

Towards an Architecture for Data Altruism in Solid

Beatriz Esteves^{1,*}

¹*Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

Abstract

This demo showcases an architecture to implement data altruism as a service using the Solid protocol and ODRL policies to grant access to personal data for altruistic purposes in a privacy-friendly manner. Policies are represented using OAC, the ODRL profile for Access Control, and DGATerms, a vocabulary with terms modelled from the European Union’s Data Governance Act (DGA), including *data altruism* concepts. In addition, we present a Solid Data Altruism application, *SoDA*, where (a) a data subject can generate a policy to share their personal data for an altruistic purpose, (b) data users can request access to datasets for altruistic purposes, and (c) data altruism organisations can use to maintain metadata regarding available datasets.

Keywords

Solid, data altruism, ODRL policies, personal data access

1. Introduction


Following the current efforts to decentralise the storage and access to data on the Web [1], the Solid protocol [2] allows its users to have their data stored on personal datastores, the “Pods”, and control which users and applications can have access to it. The regulatory agenda in Europe has followed this technological trend by putting data subjects – individuals whose personal data is being processed – in a decision-making position with regards to their data, while improving data availability and promoting trust in data intermediation services [3]. In particular, the Data Governance Act (DGA) [4] introduces the concept of *data altruism* – the voluntary sharing of data for the general interest of the public, such as improving healthcare systems or combating climate change, managed by *data altruism organisations*, non-profits who make personal (and non-personal) data available to data users who wish to use such data for altruistic purposes.

In this demo, we propose to use the ODRL profile for Access Control (OAC) [5] to create policies to determine access to data stored in Solid Pods. By using previous work on the DGATerms vocabulary [6], these policies can also be specified for specific altruistic purposes. In addition, we present an architecture and a proof-of-concept Solid Data Altruism application, SoDA, which in addition to allowing data subjects to generate these policies, allows data users to request access to datasets for altruistic purposes. The paper is organized as follows: Section 2 describes related work, Section 3 presents a description of the architecture and of the proof-of-concept demonstration, Section 4 describes the used technologies and details the data modelling used in the work and the last section presents conclusions and future work.

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece

*Corresponding author. Email: beatriz.gesteves@upm.es

 0000-0003-0259-7560 (B. Esteves)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

2. Related Work

A few solutions are already emerging to deal with DGA's requirements as it will be applicable from the 24th of September 2023. For instance, the Smart Citizen platform¹ allows citizens to collect noise and air quality data through home sensors and share that data through the platform with researchers and governments for them to develop targeted solutions for pressing environmental issues such as air pollution. The German Corona-Datenspende-App² collected health-related data from fitness bracelets and smartwatches, e.g., heart rate, body temperature, blood pressure, sleeping patterns, for researchers to monitor and identify at an early stage possible Covid-19 hotspots.

3. Demonstration

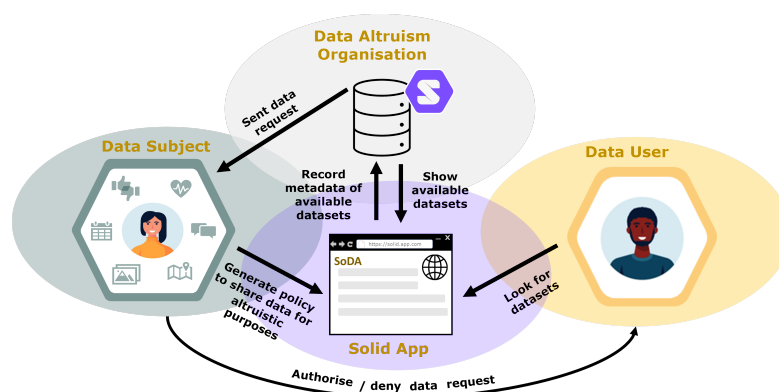


Figure 1: High-level overview illustrating an architecture to implement data altruism as a service using the Solid protocol and a Solid application, SoDA, to make available/search datasets.

The diagram in Figure 1 showcases a high-level overview of an architecture to implement data altruism as a service using the Solid protocol, with the Solid Data Altruism application – SoDA – at the centre of the system architecture. With this architecture, we aim to provide an early proof-of-concept which focuses on allowing data subjects to share personal data and data users to look for datasets available to be reused for altruistic purposes, in a privacy-friendly manner as the only information that is disclosed about the dataset is the type of data it contains and the purpose for which it can be used.

In this architecture, Solid users are identified by a WebID and store data and/or request to access data stored in Solid Pods, as prescribed by the Solid protocol. In the cases where personal data is stored in Pods, personal data protection laws apply, such as the European Union's General Data Protection Regulation (GDPR) and the DGA, with the user storing their personal data in Pods being considered a *data subject*. Moreover, both data subjects and data users manage access to data through Solid applications. In this context, we introduce *SoDA*, a Solid Data Altruism application, which allows:

¹<https://smartcitizen.me/>

²<https://corona-datenspende.de/>

```

1 PREFIX dct: <http://purl.org/dc/terms/>
2 PREFIX dcat: <http://www.w3.org/ns/dcat#>
3 PREFIX odrl: <http://www.w3.org/ns/odrl/2/>
4 PREFIX dpv: <https://w3id.org/dpv#>
5 PREFIX oac: <https://w3id.org/oac#>
6 PREFIX dga: <https://w3id.org/dgaterms#>
7 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8 PREFIX ex: <http://example.com/>
9 ex:policy-123456 a odrl:Offer ; odrl:uid ex:policy-123456 ; odrl:profile oac: ;
10   dct:creator <https://solidweb.me/userA/profile/card#me> ;
11   dct:issued "2023-07-19T17:26:35"^^xsd:dateTime ;
12   odrl:permission [
13     odrl:assigner <https://solidweb.me/userA/profile/card#me> ;
14     odrl:action oac:Read ; dpv:hasPersonalData ex:EnergyConsumption ;
15     odrl:target <https://solidweb.me/userA/energyconsumption/june2023> ;
16     odrl:constraint [
17       odrl:leftOperand oac:Purpose ;
18       odrl:operator odrl:isA ;
19       odrl:rightOperand dga:CombatClimateChange ] ] .

```

Listing 1: ODRL Offer policy set by User A that allows read-access to a dataset with EnergyConsumption data for the purpose of combating climate change.

- (a) data subjects to generate policies to share their personal data for an altruistic purpose;
- (b) data users to request access to datasets according to the type of data they contain and the purpose for which it can be used;
- (c) organisations to provide data altruism as a service, by storing metadata regarding available datasets in their own Solid Pod, without having the need to store the data themselves, following Solid’s decentralisation philosophy.

Using SoDA³, data subjects can generate data access policies related to the access to their personal data, which are stored by the data altruism organisation in a Solid Pod which only records metadata about the dataset and their access conditions. These records are then used to show the available datasets to data users using SoDA, preserving the data subjects’ privacy by only showing the type of data available and the purpose for which it can be used, without revealing the identity of the data subject. If data users find datasets that they wish to use, the data altruism organisation acts as an intermediary by sending the data request to the data subject, who then decides to authorise/deny access. More details on this demonstration are available at <https://besteves4.github.io/iswc23demo/>, including a recording of the app’s functionalities.

4. Data Modelling

In this demo, OAC⁴ is used to define legally-aligned policies to grant access to personal data stored in Solid Pods since it is an RDF-based specification that uses (i) the Open Digital Rights Language (ODRL) standard to express different types of access policies, e.g., offers,

³Source code is available at <https://github.com/besteves4/soda>.

⁴<https://w3id.org/oac#>

```

1 ex:datasets a dcat:Catalog ; dct:created "2023-06-10"^^xsd:date ;
2   dct:description "Catalogue of datasets maintained by SoDACompany" ;
3   dct:publisher ex:SoDACompany ; dcat:dataset ex:dataset_001 .
4 ex:SoDACompany a dga:DataAltruismOrganisation .
5 ex:dataset_001 a dcat:Dataset ; odr1:hasPolicy ex:policy-123456 ;
6   dpv:hasLocation <https://solidweb.me/userA/energyconsumption/june2023> ;
7   dct:publisher <https://solidweb.me/userA/profile/card#me> ;
8   dct:description "Dataset with energy consumption data of June 2023" ;
9   dcat:mediaType <https://www.iana.org/assignments/media-types/text/csv> .

```

Listing 2: Catalogue of datasets maintained by a Data Altruism Organisation.

requests or agreements, associated with data stored in decentralised datastores, and (ii) the Data Privacy Vocabulary (DPV) [7] as a controlled vocabulary for invoking privacy and data protection-specific terms. Moreover, OAC was chosen as it can be used to extend Solid’s access control list mechanism, Web Access Control (WAC) [8], to have richer access control policies where specific purposes for access can be defined, among other constraints such as restrictions on the access duration or on the types of entities, e.g., non-profit or for-profit, that can use the data. In addition, the DGATerms vocabulary⁵ is used to represent the altruistic purposes defined in the DGA, such as scientific research or combating climate change. Listing 1 presents an example of a policy set by User A, which allows data users to read the dataset stored at <https://solidweb.me/userA/energyconsumption/june2023>, which contains EnergyConsumption data as it is indicated by the `dpv:hasPersonalData` predicate, for the altruistic purpose of combating climate change.

In addition, W3C’s Data Catalog Vocabulary (DCAT) is used to maintain a catalogue of the available datasets, which allows the data altruism organisation to show available datasets to data users and send data requests in their name in a privacy-friendly manner as data users only get access to the dataset if the data subject authorises it. Listing 2 presents an example of a catalogue of datasets maintained by *SoDACompany*, a data altruism organisation. Metadata regarding the dataset storage location, the publisher of the dataset and the policy that determines access to it is also recorded in these catalogues.

5. Conclusions and Future Work

In this demo, we presented an architecture to manage data altruism activities in a decentralised setting, such as Solid, and an application that allows data subjects to generate policies regarding data they wish to make available for the public good and data users to look for such datasets.

Such an architecture would help to achieve the European Commission’s vision of having trustworthy data altruism services where data subjects are in control of who can access their data. This system needs to be complemented by future endeavours on: (i) SHACL shapes to validate the policies, (ii) usability testing to assess the app’s design choices, including scalability testing – which might require the usage of data aggregators to deal with organisations that want to access a large number of datasets, (iii) improving/automating the process of authorising/denying data

⁵<https://w3id.org/dgaterms#>

requests using technologies, such as RDF surfaces [9], to reason over the offer/request policies and contribute to the (iv) generation of immutable agreements – e.g., using existing work on integrating Verifiable Credentials into the Solid ecosystem [10] – that record the conditions for data usage that can be utilised by authorities in case the entities using the data misbehave.

Acknowledgments

This research has been supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813497 (PROTECT) and Horizon 2020 innovation action under grant agreement No. 101036418 (AURORA).

References

- [1] S. Verbrugge, F. Vannieuwenborg, M. Van der Wee, D. Colle, R. Taelman, R. Verborgh, Towards a personal data vault society: an interplay between technological and business perspectives, in: FITCE 2021, 2021. doi:10.1109/FITCE53297.2021.9588540.
- [2] S. Capadisli, T. Berners-Lee, R. Verborgh, K. Kjernsmo, Solid Protocol Version 0.10.0, W3C Community Group Draft Report (2022). URL: <https://solidproject.org/TR/protocol>.
- [3] European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - A European strategy for data, 2020.
- [4] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), 2022.
- [5] B. Esteves, H. J. Pandit, V. Rodríguez-Doncel, ODRL Profile for Expressing Consent through Granular Access Control Policies in Solid, in: 2021 EuroS&PW, 2021, pp. 298–306. doi:10.1109/EuroSPW54576.2021.00038.
- [6] B. Esteves, V. Rodríguez-Doncel, H. J. Pandit, D. Lewis, Semantics for Implementing Data Reuse and Altruism under EU’s Data Governance Act, in: To Appear on SEMANTiCS 2023 Proceedings, 2023. doi:10.5281/zenodo.8301901.
- [7] H. J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F. J. Ekaputra, J. D. Fernández, R. G. Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal, R. Wenning, Creating a Vocabulary for Data Privacy: The First-Year Report of Data Privacy Vocabularies and Controls Community Group (DPVCG), in: On the Move to Meaningful Internet Systems: OTM 2019 Conferences, volume 11877, Springer International Publishing, 2019, pp. 714–730. doi:10.1007/978-3-030-33246-4_44, Lecture Notes in Computer Science.
- [8] S. Capadisli, Web Access Control 1.0.0, W3C Candidate Recommendation (2022). URL: <https://solidproject.org/TR/wac>.
- [9] P. Hochstenbach, J. De Roo, R. Verborgh, RDF Surfaces: Computer Says No, in: 1st Workshop on Trusting Decentralised Knowledge Graphs and Web Data, 2023.
- [10] C. H.-J. Braun, T. Käfer, Attribute-based Access Control on Solid Pods using Privacy-friendly Credentials, in: Proceedings of the Poster and Demo Track and Workshop Track of SEMANTiCS 2022, 2022.