Orchestrated Co-creation of High-Quality Open Data Within Large Groups

Giuseppe Ferretti, Delfina Malandrino, Maria Angela Pellegrino, Andrea Petta, Gianluigi Renzi, Vittorio Scarano, Luigi Serra

Orchestrated Co-creation of High-Quality Open Data within Large Groups

Giuseppe Ferretti[1], Delfina Malandrino[2], Maria Angela Pellegrino[2](✉), Andrea Petta[2], Gianluigi Renzi[1], Vittorio Scarano[2], and Luigi Serra[2]

[1] Consiglio Regionale della Campania, Napoli, Italy
{ferretti.giu, renzi.gia}@consiglio.regione.campania.it
[2] Dipartimento di Informatica, University of Salerno, Italy
{dmalandrino, mapellegrino, vitsca}@unisa.it,
{andrpet, luigser}@gmail.com

Abstract. According to Open Government Data, governments should co-operate with citizens in order to co-create Open Data (OD). When large groups are involved, there is the need to orchestrate the work by clearly defining and distributing roles. Our Regional Administration - the Council of the Campania Region in Italy - claimed a motivating use case which inspired the proposed roles involved in the OD production process. We consider *validator*, *creator*, and *filler* as roles. To each role tasks and responsibilities are attached. Roles and related activities are integrated into SPOD (a Social Platform for Open Data) to guide users in producing high-quality OD by proactive quality assurance techniques.

Keywords: Open Data, Open Government, Orchestration, Roles, Large Groups, Co-creation, Quality Assurance

1 Introduction

Open Data (OD) refer to "*data which are open for free access, use and modification to be shared for any purpose*" [15]. In the last years, the e-government communities manifest great interest in OD. Therefore, many initiatives and platforms have been developed in order to publish open data sets in several different fields such as mobility, security (e.g. crime rates), economy (e.g. statistics on business creations) [23]. This interest in OD is due to the interpretation of Open Data as an essential tool for the dissemination of the Open Government principles [20, 23]. There is rich evidence stating that Open Government Data (OGD) has the potential to drive innovation [7, 20], not only because it allows an increasing level of transparency but also because it helps empower citizens and communities [20]. However, simply providing OGD does not automatically result in significant value for society [20]: the potential benefits of OGD [20, 29] will not be realized unless data are actually used. In truth, many data sets are available, but often repositories contain OD that users do not need and data sets that citizens need are not available (or not published) by Public Administrations

(PAs) [36]. By involving citizens, not only heterogeneous skills can be exploited, but also effective needs can be considered during the OD creation.

Thus, our research question is *how to support PAs and citizens (without any upper limit on the group size) in working together to publish high-quality OD.* By focusing on small groups (7-8 persons), they can exploit peer-to-peer methodologies without losing the overall picture of the rest of the group. We already had experience in managing small groups by an *agile* approach. In fact, SPOD (Social Platform for Open Data) supports an agile iterative, evolutionary, test-driven and collaborative methodology for the production of OD [13]. However, in environments in which the group size increases and there is a high diversity of partners and contributors, an *orchestrator* is needed in order to ensure valuable inputs and mitigate concerns from network actors [10]. The orchestration is a well-known strategy applied to large groups [17]. It ensures the creation and extraction of value, without the introduction of hard hierarchical authority [17]. Therefore, the participants do not work as equals but they can clearly define and distribute roles in such an *agile* way. If in the past the agile approach was considered suitable only for small groups, in the last years there is an increasing interest in exploit it also in large group management [18]. Moreover, McBride et al. [25] cite both the agile approach and the occurrence of different (motivated) stakeholders among the key factors of a co-creation process. Each stakeholder should play a specific role. Each role implies tasks and responsibilities. They must be distributed taking into account individual skills and the overall needs.

However, there is also a dark side of the orchestration: it is easy to produce data of low-level quality while working into a large group without a well-defined guide. In fact, it is easy to duplicate data or leave them incomplete when several different people are involved. The problem is raised by the difficulty to keep an overall vision of the whole data set. Also, data accuracy can be compromised if clear guidelines to produce data are not established in advance. *Completeness* and *accuracy* are two of the quality pillars [4]. Data quality issues are quoted among the principal barriers of complete exploitation of OD [7, 20, 34]. Data will cost too much to be transformed into a standard format [2]. For instance, poor data quality costs the US economy around $3.1 trillion a year [9]. Moreover, according to a survey conducted by TMMData and the Digital Analytics Association, nearly 40% of data professionals spend more than 20 hours per week accessing, blending, and preparing data rather than performing analysis. The situation becomes even more complex when the published data involved individuals' information. In that case, there is the need to ponder how to protect individual privacy to be compliant with the General Data Protection Regulation (GDPR or Regulation (EU) 2016/679) [1].

Our proposal is to scaffold PAs and citizens in working together by orchestrating the co-creation of OD in an agile way. Our approach is integrated into SPOD which already supports the agile co-creation of OD leading PAs and citizens in working together. The main *contribution* of this work is the introduction of *roles* into the OD creation process. Roles keep responsibilities and tasks clearly divided. Moreover, they lead to work orchestration. SPOD will guide participants

in easily identify tasks and responsibilities attached to played roles. Roles will be distributed according to the skills of group members and baring in mind the final goal, i.e. produce high-quality data. In order to satisfy quality requirements, one of the role (the *creator* of the data set under definition) can attach constraints and rules to each column in order to avoid trivial syntactic errors.

This paper is structured as follows: in Section 2 related work is reported; in Section 3 we present the motivational use case which inspires the proposed roles and their responsibilities; in Section 4 we detail the implemented approach and how it is embedded into SPOD; then the article concludes with the future directions and some considerations.

2 Related Work

Orchestration. Network-orchestration activities include ensuring *knowledge mobility*, *network stability*, and *innovation appropriability*, as well as coordination [30]. According to the context, these activities can be emphasized to different extents (e.g. highlighting knowledge mobility over appropriability) and can be carried out in quite different ways (e.g. by simply facilitating different activities). Multiple members may participate in these activities. Acknowledging the orchestrator roles is therefore relevant. In literature, several different types of orchestrators have been proposed. Roijakkers et al. [31] divided users into *orchestrators* and *non-players*: an orchestrator typically is an actor that has relatively strong individual incentives within networks and ecosystems that he/she aims to influence, while non-player orchestrator influences and supports the network without being an active competitor in the end market. Furthermore, the existing literature provides examples where roles and tasks are defined according to specific scenarios [19, 21, 27]. According to our motivating use case (which will be presented in Section 3), we define the *validator* role who is the legal manager of the data set content since he/she validates manually each row and adds to the data set only those semantically correct; the *creator* role who is the manager of data set constraints and defines the form to guide the data set filling; the *filler* role is in charge of populating the data set. The filler can be qualified as i) *advanced* if he/she can both populate the data set and have an overall vision of the whole data set; ii) *plain* if he/she can only suggest a new row to the validator without consulting the rest of the data set.

Co-creation by an agile approach. OD platforms can simplify the interaction among citizens and organizations giving them the opportunity to collaborate with government organizations. These platforms can be seen as collaborative environments which enable participation in collective decision-making efforts [32]. The subjects involved in this activity (e.g. citizens and PAs) work on open data set through a platform (SPOD in our case) splitting tasks by roles, respecting rules, and exploiting the community [32]. Once established the final goal (co-create high-quality open data), the operative approach must be chosen. In the last years, there is an increasing interest in dropping down the classical waterfall-like approach and adopting a more agile process [34]. Toots et al. [34, 35] propose

a framework for data-driven public service co-production. They observed that, in traditional waterfall-like models, public administrators are steering and controlling the whole process with citizen input being occasionally, but not necessarily, sought. Agile development focuses on being able to adapt quickly to changes by following an *agile* approach. Similarly, Mergel [26] points out that in traditional waterfall project management approach each phase sequentially follows the previous step. In contrast, an agile approach focuses on shorter development phases and continuous collaboration with final users in each phase. By the agile development it is possible to incrementally create, test, and improve products [26]. Every (intermediate) result can be immediately tested. By applying the agile methodology to the OD co-production, each data set can be iteratively and incrementally discussed and improved during the definition phase and it can be used in a practical use case to test on the way the fitness-of-use [13] (e.g. users can test the data set by creating visualisations).
The Agile methodology for software is iterative, incremental, and evolutionary [5]. Madi et al. [22] extracted a list of values out of the agile manifesto: *Collaboration, Communication, Working software, Flexibility, Customer-centric, Incremental, Iterative, Feedback, Speed, Simplicity, Self-organizing, and Learning.* All these values are taken into account in developing the SPOD features.

Data Quality Control. By a reactive quality control, users try to improve the quality of already published data sets and make them compliant with specific needs. Once the data set is provided, it is possible to perform data quality assessment which "*is the identification of erroneous data items* " [24]. It can be performed by data profiling techniques in order to collect statistics and information about data [28]. There are several works which analyse data set content and infer metadata and data types, from actual values [3, 11, 33]. SPOD is provided of a type inference mechanism: first we infer the data type for each value based on its content and, consequently, we attach a data type to each column [16]. Besides basic data types - like dates, numbers, and text - we infer also types related to personal information - such as Social Security Numbers (SSN), company codes, IBAN, gender, ZIP code and so on. The recognized data types are principally inspired by the personal data defined in the GDPR. *Personal data* is "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data,...* " [1]. By the type inference, we report both quality issues - accuracy, completeness - and privacy concerns - i.e. breaches of personal data in a textual description or due to the structure of the data set [16]. On the other side, guiding the OD creation by *proactive* quality assurance could reduce the subsequent effort in quality control. Our goal is to guide OD creation by a set of rules and constraints on values to avoid trivial syntactical errors. Since this approach does not prevent semantic mistakes, the *validator* role is in charge of verifying the correctness of data and deciding which rows should be dropped down and which ones become part of the data set under construction. According to the European Data Portal [14] (EDP) (at time of writing), 29,26%

are three-stars, and only 1,17% are four-stars data sets - referring to the *five stars rating system* [6]. A large number of three-star data sets on OD portals triggers the need to reduce their data quality issues. Therefore, we decided to focus on plain textual data without an attached schema.

3 Motivating use case

The need to distinguish among several roles and the identification of the proposed profiles are motivated by a concrete use case claimed by the Public Administration, the Council of the Campania Region in Italy, that we will name Motivating PA (MPA). Our MPA has established a Special Regional Committee since 2015 - named "Land of Fires" - which takes care of precise monitoring of the uncontrolled phenomenon of the occurrence of garbage scattered over a vast territory (90 municipalities between the province of Naples and Caserta). This monitoring takes place with the involvement of qualified stakeholders dedicated to the collection of both structured and non-structured data. These data concern not only the structural characteristics of the territory and its municipal resources dedicated to the problem of the rubbish fires but also the assessments and the indications of the operators about the usefulness (or not) of the legislative disposal that qualifies and encourages (also with money) the municipalities in this zone. The first objective of this commission was to verify the effectiveness of the application of the Regional Law (n. 20/2013) and several checks were carried out through the direct contact with 90 municipalities involved. Now, the ICT department of the MPA has been involved to streamline the process via automated tools. Several experiences were matured with questionnaires reported by using EU tools (EUsurvey). From the analysis of recovered data, the legislator will obtain valid tools to identify and implement more precise and timely intervention rules for the elimination of this dangerous phenomenon.

Nevertheless, the focus of MPA is now on the direct process of reporting these data as open data for citizens. The low-quality of collected data and the wide set of contributors must be tackled. In fact, actual tools only allow a simple "collect-and-send" data process. It is insufficient to support a complex mechanism of data collection, joint analysis and discussion, and publication as open data. For instance, previous experiences of required data collection were further elaborated through several successive meetings with other competent bodies and institutions in the field, with national government authorities, a list of chosen delegates from the major Municipalities involved and a series of public hearings held at the MPA site, where the (preliminary) results were presented and discussed.

The Special Regional Committee is now considering to extend the activities to a much wider audience of municipalities, involving all the towns of Campania. According to the plan to cover more than 500 towns, previous approaches are no longer sustainable. Thus, it is necessary to design a supportive environment that will guide the community in collecting, assembling, evaluating data for publication. Furthermore, automated tools for quality checking are needed. Because of the large number of participants, the orchestration is more suitable than peer

working in order to clearly separate roles and tasks. There should be the supervisor role - which can be declined as *validator* if he/she has to inspect data or as *creator* if he/she has to define constraints and rules on data. Moreover, there is the necessity to involve a big number of stakeholders who play the role of *filler*.

In conclusion, the introduction of roles and orchestration into a social platform (SPOD) is due to the necessity to coordinate a huge number of users leading them in creating high-quality OD. The research described here has been conducted in strict cooperation with the MPA officials and their ICT department. In the conclusions, we will report on the current state of the project.

4 Our orchestrated Open Data co-creation

Based on the motivational use case, we define the following orchestrator roles:

- *validator*: he/she is a super partes verifier. He/She is in charge of inspecting the content of the data set and discarding all rows conceivably semantically incorrect. Moreover, he/she is the legal manager of the data set;
- *creator*: this role corresponds to the expert in the field and/or who is able to opportunely model the data set under the definition. He/She is in charge of defining the structure of the data set and its columns specifying their data types and, if necessary, constraints on them;
- *advanced filler* who is in charge of filling in the data set and has the privilege of having an overall vision of the whole data set;
- *plain filler* who can only fill in the data set but cannot have a look of the other rows of the data set.

Each role is attached to a set of tasks. Starting from a data set, the *creator* has to define a form in order to bind a data type to each column and/or force some constraints on them. The starting point could be an empty or a partially filled in data set. The minimum data set to be used as a starting point has to expose the column header. By asking for the form creation, the creator is guided in filling in the form which can refer to all the columns or a subset of them. For each column the *creator* can choose among basic data types - such as *text*, *number*, *date* -, geo-coordinates, files - specifying between images and documents -, drop-down lists - also called *select* options. A select option can be manually populated by the creator. Otherwise, the tool offers some built-in select options, such as the list of all Italian regions, provinces or municipalities. Based on the data type, the form will guide the creator in specifying extra parameters, if necessary. For example, if the creator asks for a numeric value, he/she can also bind minimum and maximum values. It is also possible to specify constraints on values in order to automatically validate the syntax of values inserted by the filler, e.g. by selecting *email* the data set will prevent the insertion of syntactically wrong emails. Besides data types and constraints, the creator can also specify extra information, such as placeholders or tooltips, labels or descriptions, ask for mandatory fields or define a default value, which will help fillers in interpreting more easily which information should be inserted into the data set and in which format.

By correctly and deeply defining the form it is possible to minimize syntactical errors in the data set filling. Obviously, it does not prevent semantic errors. For example, by restricting the data type of a column to date the user will not be able to specify incorrect dates but there is no validation about its correctness.

The *filler* is in charge of filling in the form to populate the data set. The tool will force him/her to insert only syntactical valid inputs and prevent trivial errors which could compromise the overall data set quality. The distinction of *advanced* and *plain* filler is only on the visibility of the whole data set under construction. This distinction is due to security requirements: based on the situation there could be the need to involve a huge number of *filler* users. Since also unreliable people might be involved by accident, there is the need to avoid the suggestion of rows which can deliberately change the overall statistics of the data set. Therefore, we provide the opportunity to give access to the data set in reading mode only to reliable people - by providing them the *advanced* filler role - and preventing the access to others by the *plain* filler role. Therefore, the plain filler can only add rows to the data set without having the possibility to consult the already provided data. The advanced filler has the right to read - not modify - the whole data set under the definition.
To address also the semantic correctness of the data set, data suggested by the filler are not automatically added to the final data set. They are left in a *grey zone* until the validator check them.

The *validator* is the legal manager of the data set. Therefore, he/she has the power to decide which rows should be included into the final version of the data set, but he/she takes the responsibility of all the information which are included and also of those discarded. The validator has to manually inspect proposed data to filter out the wrong ones. Approved data will be moved from the *grey zone* to the actual data set under construction. By the validation step, also semantic errors are reduced. Syntactic and semantic checks assures high-quality OD.

In Fig. 1 the whole workflow is summarized: the creator defines the form starting from an empty or already partially filled data set; the filler proposes new candidate rows by filling in the form; the validator inspects the candidates rows and takes responsibility of the data effectively added to the data set under definition. Users are not forced by SPOD to follow these steps in this particular order. The creator can create the form at any time. The filler can start producing rows also before the form definition. The filler is not locked by the validator verification step. Therefore, users can choose the best operative approach according to their needs. In our *agile* approach, SPOD offers a set of tools (e.g. chat, co-creation rooms, form) which can be incrementally and iteratively exploited by users in any order and to any extent.

In Fig. 2 an example of form is reported. This is the layout of the form on the creator side. When the creator opens the form template, a box is created for each column. In this use case, the data set represents a citizen profile where the name, birthplace and birthdate, marital status and children number are reported. The creator decides to model the *name* as a string. By clicking on the plus icon on the right, the section delimited by a dotted line is opened. These

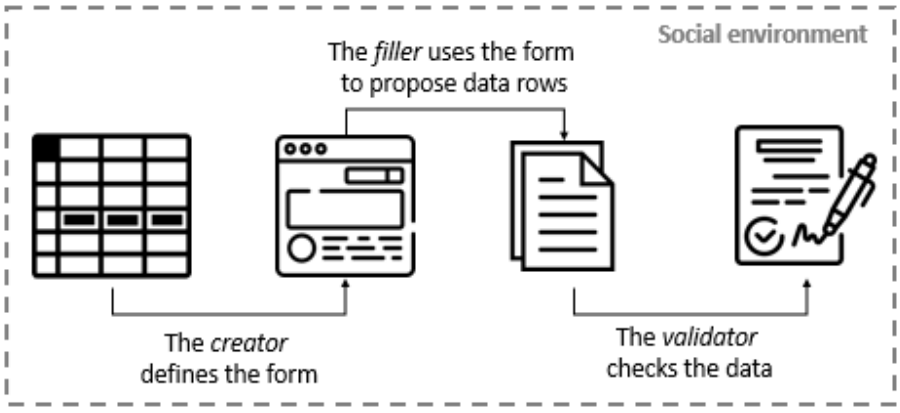

Fig. 1: Workflow of the orchestrated OD co-production process. Roles are reported in italic.

options will provide extra information to fillers during the data set populating step. The *Date of birth* is modelled as a date while the *Birthplace* is modelled as a Province. The latter represents an example of auto-filled select: the creator has to simply decide the type of the column as Province, while the filler will have access to all the available provinces. The *Marital status* column is a select filled by the creator. By choosing this type, the creator has to specify which are the valid options. The *Children number* is modelled as a number and it is possible to specify the minimum and maximum value.

The described process has been included into SPOD where citizens, PAs, associations, and every kind of stakeholder can create or join online communities of interests, discussing around OD and their visualizations [8, 12]. SPOD supports Data-Driven Discussions where citizens are engaged in participating in discussions of interest by using OD. It enables collaboration among users as a key aspect to ensure the creation of value form OD. Despite classical features of a social network - such as the wall with all the news, the possibility to share content, comment posts, chat - our platform is fully interoperable with existing OD portals. It implies that users can i) directly access data sets available on the associated open data portal; ii) create reusable visualizations; iii) share, use and reuse data sets and visualizations within the discussions in a seamless way. About privacy concern, our platform allows every public administration or organization to have its own instances for local communities. In this way, stakeholders and partners can take advantage in terms of effectiveness since a dedicated platform avoid misleading mixing of topics and helps in focusing on specific discussions. The platform provides the possibility to create visualization upon OD data sets; share and comment data sets and visualizations; conduct data-driven discuss in *agora*; create rooms of co-creation in order to create OD and share knowledge. Each participant can create a co-creation room to which other users can join. In these rooms, PAs and citizens - and different stakeholders

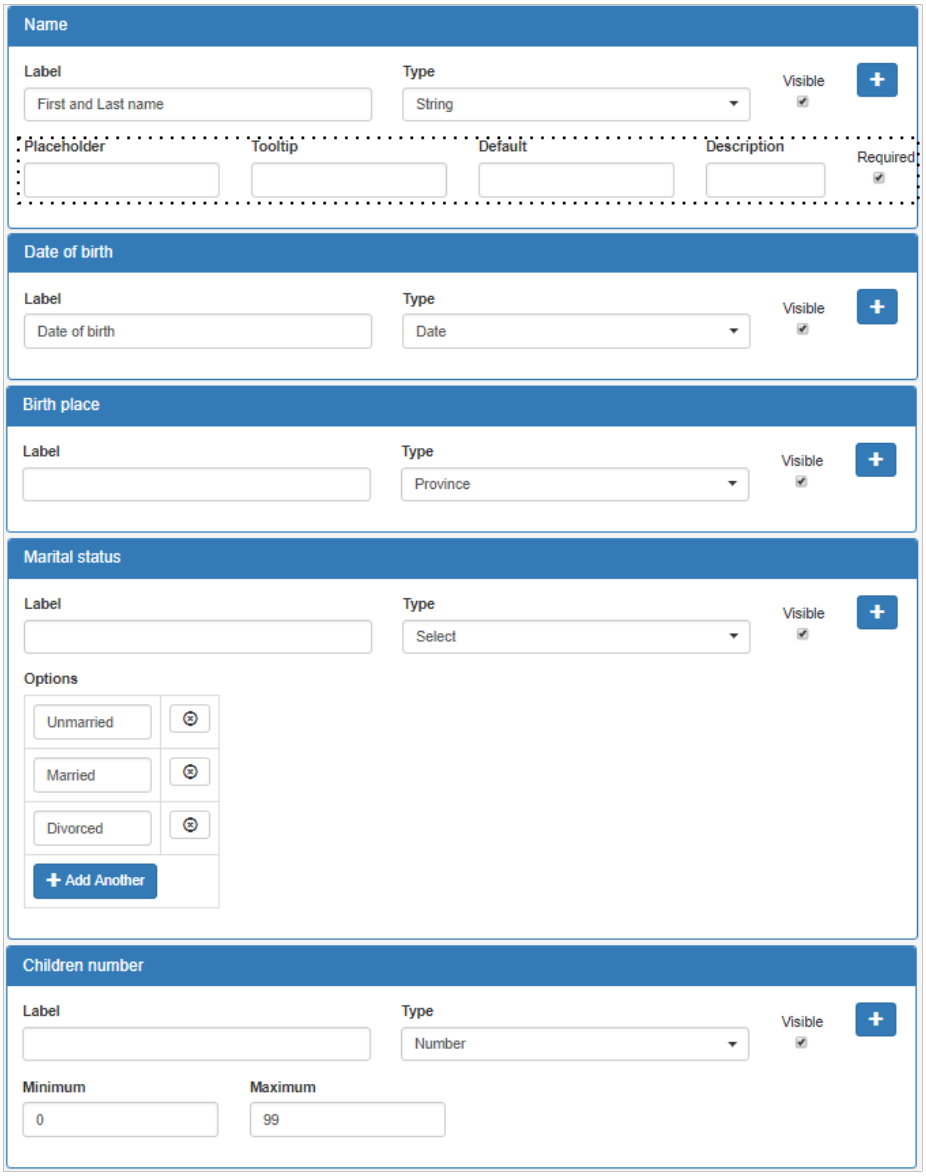

Fig. 2: Example of a form: for each column of the data set a box is provided. For each column, it is possible to specify the type, extra constraints, if necessary, and all the information which will guide the filler in the data set population step.

in general - can work together to create shared data sets. The focus is on 3-star data set - according to the 5-star rating defined by Tim Berners Lee [6]. Upon a data set, users who play the role of *creator* can define the form as described

before. Once confirmed the form, all the users who play the role of *filler* will be guided in filling in data set by the advised template. The form prevents the insertion of syntactical wrong data. Therefore, it represents a proactive quality assurance approach. Moreover, during the data set definition, every user can ask for a reactive *quality check* [16]. It is a set of tools to guarantee quality and avoid privacy issues. For each column, the quality check module infers the column data type by its actual content. Not only basic data types - such as string, date, number - are inferred but also types which try to catch the semantic value of the column content - such as region, province, municipality, name, surname, phone, email, SSN, IBAN and so on. Besides the type inference, the quality check module identifies typos and computes quality statistics considering the uniformity of column content and the completeness of values. About privacy issues, the same module detects if personal information is leaked into descriptive values and if the structure of the data set exposes a combination of information which could allow the unique identification of an individual. SPOD is online available on free at https://github.com/routetopa/spod. It can also be accessed by a mobile application. The latter can be particularly useful for *plain fillers* which can populate a data set in a practical and comfortable way simply accessing by their mobile and proposing new candidate rows. From a technological point of view, SPOD is released with an open source license. All the source code, as well as documentation, is published on GitHub at https://github.com/routetopa/spod.

5 Conclusions

Splitting the OD co-creation process into several different roles helps in detecting responsibilities for each role. According to personal skills, roles can be distributed. By the synergy of heterogeneous skills and profiles, high-quality data can be provided. In this paper, we have discussed how SPOD guides *creators* in defining a form to attach a schema (data types and rules) to data sets under the definition. This form prevents syntactic errors. Those semantic still need to be checked manually by the *validator*. The considered roles and the proposed approach are the results of cooperation with our Regional Administration MPA. We have now a working prototype that is actually being tested. We are, then, planning an evaluation phase on the field to verify if our solution completely satisfies the needs of the motivating use case, first with the previously contacted Municipalities (around 100) and then with the whole set of Municipalities in Campania. One of the future steps is to combine the proactive quality guide of the form with the reactive type inference module to help creators in defining a more complete data set profile in case of partially populated data set. The workflow will be that 1) the Creator defines a form starting from an already populated data set, 2) the *type inference* process is applied to infer the data types of each column, 3) the form is populated by the inferred data types, 4) the Creator can either adopt them or relax the suggested rules and constraints, 5) finally, the form is published and can be used by the Filler.

References

1. General data protection regulation (gdpr) (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj
2. Open data goldbook for data manager and data holders. Available on line (2016), https://www.europeandataportal.eu/sites/default/files/goldbook.pdf, European Data Portal. Last checked on January 1, 2019
3. Alur, N., Joseph, R., Mehta, H., Nielsen, J.T., Vasconcelos, D.: Ibm websphere information analyzer and data quality assessment (2007), http://www.redbooks.ibm.com/redbooks/pdfs/sg247508.pdf
4. Ballou, D.P., Pazer, H.L.: Modeling data and process quality in multi-input, multi-output information systems. Management Science **31**(2), 150–162 (1985)
5. Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., Thomas, D.: Agile manifesto (2001), https://agilemanifesto.org
6. Berners-Lee, T.: Linked data - design issues (2006), http://www.w3.org/Designlssues/LinkedData.html, last accessed on 2018/05/03
7. Chan, C.M.L.: From open data to open innovation strategies: Creating e-services using open government data. In: 46th Hawaii International Conference on System Sciences. pp. 1890–1899 (2013)
8. Cordasco, G., Donato, R.D., Malandrino, D., Palmieri, G., Petta, A., Pirozzi, D., Santangelo, G., Scarano, V., Serra, L., Spagnuolo, C., Vicidomini, L.: Engaging citizens with a social platform for open data. In: Proceedings of the 18th Annual International Conference on Digital Government Research. pp. 242–249 (2017)
9. Big Data, I., Analytics Hub: Infographics and animations: The four v's of big data (2017), https://www.ibmbigdatahub.com/infographic/four-vs-big-data
10. Dhanaraj, C., Parkhe, A.: Orchestrating innovation networks. Academy of Management Review **31**(3), 659–669 (2006)
11. Döhmen, T., Mühleisen, H., Boncz, P.: Multi-hypothesis csv parsing. In: Proceedings of the 29th International Conference on Scientific and Statistical Database Management. p. 16. ACM (2017)
12. Donato, R.D., Malandrino, D., Palmieri, G., Petta, A., Pirozzi, D., Scarano, V., Serra, L., Spagnuolo, C., Vicidomini, L., Cordasco, G.: Datalet-ecosystem provider (deep): Scalable architecture for reusable, portable and user-friendly visualizations of open data. In: 2017 Conference for E-Democracy and Open Government. pp. 92–101 (2017)
13. Donato, R.D., Ferretti, G., Marciano, A., Palmieri, G., Pirozzi, D., Scarano, V., Vicidomini, L.: Agile production of high quality open data. In: Proceedings of the 19th Annual International Conference on Digital Government Research. pp. 84:1–84:10 (2018)
14. European Commission: Open data portal (2017), https://www.europeandataportal.eu/data/it/dataset
15. European Data Portal: Protecting data and opening data (2019), https://www.europeandataportal.eu/en/highlights/protecting-data-and-opening-data
16. Ferretti, G., Malandrino, D., Pellegrino, M.A., Pirozzi, D., Renzi, G., Scarano, V.: A non-prescriptive environment to scaffold high quality and privacy-aware production of open data with ai. In: 20th Annual International Conference on Digital Government Research (2019)

17. Gausdal, A.H., Nilsen, E.R.: Orchestrating innovative sme networks. the case of "healthinnovation". Journal of the Knowledge Economy **2**(4), 586–600 (2011)
18. Gerster, D., Dremel, C., Brenner, W., Kelker, P.: How enterprises adopt agile structures: A multiple-case study. In: HICSS 2019 Proceedings (2019)
19. Hurmelinna-Laukkanen, P., Ntti, S.: Orchestrator types, roles and capabilities a framework for innovation networks. Industrial Marketing Management **74**, 65 – 78 (2018)
20. Janssen, M., Charalabidis, Y., Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Information Systems Management **29**(4), 258–268 (2012)
21. Klimas, P., Czakon, W.: Innovative Networks in Knowledge-Intensive Industries How to Make Them Work? An Empirical Investigation into the Polish Aviation Valley, pp. 133–157 (2014)
22. Madi, T., Dahalin, Z., Baharom, F.: Content analysis on agile values: A perception from software practitioners. In: Malaysian Conference in Software Engineering. pp. 423–428 (2011)
23. Martin, S., Foulonneau, M., Turki, S., Ihadjadene, M.: Open data: Barriers, risks and opportunities. In: In Proceedings of the European Conference on E-Government. vol. 58, pp. 301–309 (2013)
24. Maydanchik, A.: Data quality assessment. Bradley Beach, N.J.: Technics Publications, LLC (2007)
25. McBride, K., Aavik, G., Toots, M., Kalvet, T., Krimmer, R.: How does open government data driven co-creation occur? six factors and a perfect storm; insights from chicago's food inspection forecasting model. Government Information Quarterly **36**(1), 88 – 97 (2019)
26. Mergel, I.: Agile innovation management in government: A research agenda. Government Information Quarterly **33**, 516–523 (2016)
27. Nambisan, S., Mohanbir, S.: Orchestration processes in network-centric innovation: Evidence from the field. Academy of Management Perspectives **25**(3), 40–57 (2011)
28. Naumann, F.: Data profiling revisited. ACM SIGMOD Record **42**(4), 40–49 (2014)
29. OECD: Rebooting public service delivery - how can open government data help drive innovation? (2016)
30. Pikkarainen, M., Ervasti, M., Hurmelinna-Laukkanen, P., Nätti, S.: Orchestration roles to facilitate networked innovation in a healthcare ecosystem. Technology Innovation Management Review **7**, 30–43 (2017)
31. Roijakkers, N., Leten, B., Vanhaverbeke, W., Clerix, A., Van Helleputte, J.: Orchestrating innovation ecosystems imec. In: Proceedings of the 35th DRUID Conference (2013)
32. Ruijer, E., Grimmelikhuijsen, S., Meijer, A.: Open data for democracy: Developing a theoretical framework for open data use. Government Information Quarterly **34**(1), 45 – 52 (2017)
33. The Open Knowledge Foundation Ltd.: Library messytables link (2013), https://messytables.readthedocs.io/en/latest
34. Toots, M., McBride, K., Kalvet, T., Krimmer, R.: Open data as enabler of public service co-creation: Exploring the drivers and barriers. pp. 102–112 (2017)
35. Toots, M., McBride, K., Kalvet, T., Krimmer, R., Tambouris, E., Panopoulou, E., Kalampokis, E., Tarabanis, K.: A framework for data-driven public service co-production. In: Electronic Government. pp. 264–275 (2017)
36. Zuiderwijk, A., Janssen, M.: The negative effects of open government data - investigating the dark side of open data. In: 15th Annual International Conference on Digital Government Research. pp. 147–152