Towards a personalization of privacy preservation: AI democratizes Swedish open data

Esteban Guerrero[1], Proscovia Svärd[2] and Lili Jiang[1]

[1] Department of Computing Science, Umeå University, Sweden
{esteban, lili.jiang}@cs.umu.se
[2] Department of Information Systems Technology, Mid Sweden University, Sweden
proscovia.svard@miun.se

Abstract. Considering the positive impacts of Open Government Data (OGD) and the necessity to empower individuals in data management, we indicate data openness towards democracy in Sweden by building a data information model, and present an intelligent agent based framework of democratizing data openness by taking account of personalization and privacy preservation. Especially we provide the insights how AI enhances OGD and discuss future directions.

Keywords: open government data · agent · personalization · privacy

1 Motivation

In general, governments are undertaking different initiatives to release data from governmental institutions aiming to improve government transparency, strengthen social and commercial values, and enhance participatory governance [1]. In Sweden, a gap exists between the implementation of Open Government Data directives (*e.g.* the Public Sector Information (PSI)[3]) and its awareness, acceptance and consumption by individual Swedish citizens. *Data privacy*, as a key component of OGD, has been overlooked[4] without automatized processes, current mechanisms for *privacy preservation* adopted in Sweden lacks public oversight and personalized data control, *i.e.* a deficiency in the full democratization of open data.

The goal of building technological mechanisms as tools to empower individuals in data managment during the entire OGD Life-Cycle is complex [1], facing at least three of the following challenges: 1) understand what type of data is available, *i.e.,* a general information model of OGD for the Swedish context; 2) quantify the release of OGD, *i.e.,* generate OGD indexes; and 3) build technical infrastructure to support mechanisms for OGD management, moving from the current static Web technology towards autonomous and tailored to individuals machinery, *i.e.,* human-centered artificial intelligence (AI)-based systems for OGD.

---

[3] See PSI definition on the Swedish portal for open data: https://oppnadata.se/definitioner/ Last access: 05-May-2020

[4] See the Swedish National Archives report: "Riksarkivets kartläggning av användarbehov kring PSI-förteckningar och öppna data" http://s.cs.umu.se/h5o0a9 Last acceded on the 05-May-2020

This paper presents a framework for democratizing Swedish open government data dealing with the three aforementioned challenges using hybrid approaches of AI. Our contributions are: 1) an informational model of OGD and democracy in Sweden, aimed at structure and quantify *openness* of OGD; 2) the creation of metrics to assess data openness and privacy; and 3) a software agent-based platform for OGD management. In this paper, we provide an overall description of these three solutions laying aside formal and technical characterizations of individual mechanisms.

The outline of this paper is the following, in Section 2, an information model on open data for the Swedish context is presented. A framework to implement OGD management is presented in Section 3. A discussion regarding our approach presenting future work and expectations is introduced in Section 4.

2   Open data informational model

In this section, an open data information model is presented to indicate data openness in the Swedish case scenario. This model will contain key information of data openness to quantify openness, *i.e.,* indicators such as accessibility, timely, completeness, machine process ability, etc. [8]. Rodriguez [7] developed a Data Openness (DO) model to measure the openness of government data, which takes several indicators into considerations including data completeness, timely, primary, accessible, machine processable, nondiscriminatory, non-proprietary, and license free. Motivated by this work, we incorporate the special characters for Sweden to build Swedish data openness indicator. The aim is to extend this approach to model and measure data openness in Sweden with respect to the selected criteria of democracy indexes.

The conceptual model will guide us to define benchmarks of data openness and democracy indexes for the Swedish context. To this end, we will employ a Participatory Action Research (PAR) method to identify the problems-solutions of PSI implementation, along with institutions and organizations involved in this process [5]. Data openness [8] and democracy indexes and indicators will be investigated and extended to the Swedish case. We will use statistics to identify relations among information parameters (indexes and indicators).

3   An AI agent based autonomous open data platform

In general, the OGD information management is a complex task where governmental institutions release information that is partially linked to a central governmental organization called Öppna Data `https://oppnadata.se/`. In our AI-based platform, we follow the same distributed approach, where agent-based federated software manages OGD. This modular and federated architecture of this platform, aims towards technologically back the awareness, acceptance, consumption and adoption of OGD.

To simplify the presentation, the agent-based federated *nodes* can be seen as independent entities managing their inputs: *i*) OGD from Swedish gov. institutions, *ii*) authority directives and rules (*e.g.* PSI), and *iii*) the individual
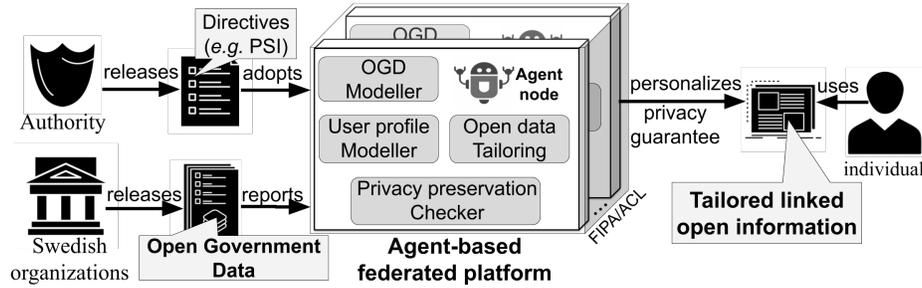
Fig. 1. An AI Agent-based framework for preserving OGD privacy in Swedish context.

queries made by users (see Figure 1). The agent-based platform is in charge of
1) defining information models for every entity manipulating OGD in order to
semantically link OGD; 2) manage OGD privacy (*.e.g* using privacy preservation
techniques); 3) tailor OGD requests to users (*e.g.* based on access privileges);
and 4) reporting OGD use, publication, and statistical OGD management to the
central authority. As state-of-the-art agent-based technology, nodes in the plat-
form are endowed with agent-to-agent communication channels (*e.g.* FIPA-ACL
[6]) aimed to provide distributed deliberation capabilities.

Key in this project, is to guarantee levels of privacy to users and providers of
open data. Personalized privacy preservation details the technological AI-based
mechanisms for tailoring privacy algorithms, e.g., differential privacy approaches
[2]. A trade-off between data utility and privacy control will be explored. In the
Swedish OGD context, there are organizations with different levels of privacy
protection, *e.g. Skatteverket* vs. *Kungliga Biblioteket*. The main problem is the
unification of privacy specifications for all the Swedish organizations. Our pre-
vious work on differential privacy [9] can be used to address the privacy budget
issue for organizations publishing open data.

The task of privacy personalization is oriented to design AI-based mecha-
nisms to guarantee data privacy to OGD producers and consumers. We will
extend the current literature about e-government openness index by adding two
components: privacy preservation checker based on our previous work [10,9] and
OGD tailoring to provide personalized users' privacy support considering our
previous experience on AI-based coaching systems [3,4].

4   Discussion of significance and future work

Access to open data promotes the citizens' engagement in governance issues but
it also has a potential to improve their financial standing through its exploitation
to create new electronic services. The fact that open data is still concentrated
in the hands of a few individuals in the society requires more research to be
undertaken to promote it as a democratic endeavour. Currently, the majority of
the citizens are outside of this development though they are included in policy
wise. If Sweden is to avoid a data and digital divide and genuinely promote
open data to non-experts, there is a need for more empirical studies that will

promote an understanding of the challenges and to bridge the knowledge gap that currently exists on the implementation of the PSI directive, publication of open data, its use, consumption and adoption and the data and digital divide. At present, research confirms that though the rhetoric on open data gives an impression that it is open and available to all citizens, that is not the case. The study presented in this paper is therefore important to the society as a whole and the policy makers. Open government initiatives are about creating inclusive government institutions and inclusivity can only be achieved through participation.

For future work, we will frame a use case scenario to create a benchmark of information testing our theoretical proposals. Afterwards, the experimental study will will be conducted to test the AI-agent framework using the Swedish use case scenario. Synthetic data could be utilized to test AI-algorithms of personalized privacy preservation. In this direction, we expect to obtain feedback to iteratively adjust our theoretical approaches, e.g., integrating new parameters of information to the openness data index.

References

1. Attard, J., Orlandi, F., Scerri, S., Auer, S.: A systematic review of open government data initiatives. Government Information Quarterly **32**(4), 399–418 (2015)
2. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science **9**(3–4), 211–407 (2014)
3. Guerrero, E., Lu, M.H., Yueh, H.P., Lindgren, H.: Design Principles and Action Reflection for Agent-Based Assistive Technology. In: International Workshop on Artificial Intelligence in Health. pp. 84–98. Springer (2018)
4. Lindgren, H., Guerrero, E., Janols, R.: Personalised persuasive coaching to increase older adults' physical and social activities: a motivational model. In: International Conference on Practical Applications of Agents and Multi-Agent Systems. pp. 170–182. Springer (2017)
5. Minkler, M.: Using participatory action research to build healthy communities. Public health reports **115**(2-3), 191 (2000)
6. Poslad, S., Buckle, P., Hadingham, R.: The fipa-os agent platform: Open source for open standards. In: proceedings of the 5th international conference and exhibition on the practical application of intelligent agents and multi-agents. vol. 355, p. 0 (2000)
7. Rodríguez-Bolívar, M.P.: Approaches to visualising linked dameasuring e-government efficiency
8. Veljković, N., Bogdanović-Dinić, S., Stoimenov, L.: Benchmarking open government: An open data perspective. Government Information Quarterly **31**(2), 278–290 (2014)
9. Vu, X.S., Jiang, L.: Self-adaptive privacy concern detection for user-generated content. arXiv preprint arXiv:1806.07221 (2018)
10. Vu, X.S., Jiang, L., Brändström, A., Elmroth, E.: Personality-based knowledge extraction for privacy-preserving data analysis. In: Proceedings of the Knowledge Capture Conference. pp. 1–4 (2017)