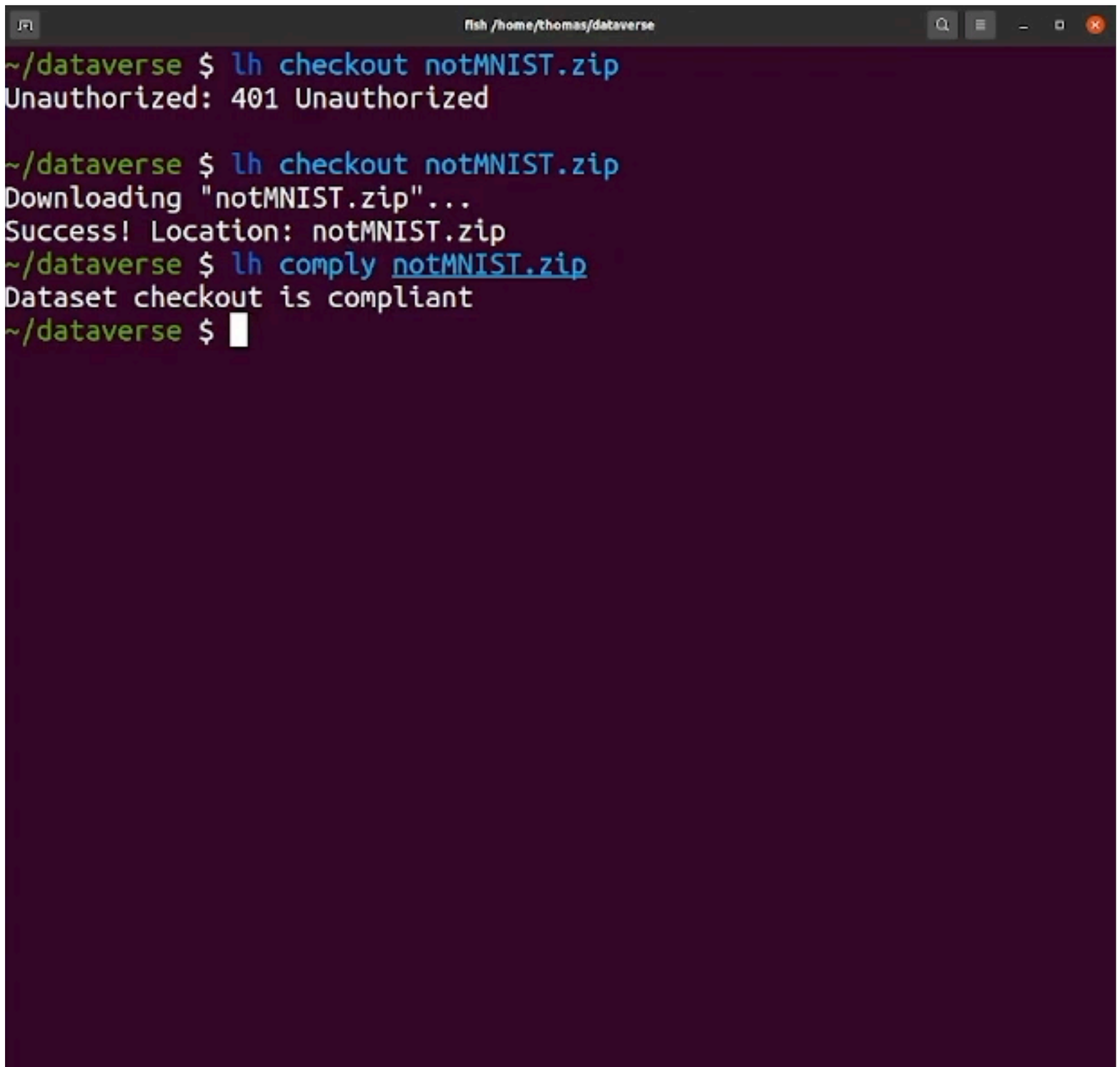
- Integrates with existing authentication services
- Public list of available datasets (discovery)

Lohpi workflow



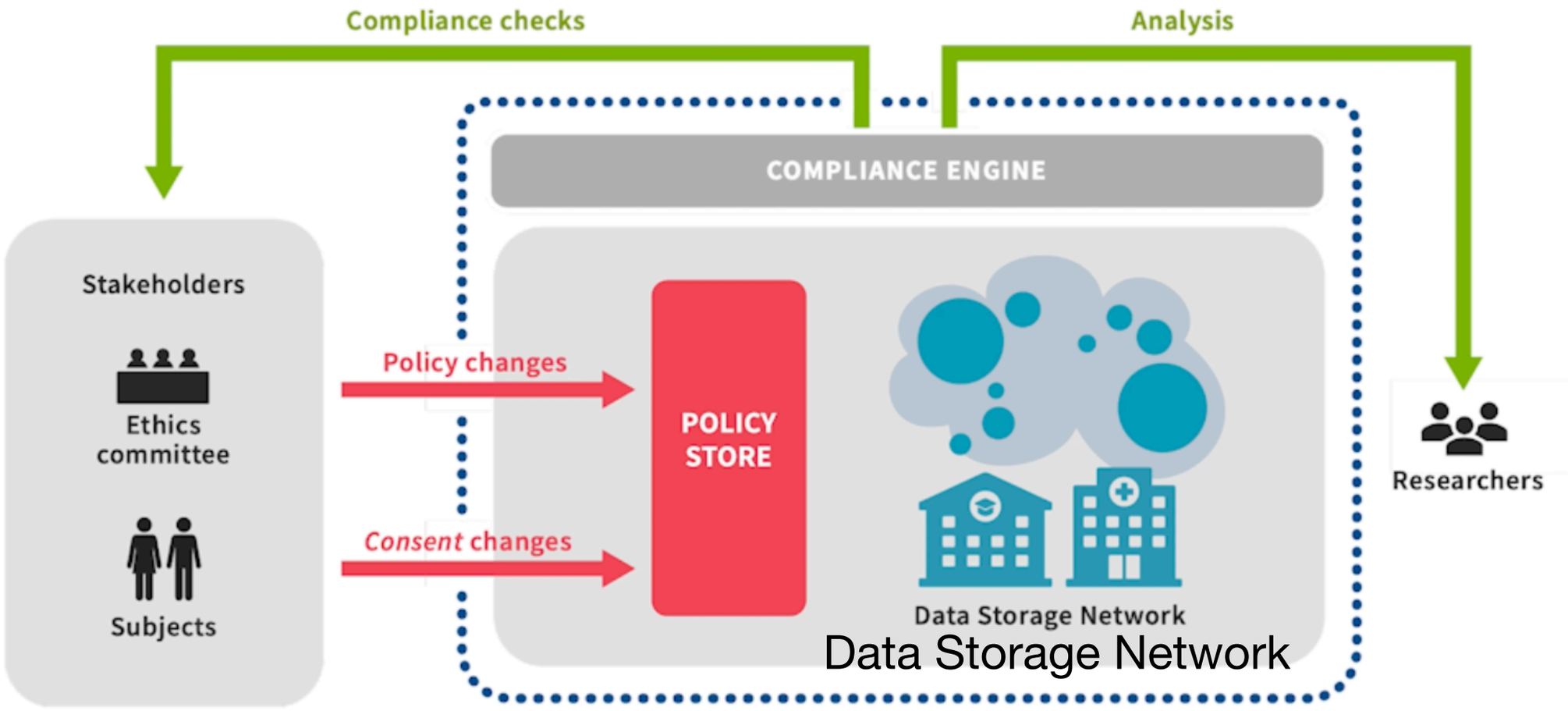
- Integrates with existing authentication services
- Public list of available datasets (discovery)
- Seamless data checkouts

Lohpi workflow



- Integrates with existing authentication services
- Public list of available datasets (discovery)
- Seamless data checkouts
- Up-to-date data security policies

Lohpi architecture



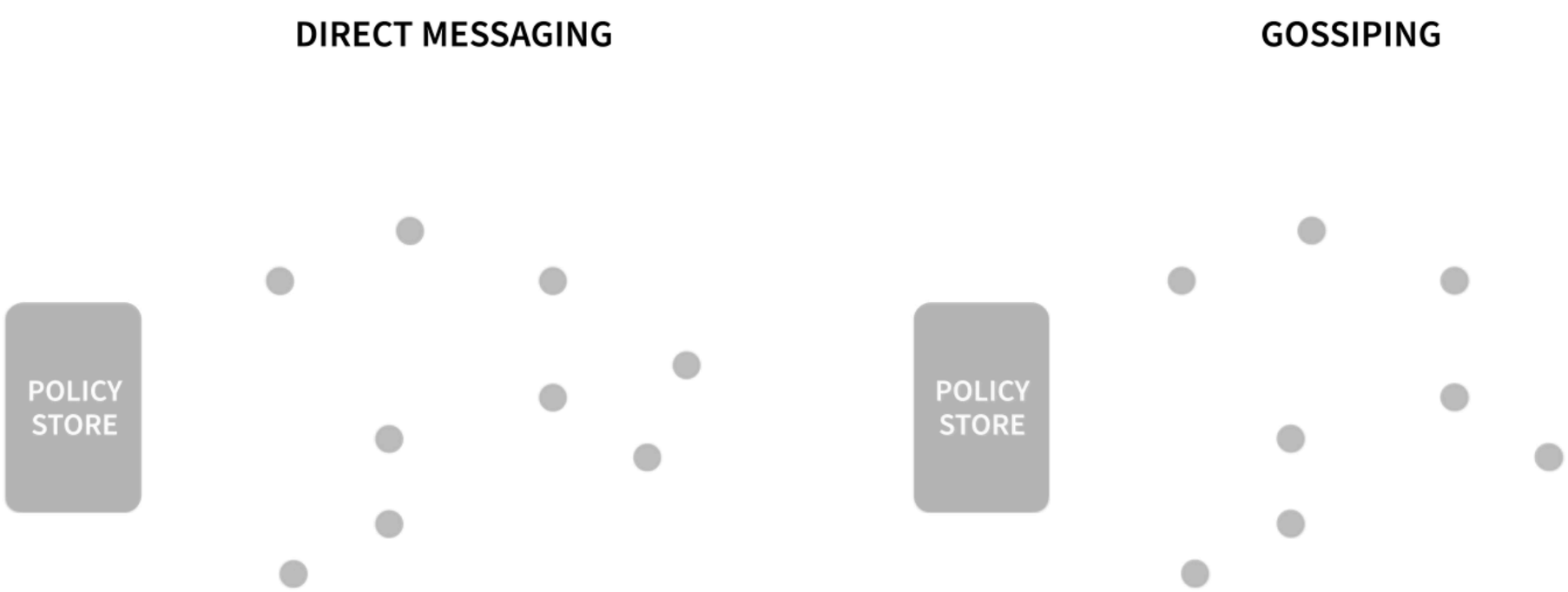
Sharma, Aakash, et al. "Up-to-the-minute Privacy Policies via gossips in Participatory Epidemiological Studies." Frontiers in big Data 4 (2021).

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.

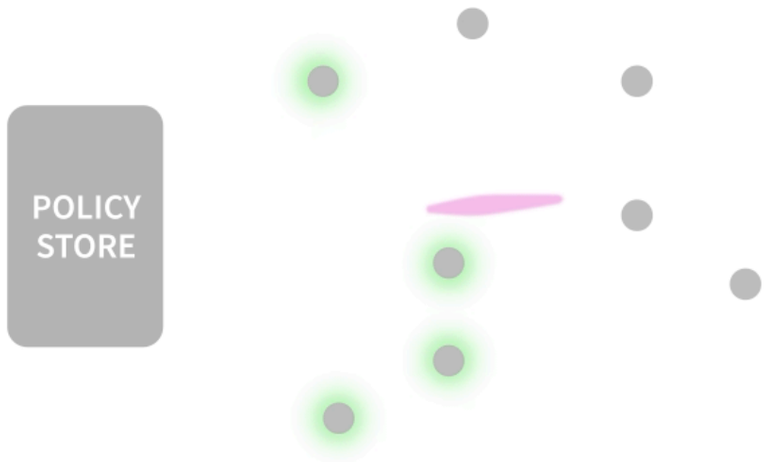
Johansen, H. D., Renesse, R. V., Vigfusson, Y., & Johansen, D. (2015). Fireflies: A secure and scalable membership and gossip service. ACM Transactions on Computer Systems (TOCS), 33(2), 1-32.

Remove bottleneck

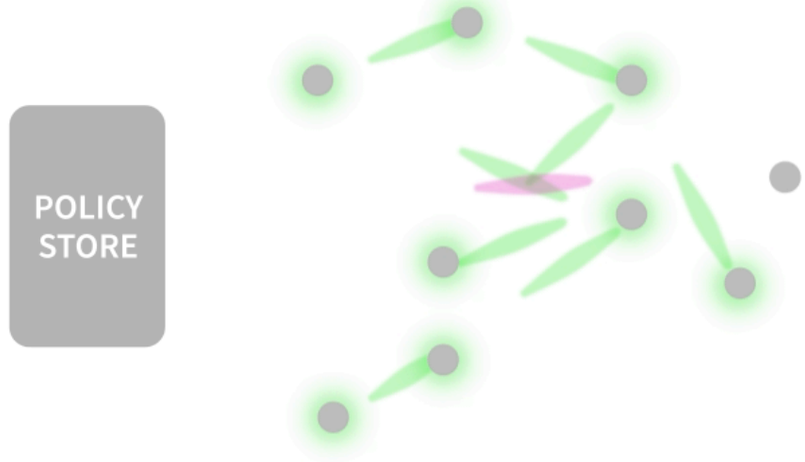


Scalability

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…
})…

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. 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.

2. Salerno, Jennifer, et al. "Ethics, big data and computing in epidemiology and public health." Annals of Epidemiology 27.5 (2017): 297-301.

3. 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.

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

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

6. 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.

7. Sharma, Aakash, et al. "Up-to-the-minute Privacy Policies via gossips in Participatory Epidemiological Studies." Frontiers in big Data 4 (2021).

LOHPI Team

Thomas Bye Nilsen

thomas.bye.nilsen @uit.no

Aakash Sharma

aakash.sharma @uit.no

Dag Johansen

dag.johansen @uit.no

Håvard D. Johansen

havard.johansen @uit.no

UiT The Arctic University of Norway

Lab→NN

SpareBank 1 NORD-NORGE

Lohpi logo designed by Katja Pauline Czerwinska