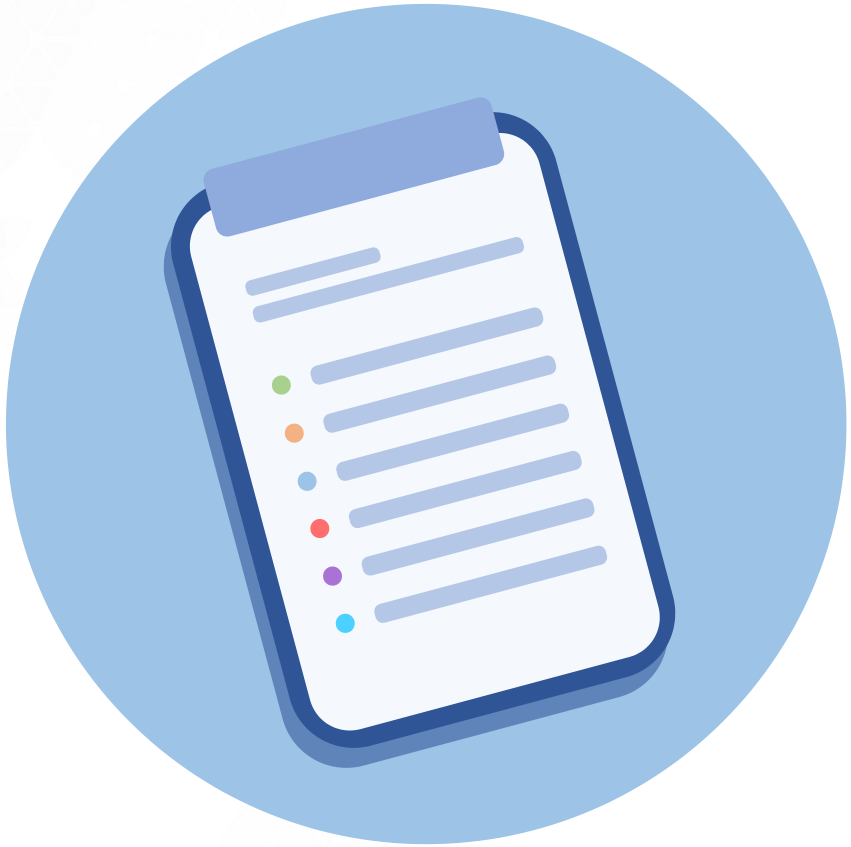
Evolveum

MidPrivacy: Identity Provenance

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

Slovakia, 3 September 2020

Agenda



- Introduction and goals
- Project progress
- Deliverables and achievements
- Demo
- Outcomes, discoveries, challenges
- Future work & Inspiration

Evolveum

NGI TRUST

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

Introduction to the project

Data provenance is one of the fundamental problems of data protection.

Data protections regulations and practices ask for transparency and accountability.

Data provenance was chosen as the primary goal of the first phase of MidPrivacy initiative because it brings a solid foundation to build full suite of privacy-enhancing features in the future.

Existing midPoint data model functionality will be enhanced with the meta-schema capabilities to track origin (provenance) of every data item.

MidPoint could be one of the first IDM systems to support user-centric data protection functionality.

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

MidPrivacy project goals

Expected project outcomes:

- Adapt existing or design new **data modeling (schema) language** to support meta-schema capabilities required to support data provenance features.
- Implement **prototype libraries** for processing the meta-schema to evaluate feasibility of this approach.
- Use meta-schema to **process provenance-annotated personal data**. Implement prototype functionality in midPoint.
- Implement **prototype user interface** to present the provenance data to the user.
  The purpose of this prototype is to evaluate whether the complex meta-data can be presented to a user in an understandable and intuitive way, thus supporting transparency and user intervenability in personal data protection. This prototype user interface can be used in future usability testing with a potential to be fully productized.
- **Evaluate market potential** for data protection features in IDM systems in two different ways:
  1) Conduct a quick study of market demand for data protection features (e.g. by using surveys, on-line and personal discussions and similar means).
  2) Use prototypes created in this project as a basis for further discussions with potential customers and users, evaluating potential for full productization of data protection mechanisms in midPoint.

Evolveum

NGI TRUST

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

Potential revenue

Vertical/Type	Recurring Subscription Revenue (mio EUR/year)	Subscription Recipients	Project Revenue (mio EUR)	Project Revenue Recipients
Enterprise	8	Evolveum	79,8	Internal/ Evolveum/ Partners
InCommon Members/Higher Education	4	Evolveum	26,6	Internal/Partners
Government/ Municipalities (citizens management)	4,5	Evolveum	30	Evolveum/ Partners
TOTAL	16,5	Evolveum	136,4	

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

Work done in the project

Analysis: We have gathered use cases for data provenance to provide input for data structure design later in the project.

Completion of solution architecture by filling in all the low-level design details. This included work on Axiom modeling language, metadata schemas (especially identity provenance metadata schema) and design of metadata handling and metadata value mapping.

Design and prototype of Axiom language: We have started designing new data modeling language: Axiom. The plan is to use Axiom language to model provenance meta-data and to apply them to existing midPoint schemas.

Implementation of prototype code and integrating the code into midPoint. Major areas were: implementation of Axiom-processing code to support metadata models (schema) and implementation of metadata mappings. Development of the code for both parts was finished, the code was integrated into midPoint core code base.

Survey of intended use of metadata by midPoint community. we have received only a small number of responses to consider the results relevant to make sound decisions about metadata schemas. Despite that we were able to get at least some informal insights from the survey results.

Metadata user interface: User interface to display metadata to users (mostly for system administrators).

Validation and improvements: We have tested the code and validated the functionality by preparing testing configurations. This lead to identification of several deficiencies, which in turn lead to code and design improvements.

Evolveum

NGI TRUST

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

Major achievements (Milestone 3) – NOT REACHED YET

- Metadata user interface (GUI)
  - Several design and implementation iterations
- Metadata schemas: multiplicity
  - Probably the most important discovery
- Axiom – code generator prototype
  - Needed for maintainability
- Testing and validation
  - Validation lead to improvements
- Preparing the demo

M3

Evolveum

NGI TRUST

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

Milestones in the project

Project Milestones	Title	Date achieved
MS0	Project start	16. March 2020
MS1	Meta Schema prototype	15. May 2020
MS2	Meta-schema integrated into midPoint core	15. July 2020
MS3	Project finish	15. September 2020

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

MS1/MS2 Deliverables in the project

Deliverables	Milestone	Responsible person	Start date	Date achieved	% of completion
Research data modeling and schema languages	Meta Schema prototype (MS1)	Anton Tkacik	16. March 2020	30. April 2020	**100%**
Requirements gathering and market research	Meta Schema prototype (MS1)	Slavek Licehammer	16. March 2020	15. May 2020	**100%**
Implementation of Meta-schema prototype	Meta Schema prototype (MS1)	Anton Tkacik	1. April 2020	15. May 2020	**100%**
Design Solution Architecture	Meta-schema integration into midPoint core (MS2)	Radovan Semancik & Pavol Mederly	1. April 2020	15. June 2020	**100%**
Integrate the meta-schema prototype into midPoint core	Meta-schema integration into midPoint core (MS2)	Anton Tkacik & Pavol Mederly	16. May 2020	15. July 2020	**100%**

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

M3 Deliverables in the project

Deliverables	Milestone	Responsible person	Start date	Planned due date
Finish implementation of user interface	MS3 Project Finish	Katarina Valalikova	16. July 2020	15. August 2020
Test and validate metadata processing code in midPoint	MS3 Project Finish	Anton, Katarina, Pavol, Slavek, Radovan	16. July 2020	31. August 2020
Gather and evaluate user feedback	MS3 Project Finish	Slavek Licehammer	16. July 2020	1. September 2020
Solution review and evaluation	MS3 Project Finish	Radovan Semancik	1. September 2020	15. September 2020

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

Deliverables of MS1 done in the project

MS1	Meta-schema prototype	Start Month: 16. March 2020	End Month: 30. April 2020
Objective	Look for candidate language that can be adapted for meta-schema purposes. In case that such alternative is not feasible, prepare a design for prototype of new data modeling language. Note: design of full-featured data modeling language is beyond the scope of this project. We are looking just to address the needs of meta-schema for the prototyping purposes. Design of a fully-featured data modeling language could be a follow-up activity after this project phase is finished, if needed.		
Related Deliverable	**Research data modelling and schema languages**		
Participants	Anton Tkacik		
% of completion	**100%**		

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

Deliverables of MS1 done in the project

MS1	Meta-schema prototype	Start Month: 16. March 2020	End Month: 15. May 2020
Objective	Analyze and discuss the needs of potential user communities, especially requirements for data protection and trust in identity management. Evaluate market potential.		
Related Deliverable	**Requirements gathering and market research**		
Participants	Slavek Licehammer		
% of completion	**100%**		

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

Deliverables of MS1 done in the project

MS1	Meta-schema schema prototype	Start Month: 1. April 2020	End Month: 15. May 2020
Objective	Implement prototype libraries for processing the meta-schema to evaluate feasibility of this approach.		
Related Deliverable	**Implementation of Meta- Schema prototype**		
Participants	Anton Tkacik		
% of completion	**100%**		

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

Deliverables of MS2 done in the project

MS2	Meta-schema integration into midPoint core	Start Month: 1. April 2020	End Month: 15. June 2020
Objective	Evaluate feasibility of use of schema languages for meta-schema purposes, design integration approach with midPoint code – ensuring that midPoint architecture and conceptual design is maintained and that the requirements are reflected.		
Related Deliverable	**Design Solution Architecture**		
Participants	Radovan Semancik & Pavol Mederly		
% of completion	**100%**		

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

Deliverables of MS2 in the project

MS2	Meta-schema integration into midPoint core	Start Month: 16. May 2020	End Month: 15. July 2020
Objective	Replace or augment schema-processing code currently used by midPoint. The goal is to enable the meta-schema capabilities to enable storage and processing of identity provenance meta-data in midPoint.		
Related Deliverable	**Integrate the meta-schema prototype into midPoint core**		
Participants	Anton Tkacik & Pavol Mederly		
% of completion	**100%**		

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

Deliverables of MS2/M3 in the project

MS3	Project Finish	Start Month: 16. March 2020	End Month: 15. August 2020
Objective	Implementation of prototype user interface to present the provenance data to the user as an extension of current midPoint user interface		
Related Deliverable	***Implementation of prototype user interface**		
Participants	Katarina Valalikova		
% of completion	**100%**		

* This deliverable has started earlier. Original start date was 15th of June 2020.

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

Deliverables of MS3 in the project

MS2/MS3	Project Finish	Start Month: 15. July 2020	End Month: 31. August 2020
Objective	Validation and Testing of meta schema		
Related Deliverable	**Meta-schema validation & testing**		
Participants	Katarina Valalikova, Radovan Semancik, Pavol Mederly, Slavek Licehammer, Anton Tkacik		
% of completion	**100%**		

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

Deliverables of MS3 in the project

MS3	Project Finish	Start Month: 16. July 2020	End Month: 01. September 2020
Objective	Present the prototype implementation to user communities to gather their feedback, suggestions and possible improvement ideas		
Related Deliverable	**Gathering & Evaluation of user feedback**		
Participants	Slavek Licehammer		
% of completion	**100%**		

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

Deliverables of MS3 in the project

MS3	Project Finish	Start Month: 01. September 2020	End Month: 15. September 2020
Objective	Evaluation of the outcomes to validate whether project goals were fulfilled. Evaluation of user feedback. Planning follow-up activities.		
Related Deliverable	**Review of the project outcomes**		
Participants	Radovan Semancik		
% of completion	**25%**		

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

Spent Persons Months in the project

Name of dedicated person	Organization	Planned Person-Month	Start date	Due date	Man-Days spent	Spent Person Month	Outstanding Person Month
Radovan Semancik	Evolveum	2.4 FTE	16. March 2020	31. August 2020	41.52	2.08 FTE	0.32 FTE
Katarina Valalikova	Evolveum	1.5 FTE	16. March 2020	31. August 2020	40.34	2.02 FTE	- 0.52 FTE
Pavol Mederly	Evolveum	1.8 FTE	16. March 2020	31. August 2020	57.01	2.85 FTE	- 1.05 FTE
Anton Tkacik	Evolveum	6 FTE	16. March 2020	31. August 2020	117.13	5.86 FTE	0.14 FTE
Slavek Licehammer	Evolveum	1.8 FTE	16. March 2020	31. August 2020	32.38	1.62 FTE	0.18 FTE
TOTAL		13.5 FTE			288.38	14.43 FTE	- 0,93 FTE

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

Budget

Organization	Total budget (EUR)	Budget used (EUR)	% of Budget Used
Evolveum	75.000,-	75.000,-	100%

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

Demo

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

Outstanding tasks

- **Public demo/workshop:** present results to the community, postponed to September.
- **Gather and evaluate user feedback** the plan is to focus the evaluation on midPoint community.
- **Solution review and evaluation (“wrap up”):** describe the outcomes, final report, update project page, etc.

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

Final outcomes of the project

- Axiom data modeling language

  Very unique aspects (inframodel, metadata), but still early draft

- End-to-end provenance metadata in midPoint

  Prototype-quality code, but some parts may be ready for production use

- MidPoint-specific metadata schemas

  Provenance, storage, transformation, process, ...

- User-customizable metadata

  Both schema and behavior

- Metadata support in midPoint user interface

More details about the outcomes: https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/outcomes/

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

Discoveries

- Data *inframodel* (Axiom concept)

  Model of underlying data structure, allows specification of metadata in data modeling

- Metadata multiplicity (and related concepts)

  Metadata has to be multi-valued, provenance plays a prominent role

  Concept of *yield*

- Intimate relation of metadata to data protection and portability

  Yield closely related to *basis for data processing*

- Metadata user experience challenge

  Multiplicity makes UX really challenging. How to make it understandable for ordinary users?

More details about the outcomes, including discoveries: https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/outcomes/

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

Side Effects

- Identity Metadata in a Nutshell

  Introductory document, a “tutorial” to boost understanding of metadata

- Axiom concepts and design details

  Axiom design goes beyond metadata, potential for future development

- Blog posts, community awareness, etc.

More details about the outcomes, including side effects: https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/outcomes/

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

Data Provenance Challenges in the project

- Axiom language design

  Namespaces, model indentification, metamodel/inframodel, subtyping, augmentation, …

- Provenance metadata schemas

  No real standard, have to start from scratch

- Metadata: multiplicity, equality and deltas

  Perhaps the most challenging and surprising part of the project

- Metadata mappings

  Suffering from the “multiplicity” problem

  Had to fit in existing midPoint code

- User interface (UX)

More information about data provenances challenges: https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/challenges/

Evolveum

NGI TRUST

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

Future Work & Inspiration

- Metadata & provenance research and schema evolution

  Standardized/unified provenance schema. Further validation/formalization of “multiplicity” concepts: yield, acquisition

- Metadata storage and use

  Metadata queries and indexing. Authorizations. Compatibility.

- User experience

  How to present metadata to ordinary users? How to make it understandable?

- Data protection

  Basis for data processing, metadata are basic building block.

- Data portability

  Provenance is related to portability. How to proceed? DAPSI?

- Commercialization

  Low commercial interest so far. How to fund further development?

More details about future work: https://docs.evolveum.com/midpoint/midprivacy/phases/01-data-provenance-prototype/future-work/

Evolveum

NGI TRUST

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

Thank you for your time

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