Management of Big Data in the public sector – system level risks and design principles

Agnieszka Rychwalska[1], Geoff Goodell[2], Magdalena Roszczyńska-Kurasińska[1].

(1) The Robert Zajonc Institute for Social Studies, University of Warsaw, Stawki 5/7, 00-183 Warsaw, Poland, a.rychwalska@uw.edu.pl, (2) University College London, Malet Place Engineering Building, London WC1E 7JE, United Kingdom, g.goodell@ucl.ac.uk

While Big Data has been a pivotal factor of success for many global, private companies, application of Big Data analytics in the public sector is still uncommon. Many public policy researchers and practitioners voice the sentiment that Big Data should be used in public policy to realize similar advantages that have boosted the private sector [1]. However, there are multiple issues preventing such development – from digitalization problems, to data format incompatibility, to lack of regulation and privacy concerns, among others [2]. Here we argue that before Big Data is used to guide and deploy policies, the system level risks of centralized data management – so far rarely recognized in public policy research - need to be acknowledged by policy makers and properly mitigated.

Drawing from an analysis of the systemic problems resulting from applications of Big Data analytics in the private sector – such as polarization of opinions and cascades of false news adoption – we conclude that the practices of data aggregation and centralization can lead to a loss of diversity in the system and a breakdown of self-regulation. These two effects can be devastating for democratic polities which by definition rely on the principles of plurality, decentralization, and accountability of representatives. We argue that in public services, inappropriate management of citizen data can be used to compromise democratic processes through the manipulation of incentives, opinions, and barriers to access. The deleterious consequences of misuse of Big Data in the administration of public services are further exacerbated because the data might be particularly sensitive and detailed, and also because the users might not have a legitimate choice about whether to share their data.

We propose a set of design constraints that should underlie data systems in public services and which can serve as a guideline or benchmark in the assessment and deployment of ICT-mediated services. The principles include, among others, minimizing control points and non-consensual trust relationships, empowering individuals to manage the linkages between their activities and empowering local communities to create their own trust relations.

We further propose a set of generic and generative design primitives that fulfil the proposed constraints and exemplify best practices in the deployment of platforms and applications that deliver services in the public interest. For example, blind tokens and attribute-based authorization may prevent the undue linking of data records on individuals. We suggest that policymakers could adopt these design primitives and best practices as standards by which the appropriateness of candidate technology platforms can be measured in the context of their suitability for delivering public services.

References

[1] Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data Applications in the Government Sector. Commun. ACM, 57(3), 78–85.

[2] Desouza, K. C., & Jacob, B. (2017). Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration & Society, 49(7), 1043–1064.