Evolveum

Data Provenance and Metadata Management in IdM
On-line Workshop

Slávek Licehammer & Radovan Semančík

September 2020

Agenda

- Provenance metadata introduction
- Demo
- MidPrivacy initiative
- Discussion

Evolveum

TRUST

This project has received funding from the European Unions Horizon 2020 research and innovation program under the NGI_TRUST grant agreement no 825618

Metadata: Data About Data

Data

```
{
  "username" : "jdoe",
  "fullName" : "Jane Doe",
  "title" : "Data Protection Specialist"
}
```

Metadata

```
{
  ...
  "fullName" : {
    "@value" : "Jane Doe",
    "@metadata" : {
      "timestamp" : "2020-06-22T15:29:28Z",
      "origin" : "federation",
      "actor" : "idp.example.edu"
    }
  }
  ...
```

Why Are Metadata Important?

- Transparency
- Accountability
- Data protection: basis for data processing
- Data assurance (reliability)
- ...
- Troubleshooting

Metadata: Whole New Dimension

```
{
  "username" : "jdoe",
  "fullName" : "Jane Doe",
  "title" : "Data Protection Specialist"
}
```

```
{
  "username" : {
    "@value" : "jdoe",
    "@metadata" : {
      "timestamp" : "2020-06-22T15:29:35Z",
      "origin" : "system-generated",
      "actor" : "sync-agent-0543"
    }
  },
  "fullName" : {
    "@value" : "Jane Doe",
    "@metadata" : {
      "timestamp" : "2020-06-22T15:29:28Z",
      "origin" : "federation",
      "actor" : "idp.example.edu"
    }
  },
  "title" : {
    "@value" : "Data Protection Specialist",
    "@metadata" : {
      "timestamp" : "2020-06-24T15:31:06Z",
      "origin" : "user-provided",
      "actor" : "asmith"
    }
  }
}
```

Provenance Metadata

Provenance: Origin; Where did it come from?

```
{
  "username" : {
    "@value" : "jdoe",
    "@metadata" : {
      "timestamp" : "2020-06-22T15:29:35Z",
      "origin" : "system-generated",
      "actor" : "sync-agent-0543"
    }
  },
  "fullName" : {
    "@value" : "Jane Doe",
    "@metadata" : {
      "timestamp" : "2020-06-22T15:29:28Z",
      "origin" : "federation",
      "actor" : "idp.example.edu"
    }
  },
  ...
```

Automatically generated by the system on June 22 2020 within process *sync-agent-0543*

Acquired from identity federation on June 22 2020, received from identity provider *idp.example.edu*

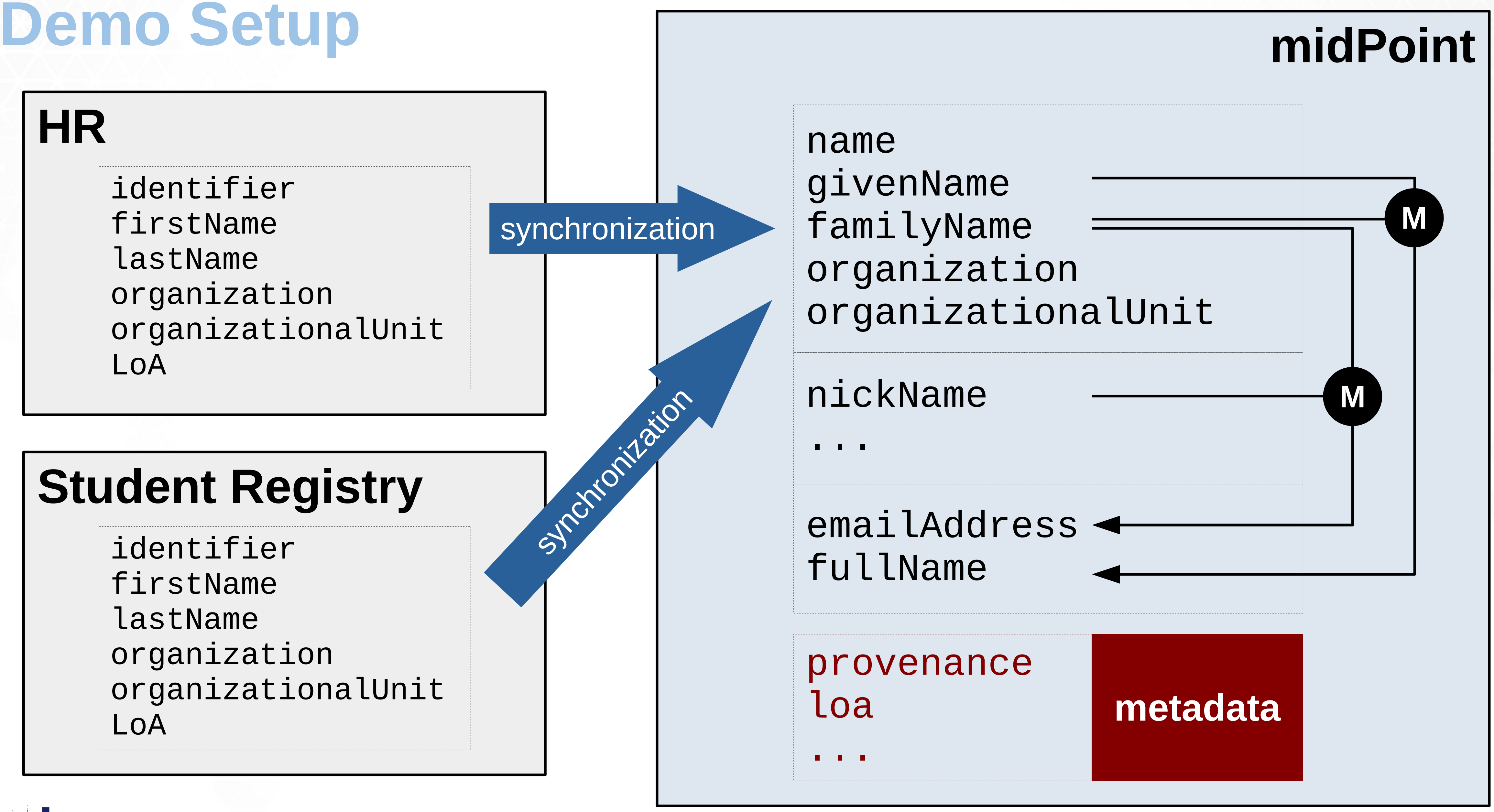
Evolveum

Disclaimer: This example is simplified. Reality is (much) more complex.

midPoint

- Identity Management and Identity Governance platform
  - Provisioning, administration, RBAC, organizational structure, audit, ...
- 100% open source
  - Apache license & EUPL
- Maintained by Evolveum
  - Full-time development and engineering teams at Evolveum
  - Community contributions
  - Professional support from Evolveum & partners

Demo Setup



Evolveum

DEMO

MidPrivacy Initiative

https://docs.evolveum.com/midpoint/midprivacy/

- Phase 0: Data protection experiments
  - Experiments with *basis for data processing* and *consent*
  - Done before GDPR came to force, but there was very little demand.
- Phase 1: Data Provenance Prototype
  - Provenance metadata as a basic foundation for data protection
  - NGI_TRUST funding
- Phase 2: ???
  - Depends on funding
- Phase 3: ???
  - Depends on funding

MidPrivacy Phase 1: Data Provenance Prototype

https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/

- Improvements to midPoint
- Maintaining meta-data for every *value*
- Visibility, accountability: foundation for data protection
- New schema language: Axiom
- Side effect: improve diagnostics
- Prototype targeted for midPoint 4.2
- Funding: NGI_TRUST

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the NGI_TRUST grant agreement no 825618.

Data Protection and Privacy

- Metadata as a foundation
  - We can process data only if we have a *basis for processing*
  - Storing *basis for data processing* in metadata (directly or indirectly)
  - Transparency and accountability
- Challenges and open questions remain
  - Axiom 0.1 to Axiom 1.0
  - Metadata multiplicity problem
  - User experience – especially for end users

See also https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/provenance-origin-basis/

Future

- Where do we go from here?
  - This is a very successful prototype, but it is just a prototype
  - How to make it production-ready?
- Data protection is needed, but it is difficult to “sell”
  - We know that those features are absolutely crucial for everybody
  - But getting paid for them *directly* is very hard
  - Long-term systematic approach is needed

See also https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/provenance-origin-basis/

Conclusion

- Data protection & privacy is a complex matter
  - Even more complex that we have anticipated
- Metadata as a foundation
  - Provenance metadata play a major role
- Innovation
  - Working prototype in midPoint 4.2
- Still a long way to go ...

Discussion

Questions & Answers

Thank you for your time

See other talks at https://docs.evolveum.com/talks

Also **follow us** on our social media for further information!

/Evolveum /Evolveum /Evolveum @Evolveum /Evolveum

Evolveum