



Classification of Open Government Data Solutions' Help: A Novel Taxonomy and Cluster Analysis

Jonathan Crusoe^{1,2}(✉)  and Antoine Clarinval³ 

¹ Swedish Center for Digital Innovation, Department of Applied IT,
University of Gothenburg, Gothenburg, Sweden

jonathan.crusoe@ait.gu.se

² Swedish School of Library and Information Science, University of Borås,
Borås, Sweden

jonathan.crusoe@hb.se

³ Namur Digital Institute, University of Namur, Namur, Belgium

antoine.clarinval@unamur.be

Abstract. Open Government Data (OGD) pose that public organisations should freely share data for anyone to reuse without restrictions. However, the rawness of this data proves to be a challenge for data or information seekers. OGD-based solutions, such as interactive maps and dashboards, could help seekers overcome this difficulty and use OGD to satisfy needs, helping them to work effectively, solve problems, or pursue hobbies. However, there are several challenges that need to be considered when designing solutions, such as seekers wanting to solve problems rather than consuming information and aiming for quick wins over quality. Previous research has classified OGD solutions, focusing on general concepts. The next step is to reveal helpful patterns in OGD solutions, helping seekers. This paper presents a taxonomy with 24 criteria to classify these patterns. It was tested on 40 OGD solutions, and the resulting classifications were grouped in a cluster analysis, identifying 16 key criteria and 6 clusters. The clusters are (1) simple-personalised, (2) proactive multi-visual, (3) lightly-facilitated exploration, (4) facilitated data-management, (5) facilitated information exploration, and (6) horizon solutions. One unexpected finding is that helpful patterns do not cluster following themes, types, or purposes of solutions. Another finding is that the importance of key criteria varies between the clusters.

Keywords: Open Government Data · solution · taxonomy · classification · cluster analysis · information behaviour

1 Introduction

Open Government Data (OGD) pose that data from public organisations should be made freely available for anyone to reuse. These data need to be used to unlock

benefits [19] or risk being a costly burden [14]. The potential benefits of OGD include improvements in accountability, value creation, and service development [29]. People who could gain from these benefits are **seekers** who utilise data or information (content) in their everyday life [30] to satisfy various needs [5]. OGD is most often made available through online portals in raw formats such as CSV. The rawness of OGD can make them difficult to use for any meaningful purpose [35]. Therefore, OGD solutions (e.g., interactive maps and dashboards) have a key part in helping seekers understand and act on OGD [16]. **Help** means to make it possible or easier for someone by doing part of the work or by providing, for example, advice or support [33]. OGD solutions could help seekers work effectively, solve problems, or pursue hobbies [25]. However, a challenge is to design solutions to help seekers who tend to prioritise ease of access and use over quality, aim for quick wins, and can find it difficult to express their needs, asking for the wrong content [3, 8, 25]. Consequently, it is important to consider seekers when designing OGD solutions [25, 27].

Previous OGD research has attempted to classify OGD solutions to understand reuse. For example, by their ability to transform data into information [6], as services along criteria like data, themes, and topics [9], and by domain and features [20]. Janssen and Zuiderwijk [17] analysed solutions from a business model perspective with a focus on the source of value and Crusoe [5] identified 23 ways OGD solutions may help seekers. These classifications provide general ideas about the possibilities of solutions being helpful for seekers but open questions about how this helpfulness has been achieved in the design of OGD solutions. As a result, it is time to take the next step to reveal possibly helpful patterns in the design of OGD solutions. We define a **helpful pattern** as the combination of help provided by a solution for a seeker. A helpful pattern could be that a solution acquires, filters, and visualises data for a seeker [see 10].

The paper's objective is to construct a taxonomy (a classification of empirical entities [2]) for helpful patterns in the design of OGD solutions. Hunke et al. [15] explain that a taxonomy can be used to design new solutions by revealing their "anatomy" and key features or properties. Rizk et al. [28] add that a taxonomy can bring understanding to key aspects of utilising data in the design and delivery of solutions. As criteria, our taxonomy must cover the seekers' needing, seeking, using, and distributing of content [5] and be able to classify a broad range of OGD solutions [e.g., 6, 17]. We started the research by synthesising a tentative taxonomy from previous research, which was then refined through iterations of classifying 40 OGD solutions. The research ended with a cluster analysis of these solutions, helping to test the taxonomy and identify clusters of helpful patterns and key criteria. This paper contributes towards explaining how OGD solutions can be designed to be helpful for seekers, and as such realise benefits through the satisfaction of needs.

2 Related Works

In their daily non-professional life, seekers frequently encounter problems that they solve by seeking information related to, e.g., healthcare or hobbies.

According to Savolainen [30], the behaviour undertaken to solve these problems comprises three steps.

First, the evaluation of the importance of the problem. Second, the selection of content sources, such as people, libraries, and digital solutions [32]. Recently, the potential of OGD as a content source has been studied in the context of seekers' everyday life [18]. However, the rawness of OGD makes it hard to directly use to solve a problem, which is a challenge for seekers. As a result, solutions based on OGD are being developed to help seekers. Third, the seekers seek orienting and practical content, which can, for example, be done through active seeking, active scanning, non-directed monitoring, and by proxy [23]. This paper focuses on OGD solutions as a content source, emphasising the third step of [30], namely how OGD solutions help seekers seek information. As a result, we want to understand the design of OGD solutions in relation to seekers. Previous research has classified OGD solutions from three broad perspectives: (1) provision, (2) solution, and (3) usage.

The provision perspective focuses on actors as providers of solutions in some contexts. Gebka and Castiaux [11] have identified roles taken by public organisations, projecting expected roles onto the seekers. The classification of Davies [6] grouped solutions as the ability to transform data into facts, data, information, interfaces, and services. Janssen and Zuiderwijk [17] viewed solutions from a business model perspective, classifying them as single-purpose, interactive, information aggregators, comparison models, repositories, and service platforms. Azkan et al. [1] covered several criteria, such as main value, data types, and payment mode. Similarly, Paukstadt et al. [26] provided criteria like payment mode, pricing model, and value proposition.

The solution perspective has its focus on describing solutions. Foulonneau et al. [9] arranged solutions following criteria like data, themes, and topics. Mainka et al. [20] covered, for example, features and type. Hunke et al. [15] included criteria, such as data generator, data target, and analytic type. On the other hand, Rizk et al. [28] used criteria like data acquisition mechanisms, data exploitation, and insights utilisation. They identified three solution groups: distributed analytic intermediaries, visual data-driven services, and analytic-embedded services. Shneiderman [31] understands criteria as tasks solutions can help seekers with, such as giving an overview of, zooming in on, and filtering content.

The usage perspective approaches solutions from the view of seekers. Virkar et al. [34] classified seekers' usage of legal information solutions, which can be to compare laws and follow legal developments. Crusoe [5] conceptualised solutions following four behaviours of seekers: needing, seeking, using, and distributing content. A solution can help seekers encounter needed content, but also formulate their needs (*needing*). It can also help them find or discover content (*seeking*) while making it easier to understand by representing data, supporting interpretations or adapting its help (*using*). The solution can enable seekers to share or spread content (*distributing*). However, previous classifications of OGD solutions do not explain how functions and properties can be combined into helpful patterns to help seekers satisfy their needs for content. It can, as such, be difficult

to construct complete designs for OGD solutions and explain how these designs can satisfy seekers' needs.

3 Research Approach

This research constructed a taxonomy for helpful patterns in the design of OGD solutions. It followed a qualitative artefact study using a qualitative approach [12]. An artefact study generates empirical material about solutions' functions and properties but provides limited information about whether they produce desired results for seekers [12]. However, a qualitative approach gives a deeper understanding of the solutions [24], helping to refine the taxonomy, which is our motivation for following this approach. The taxonomy is made for constructed types, a set of criteria with empirical reference that serve as the basis for comparison of empirical cases [2]. These criteria were the reasons for grouping solutions [21], referring to their functions and properties. We decided to use binary criteria for whether a pattern had a certain help or not. It made the taxonomy possess more criteria but allowed for freer identification of patterns and a reduction in interactive complexity among criteria. This choice also enabled the calculation of objective similarity levels between the helpful patterns [2]. This research followed four steps, iterating between the second and the third: (1) construct an initial taxonomy, (2) select OGD solutions, (3) classify solutions and refine the taxonomy, and (4) test the taxonomy with cluster analysis. The iterations aimed for saturation in the construction of the taxonomy, meaning further data collection no longer sparked new insights nor revealed new criteria [4].

First, we discussed previous research that could help to construct a taxonomy based on previous knowledge. We decided to start by synthesising previous work from multiple fields (e.g., Human-Computer Interaction, Information Behavior, Open Government Data), using [23,30,31], and [5]. Individually, researchers created a conceptual map of how concepts and previous research could be related, which was discussed among them afterwards. The discussion resulted in a tentative list of 25 criteria. Each criterion was named and provided with inclusion criteria and examples. If necessary, exclusion criteria were formulated. Furthermore, a [conceptual tree diagram](#) was created to support the classification process. At the centre was a general question (i.e., How is the pattern helping the seeker?), which was then divided into more specific questions with the leaves as the criteria. A researcher could follow and answer these questions to identify applicable criteria. When classifying a solution, colouring the leaves gave an overview of the solution's helpful pattern.

Second, we retrieved a list of 74 solutions identified in [5], enabling us to test the conceptualisation of [5] and provide new insights into previously studied solutions. We chose this list since it was easily accessible and known to contain relevant solutions. Following purposive sampling [7], we started with the solutions presented as good examples, believing them to be easy to classify and have clear helpful patterns. Then, we selected solutions based on the perceived ability to be a negative case or verification, helping to refine the taxonomy. However, some solutions were no longer active, as such we attempted to access them through the Wayback Machine, bringing back six solutions.

Third, the classification started with a small set of agreed solutions. Individually, we classified these solutions using the criteria list and the conceptual tree diagram. We tested each solution, classifying a helpful pattern. We discussed our classifications and underlying reasons, refining the taxonomy and correcting any errors. This step started with classifying a few solutions in quick iterations, allowing for rapid refinement of the taxonomy. When the taxonomy became stable, we increased the number of solutions to classify within one iteration. We reached saturation once 40 solutions had been classified. The taxonomy reduced analytical drift and the sharing of classifications allowed for cross-checking, contributing to research reliability [4]. The discussions of the taxonomy and classifications allowed reflexivity for the researchers, contributing to research validity [22]. At the end of this step, 40 helpful patterns had been classified with a taxonomy of 24 criteria. 11 criteria differed from the tentative taxonomy from the first step.

Fourth, we used cluster analysis to test the taxonomy, aiming to cluster solutions into homogeneous groups based on similarities in their helpful patterns [2]. It is important that criteria help us to group and differentiate between solutions [2, 21]. The intent is to minimise differences between solutions within a group while maximising differences between groups [2]. Following this reasoning, we started this step by removing any criteria that we considered too common or uncommon amongst the helpful patterns, as they do not help us differentiate between solutions. We used subjective thresholds of 0.2 and 0.8 (i.e., corresponding to 20% and 80% of classified solutions having a given help), identifying 16 of 24 criteria as key. We applied Gower and Legendres' S9 method to calculate a distance matrix for the criteria, as it is made for binary data and provides high resolution [13]. We then applied divisive cluster analysis [2], which results were visualised as a dendrogram, helping us to determine a cluster number of 6. A **cluster** represents a group of solutions with similar helpful patterns. In order to interpret the clusters, we created a heatmap to represent the proportion of criteria amongst these clusters. Each tile in the heatmap presents the proportion of help for a given cluster. We removed any tile with a value between 0.2 and 0.8 to highlight similarities and differences of the clusters, clarifying any particularities. We then studied the helpful patterns and clusters. If any group contained an odd or puzzling combination of solutions, we revisited the classifications and verified them, helping to reduce errors further. This approach to taxonomy construction has helped to validate the final taxonomy, as it has been tested on a heterogeneous sample of OGD solutions by two researchers.

4 Results

4.1 A Taxonomy for OGD Solutions

The taxonomy comprises 24 criteria and is presented in Table 1 and 2. For each criterion, the definition, examples, the proportion of the 40 classified solutions checking the criteria, and whether it is selected as key or not are indicated. The key criteria are those with a proportion between 0.2 and 0.8. There are 16 key criteria, representing meaningful similarities and differences.

Table 1. Criteria for the taxonomy of helpful patterns. For each criterion, the definition, examples, proportion of the 40 classified solutions checking the criteria, and whether it is selected as key or not are indicated (Part 1, continued in Table 2).

	Criteria and examples	Proport.	Key	
NEEDING	ENCOUNTER	Setup – Proactive encounters where data or its presentation has an initial structure upon arrival. <i>For example: start by presenting data for a given year or how data is grouped before the user</i>	0.550	■
		Suggest – The solution suggests content to the seeker. <i>For example: Top rankings, alternatives, recent content, and “read more here”-links</i>	0.525	■
		Reveal – Proactive encounters as highlights or conclusions of data. The solution conveys to the seeker an interpretation of the data. <i>For example: articles, blog posts, or content panels</i>	0.375	■
	FORMULATE	Nudge – The seeker is nudged towards certain actions. Not instructions nor manuals. <i>For example: “Click on a country” or search field with example keywords. It can also be more subtle like a label to explain what type of content to input</i>	0.375	■
		Answer – Presents a question that the pattern aims to answer. It can be intertwined with navigational structures. <i>For example: “What energy do we produce in the EU?” answered with text, statistics, and diagrams</i>	0.175	
SEEKING	Border – The seeker can draw borders between all content and some content through simple keywords, categories, or parameters. Go beyond predefined divisions. <i>For example: dashboards and filters</i>	0.950		
	Sift – The seeker has access to all or most of the data. <i>For example: a list of data that can be scrolled or various categories with data</i>	0.775	■	
	Detail – The seeker can request more information about some content amongst other content. It is not clicking on items in a search result list. <i>For example: interactive map where locations can be selected to show some of its information</i>	0.750	■	
	Herd – Data are stored in several datasets, which can be connected to the navigational structure. The seeker manages content as buckets or groups, which can come in categorical hierarchies. The seeker can make clear distinctions and selections between datasets. <i>For example: open data portals or interactive maps with data inventories</i>	0.750	■	
	Pull – Data are pulled out from a “black box” using various functions. It is not a filter functionality, rather the seeker does not know the large dataset behind the presented data. <i>For example: search bar, showing results</i>	0.500	■	
	Traverse – The seeker traverses the relationships within the content. Content is similar to a web, network, or layers of aggregation. Data within dataset(s) can follow a hierarchical structure. It is possible that several datasets are combined to produce this effect. It is not tagged data. <i>For example: animal taxonomies/typologies, some aggregated statistics, politicians and their parties within a governing body</i>	0.400	■	

Table 2. Criteria for the taxonomy of helpful patterns. For each criterion, the definition, examples, proportion of the 40 classified solutions checking the criteria, and whether it is selected as key or not are indicated (part 2).

	Criteria and examples	Proport.	Key
REPRESENT	Facts – The content presents facts and information to the user. <i>For example: tables and summary numbers</i>	1.000	
	Comparative – The content makes similarities or differences clearly visible. <i>For example: bar charts and map charts</i>	0.825	
	Relative – Content reveals relations, occurrences, densities, relativities, or concentrations. <i>For example: interactive maps, word clouds, or heat maps</i>	0.625	■
	Movement – The content highlights the transfers of data between nodes or changes in data over time. <i>For example: Sankey diagrams, line charts, or animations</i>	0.525	■
USING SUPPORT	Clarify – Explanations and descriptions of content, easing the understandability. <i>For example: change language, colours to highlight weight or importance, explanations of terminology or patterns, or about pages</i>	0.950	
	Facilitate – Make social communication or community building possible or easier. Clear socio-technical purpose. It can be donations, but not pure contact information to developers without a clear purpose. <i>For example: forums or contact information to politicians</i>	0.750	■
	Elucidate – Makes the content clearer, enabling the drawing of different conclusions from the same content. <i>For example: side-by-side visualisations of the same data and map layers</i>	0.350	■
ADAPT	Personalise – Personalise the offered help by changing content or providing personalised content. Add customisation to the offered help. <i>For example: drawing tools, measuring tools, or asking for notification on changes in content, put items in a basket</i>	0.400	■
	Acquire – The seeker can upload or request content. <i>For example: upload map layers or request evaluation of products</i>	0.175	
	Record – The actions of the seeker can be undone, replayed, or viewed. A history is kept of the seeker’s actions. A seeker could save states. <i>For example: button to undo the last action</i>	0.125	
DISTRIB.	Extract – The content can be extracted from the solutions. <i>For example: download a visualisation or underlying data; or print a map</i>	0.825	
	Refer – The seeker can refer to specific content, which is still part of the solution. <i>For example: share a link to it or over social media</i>	0.675	■
	Embed – Other solutions can embed the solution’s content. <i>For example: iframes or API</i>	0.300	■

4.2 Divisive Cluster Analysis

Only the 16 key criteria were used in the divisive cluster analysis. In total, the analysis returned 6 clusters represented as a dendrogram in Fig. 1. It visualises the distances between helpful patterns based on their similarities and differences. Solutions belonging to the same cluster present similarities in their helpful patterns and are coloured alike. Cluster 1 (colored in red) groups 6 solutions. Cluster 2 (chartreuse) has 7 solutions, Cluster 3 (salmon) has 4, Cluster 4 (blue) is the most populated cluster with 12 solutions, Cluster 5 (purple) has 4, and Cluster 6 (orange) has 7. Table 3 lists the 40 classified solutions sorted by cluster and the checked criteria for each.

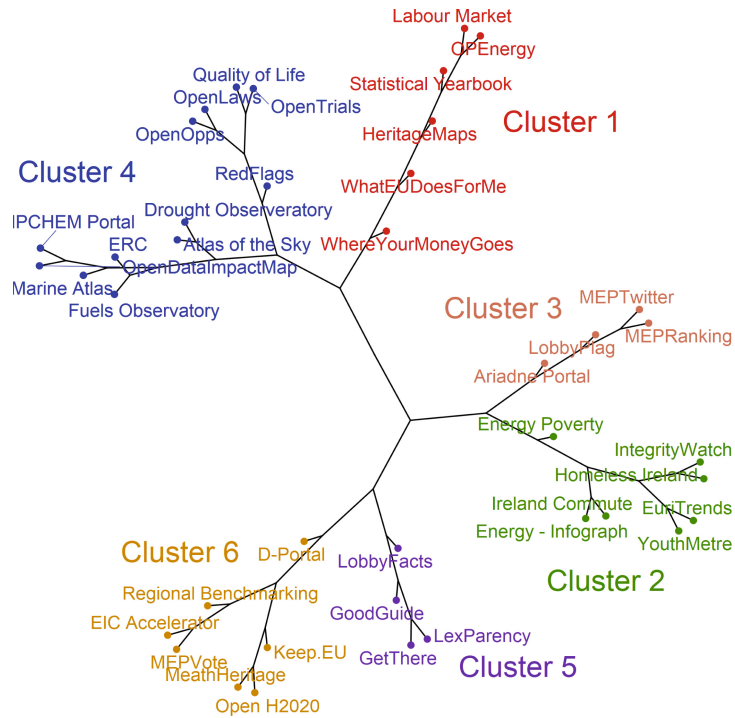


Fig. 1. Dendrogram resulting from the divisive cluster analysis. (Color figure online)

In order to interpret these clusters, it is necessary to analyse the proportion of each criterion, at the level of each cluster. Figure 2 presents the key criteria proportions amongst the six clusters. It shows, for example, that solutions in Cluster 1 always have “Herd” and never “Facilitate” and have a high (resp. low) likelihood to possess, for example, “Detail” (resp. “Embed”). The most demanding cluster in terms of criteria to check is Cluster 3. Its solutions must possess 9 criteria. On the contrary, no criterion is a must-have in Cluster 6, but several criteria have a high likelihood.

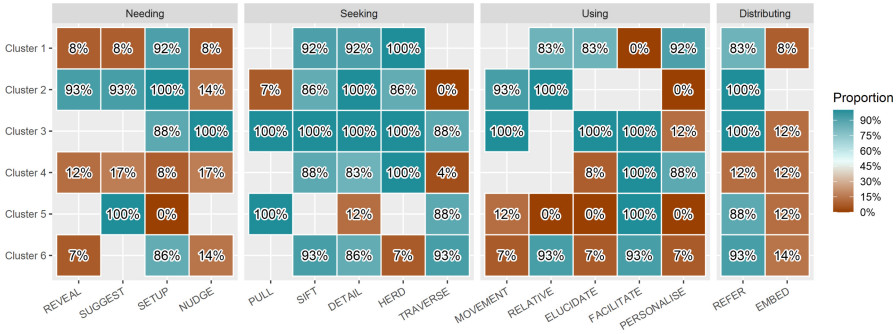


Fig. 2. Key criteria proportions amongst clusters.

Given the proportions of the key criteria, the following interpretations can be given for the six clusters.

Cluster 1 – Simple-personalised help. Solutions in this cluster follow a simple pattern compared to other clusters, as they seldom allow seekers to embed content and give limited help to formulate needs or encounter content. They focus on providing personalised information and visualisations, becoming a base or frame for interpretation [5]. They can help seekers see data from various perspectives, as such draw different conclusions.

Cluster 2 – Proactive multi-visual help. These patterns use various ways to visualise data as part of one or more datasets. The patterns are proactive or at least active, seeking to satisfy the seeker’s need for data or information [5], meaning they can provide conclusions or guide a seeker’s attention towards meaningful insights. This cluster matches visual data-driven services from [28], which visualise data and use storytelling to communicate insights to seekers. However, storytelling is less emphasised in our taxonomy.

Cluster 3 – Lightly-facilitated exploration help. The third cluster patterns aim to help the seeker explore its datasets from multiple perspectives. Some of these patterns allow the seeker to explore relationships within a dataset or details about data. They have some degree of facilitation where the seeker can provide feedback or ask questions.

Cluster 4 – Facilitated data-management help. Solutions in this cluster help seekers manage some larger dataset(s) while facilitating social interactions. While the patterns tend to allow for personalisation, the visualisations are often simple with limited ability to distribute. They seldom reveal any highlights or conclusions in the data.

Cluster 5 – Facilitated information exploration help. These helpful patterns have a limited ability to visualise data. Instead, they focus on information and any related internal connections. This information is often socially complex, such as health information about products, lobbying in the EU, and coordination of lift sharing. Social facilitation can range from community building to feedback or Q&A. While some level of dialogue [17] is possible, the patterns may

Table 3. List of classified solutions and checked criteria (* denotes key criteria).

Name	Reveal*	Suggest*	Setup*	Answer	Nudge*	Pull*	Stim*	Detail*	Border	Head*	Traverse*	Facts	Comparative	Movement*	Relative*	Elicitate*	Clarify	Facilitate*	Acquire	Record	Personalise**	Refer*	Embed*	Extract
Cluster 1	WhereYourMoneyGoes		■			■	■	■	■	■	■	■	■	■	■	■	■							■
	HeritageMaps			■		■	■	■	■	■	■	■	■	■	■	■	■		■	■	■		■	■
	WhatEUDoesForMe		■	■			■	■	■	■	■	■	■	■	■	■	■					■	■	■
	Labour Market			■			■	■	■	■	■	■	■	■	■	■	■					■	■	■
	OPEnergy	■	■				■	■	■	■	■	■	■	■	■	■	■					■	■	■
Cluster 2	YouthMetre	■	■	■			■	■	■	■	■	■	■	■	■	■	■						■	■
	Energy - Infograph	■		■			■	■	■	■	■	■	■	■	■	■	■						■	■
	Ireland Commute	■	■	■				■	■	■	■	■	■	■	■	■	■			■			■	■
	EnergyPoverty		■	■	■			■	■	■	■	■	■	■	■	■	■						■	■
	EuriTrends	■	■	■	■			■	■	■	■	■	■	■	■	■	■			■			■	■
	Homeless Ireland	■	■	■				■	■	■	■	■	■	■	■	■	■						■	■
IntegrityWatch	■	■	■				■	■	■	■	■	■	■	■	■	■						■	■	
Cluster 3	Ariadne Portal	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■				■	■	■	■
	MEPRanking				■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	MEPTwitter		■	■			■	■	■	■	■	■	■	■	■	■	■						■	■
	LobbyPlag	■	■	■			■	■	■	■	■	■	■	■	■	■	■						■	■
Cluster 4	Atlas of the Sky		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■				■	■	■	■
	OpenOpps			■	■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	Marine Atlas		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	IPCHEM Portal	■	■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	OpenTrials				■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	Fuels Observatory		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	ERC				■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	OpenLaws				■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	RedFlags	■				■	■	■	■	■	■	■	■	■	■	■	■						■	■
	OpenDataImpactMap	■	■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
Cluster 5	GetThere		■		■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	LexParenity		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	GoodGuide	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	LobbyFacts	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■						■	■
Cluster 6	Keep.EU	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■						■	■
	MEPVote			■		■	■	■	■	■	■	■	■	■	■	■	■						■	■
	EIC Accelerator		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	Open H2020		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■
	MeathHeritage		■	■		■	■	■	■	■	■	■	■	■	■	■	■						■	■
	D-Portal	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■					■	■	■
Regional Benchmarking		■			■	■	■	■	■	■	■	■	■	■	■	■						■	■	

encourage it to be outside the solutions, for example by providing contact information or sharing content on social media.

Cluster 6 – Horizon solutions. These solutions share most key criteria, but there is little agreement. They are complex to some degree and specialised. This cluster indicates that there are more clusters to be identified. It could also signal innovative designs, as help is combined in new or unique ways. Similar to [28], our taxonomy does not address the structure of helpful patterns. This cluster could be a result of this limitation, and as such, opens new avenues for future research.

5 Discussion

5.1 Novelty of the Contributions

Previous research had classified solutions, following a provision [e.g., 1,6,11], solution [e.g., 9,28,31], or usage [e.g., 5,34] perspective. Our research bridges the latter two, meaning we attempt to classify features and properties from the perspective of seekers. This approach makes our research original within the context of previous OGD research. While it is similar to [34], which focus on classifying seekers, our work is oriented towards solutions. Our research is a step towards designing OGD solutions that can be helpful for seekers but also to evaluate how OGD solutions have attempted to help seekers. It opens questions about possible matches and mismatches between seekers and solutions. Moreover, we constructed a novel taxonomy comprising 24 criteria able to classify helpful patterns in the design of OGD solutions. It enabled the identification of 6 clusters among 40 helpful patterns. 16 of the taxonomy’s 24 criteria were key in differentiating and understanding these clusters. The successful identification and interpretation of helpful pattern clusters serve as a test of the taxonomy.

Most of the initial criteria provided in [5,23,30,31] were identified to some degree amongst the helpful patterns. Analysing the proportion of each gives interesting insights into how OGD solutions currently help seekers. First, the rarity of the criteria varies (e.g., 17.5% of solutions have “Acquire” and 82.5% have “Comparative”). Nonetheless, the proportions are mostly included within the 0.2–0.8 range, meaning that most of the criteria are neither too common nor too uncommon. Such criteria can be found in all four categories (i.e., needing, seeking, using, and distributing). This shows the diversity among OGD solutions and reinforces the need for a detailed taxonomy to characterise how they help data and information seekers. Second, the proportion of the criteria within the “Needing” category varies between 0.175 and 0.55. At the same time, the proportions of the criteria in the “Seeking” category and of those related to visualisations in the “Using” category are overall higher. This difference indicates that while OGD solutions help seekers look for and use information, few of them help seekers encounter information or formulate needs for information. Third, few solutions allow for personalising (40%) help and content, and even fewer (12.5%) satisfy the “Record” criteria, which has long been recommended in the literature [31]. Fourth, only 30% of OGD solutions allow seekers to embed the solution’s content into other solutions. This lack indicates that it is difficult to build OGD solutions based on other OGD solutions, which would be another approach to increasing the value of OGD instead of working with the raw data directly.

An unexpected finding is that solutions perceived to have similar themes, types, or purposes can be designed following different helpful patterns. We expected similar solutions (e.g., OGD portals and interactive maps) to form clusters or at least follow the classes identified in [17]. It adds to our understanding of taxonomies specialised towards certain fields [e.g., 1,26,28] by explaining why digital solutions, like OGD solutions, can be difficult to classify. For example,

Janssen and Zuiderwijk [17] classified solutions as single-purpose, interactive, information aggregators, comparison models, repositories, and service platforms, while [6] grouped them as facts, data, information, interfaces, and services. It is possible to identify some of these classes among our classified OGD solutions, but none of the identified clusters represents them. There is, as such, a possible disconnect between the combinations of functions and properties that can help seekers and the themes, types, or purposes of solutions. Therefore, the application of helpful patterns could be a new fruitful approach to studying and designing solutions for seekers.

We identified 16 key criteria with proportions ranging between 0.2 and 0.8. It led us to another finding, as the key criteria do not play an equally important role in each cluster (see Fig. 2). For example, Cluster 1 has criteria regarding visualisations with varied expressions, while “Herd” is required. In contrast, Cluster 5 has “Herd” with varied expressions, while criteria regarding visualisations are mostly non-existent. It means that to understand and study some solutions, certain criteria come into the foreground, while others are in the background. The focus may be difficult to make based on the perceived similar themes, types, or purposes among solutions, as the helpful patterns may be different. Consequently, taxonomies with few criteria [e.g., 6, 17, 28] may attribute importance to properties and functions that can be relevant for some solutions, but not others. This finding gives us a new insight into the complexity of solutions, but also possible limitations in classifying them.

5.2 Implications of the Contributions

The two contributions of this research, that is, the taxonomy and the six clusters identified from the 16 key criteria, have utility for researchers and for practitioners. In general, they can be used to describe and analyse helpful patterns in existing solutions. They can also be used to design them. **Researchers** can use the taxonomy to guide data collection or support the analysis of solutions. The 6 clusters can act as the basis for empirical comparisons, providing a stepping stone towards theory development. The 16 key criteria can guide the researcher’s attention towards functions and properties that are important to differentiate between solutions but also help to identify functions and properties important to consider when studying specific solutions. **Public organisations** providing OGD can apply the taxonomy as well, on their OGD portal. This would help them to understand how the solution can help seekers, revealing potential areas of improvement. The identified clusters can give them an idea of what solutions could be built from the provided OGD, which can inform relevant help features to include in the OGD portal. **OGD reusers** can use the taxonomy as a basis to brainstorm about innovative designs, helping them to consider important areas. The reusers can also use the taxonomy to evaluate solutions, as it opens to identifying any possible deficiencies or impediments.

However, it must be noted that the utility of the taxonomy depends on the complexity of the solutions being classified. Solutions with tightly related properties and functionalities produce better classifications, while solutions with

varied dynamic content (e.g., blogs or descriptions in metadata; allowing for unpredictable variations in help) or specialised parts (e.g., a solution that has a forum, a dashboard, and a news section) can lead to unbalances or gaps in the classification. On the other hand, some of the classified OGD solutions had properties or functionalities difficult to detect for the authors (e.g., hidden within multiple layers of menus or small icon buttons at unexpected locations), which led to classification errors needing to be discussed among researchers. Therefore, we recommend that classification is done independently by at least two individuals and discussed afterwards to lower analytical drift and support reflection. Some errors in our classification also emerged from conceptual unclarity arising from misinterpretable functions or properties (e.g., a text field is presented as a search bar, indicating “Pull”, but was used to filter a list of items, as such being “Sift”). It is, therefore, important to understand functions and properties by how they attempt to help the seeker rather than how they describe themselves. Moreover, in our cluster analysis, we could not find any cluster that matches one pattern mentioned by [5]: contextualisation of help to the life of a seeker. Rather, it is spread out over several clusters, meaning the taxonomy may need further refinement towards capturing properties and functionalities that work to contextualise data or information.

5.3 Limitations and Future Research

The research presented in this article has several limitations. We used subjective thresholds for the key criteria and heatmap tiles in Fig. 2, meaning that other clusters may be identifiable among our classified solutions. However, after inspection, the identified clusters contained similar helpful patterns, giving important insights into the helpfulness of OGD solutions. Moreover, a delimitation is that solutions with dynamic content can introduce criteria while being difficult to detect and classify. It relates to a limitation of the taxonomy, as it is not constructed to handle the structure of patterns. If a solution offers different help at various locations, its classification presents these as equally important and related, which is a future research avenue. While the taxonomy construction reached saturation, the taxonomy was only tested with 40 classified solutions. Future research could apply the taxonomy to a larger sample of solutions, going beyond those previously identified by [5], giving insight into how OGD public organisations and reusers have tried to help seekers, but also identify missed opportunities and innovative designs. Another avenue could be to evaluate helpful patterns involving seekers.

6 Conclusion

Our main contribution is the theoretically grounded and empirically tested taxonomy for helpful patterns in the design of OGD solutions. The taxonomy consists of 24 criteria where 16 were identified as key by classifying 40 OGD solutions for their helpful patterns. The helpful patterns were grouped into 6 clusters following the 16 key criteria, which are (1) simple-personalised help, (2) proactive

multi-visual help, (3) lightly-facilitated exploration help, (4) facilitated data-management help, (5) facilitated information exploration help, and (6) horizon solutions. Another finding is that the importance of key criteria varies between the clusters. We expected helpful patterns to cluster following themes, types, or purposes of solutions, which was not the case, as different solutions provide similar helpful patterns.

References

1. Azkan, C., Iggena, L., Gür, I., Möller, F., Otto, B.: A taxonomy for data-driven services in manufacturing industries. In: Proceedings of the 26th Pacific Asia Conference on Information Systems, p. 184. Association for Information Systems (2020)
2. Bailey, K.D.: *Typologies and Taxonomies: An Introduction to Classification Techniques*. Sage, Thousand Oaks (1994)
3. Bates, M.J.: Information behavior. In: *Encyclopedia of Library and Information Sciences*, vol. 3, pp. 2381–2391 (2010)
4. Creswell, J.W., Creswell, J.D.: *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, Thousand Oaks (2017)
5. Crusoe, J.: How may an OGD solution help you?-an information behaviour perspective. In: Janssen, M., et al. (eds.) EGOV 2022. LNCS, vol. 13391, pp. 181–195. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-15086-9_12
6. Davies, T.: Open data, democracy and public sector reform - a look at open government data use from data.gov.uk (2010)
7. Denscombe, M.: *The Good Research Guide: For Small-Scale Social Research*. McGraw Hill, New York (2010)
8. Erdelez, S.: Information encountering: it's more than just bumping into information. *Bull. Am. Soc. Inf. Sci. Technol.* **25**(3), 26–29 (1999)
9. Foulonneau, M., Martin, S., Turki, S.: How open data are turned into services? In: Snene, M., Leonard, M. (eds.) IESS 2014. LNBIP, vol. 169, pp. 31–39. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04810-9_3
10. Fry, B.J.: *Computational information design*. Ph.D. thesis, Massachusetts Institute of Technology (2004)
11. Gebka, E., Castiaux, A.: A typology of municipalities' roles and expected user's roles in open government data release and reuse. In: Scholl, H.J., Gil-Garcia, J.R., Janssen, M., Kalampokis, E., Lindgren, I., Rodríguez Bolívar, M.P. (eds.) EGOV 2021. LNCS, vol. 12850, pp. 137–152. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-84789-0_10
12. Goldkuhl, G.: The generation of qualitative data in information systems research: the diversity of empirical research methods. *Commun. Assoc. Inf. Syst.* **44**, 572–599 (2019)
13. Gower, J.C., Legendre, P.: Metric and euclidean properties of dissimilarity coefficients. *J. Classif.* **3**, 5–48 (1986)
14. Hossain, M.A., Dwivedi, Y.K., Rana, N.P.: State-of-the-art in open data research: insights from existing literature and a research agenda. *J. Organ. Comput. Electron. Commer.* **26**(1–2), 14–40 (2016)
15. Hunke, F., Engel, C.T., Schüritz, R., Ebel, P.: Understanding the anatomy of analytics-based services-a taxonomy to conceptualize the use of data and analytics in services. In: Proceedings of the 27th European Conference on Information Systems, pp. 1–14. Association for Information Systems (2019)

16. Hunnius, S., Krieger, B.: The social shaping of open data through administrative processes. In: Riehle, D., et al. (eds.) *Proceedings of the International Symposium on Open Collaboration*, pp. 1–5. Association for Computing Machinery (2014)
17. Janssen, M., Zuiderwijk, A.: Infomediary business models for connecting open data providers and users. *Soc. Sci. Comput. Rev.* **32**(5), 694–711 (2014)
18. Koesten, L.M., Kacprzak, E., Tennison, J.F., Simperl, E.: The trials and tribulations of working with structured data: -a study on information seeking behaviour. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1277–1289. Association for Computing Machinery (2017)
19. Lee, D.: Building an open data ecosystem: an irish experience. In: Estevez, E., Janssen, M., Soares Barbosa, L. (eds.) *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pp. 351–360. Association for Computing Machinery (2014)
20. Mainka, A., Hartmann, S., Meschede, C., Stock, W.G.: Mobile application services based upon open urban government data. In: *Proceedings of the 2015 iConference*, pp. 1–15. iSchools (2015)
21. Marradi, A., et al.: Classification, typology, taxonomy. *Qual. Quant.* **24**(2), 129–157 (1990)
22. Mays, N., Pope, C.: Assessing quality in qualitative research. *BMJ* **320**(7226), 50–52 (2000)
23. McKenzie, P.J.: A model of information practices in accounts of everyday-life information seeking. *J. Doc.* **59**(1), 19–40 (2003)
24. Myers, M.D.: *Qualitative Research in Business and Management*. Sage Publications Limited, Thousand Oaks (2013)
25. Nicholas, D., Herman, E.: *Assessing Information Needs in the Age of the Digital Consumer*. Routledge, Milton Park (2010)
26. Paukstadt, U., Strobel, G., Eicker, S.: Understanding services in the era of the internet of things: a smart service taxonomy. In: *Proceedings of the 27th European Conference on Information Systems*, pp. 1–18. Association for Information Systems (2019)
27. Pettigrew, K.E., Fidel, R., Bruce, H.: Conceptual frameworks in information behavior. *Ann. Rev. Inf. Sci. Technol.* **35**(43–78) (2001)
28. Rizk, A., Bergvall-Kåreborn, B., Elragal, A.: Towards a taxonomy for data-driven digital services. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 1076–1085. Association for Information Systems (2018)
29. Safarov, I., Meijer, A., Grimmelikhuijsen, S.: Utilization of open government data: a systematic literature review of types, conditions, effects and users. *Inf. Polity* **22**(1), 1–24 (2017)
30. Savolainen, R.: Everyday life information seeking: approaching information seeking in the context of “way of life”. *Libr. Inf. Sci. Res.* **17**(3), 259–294 (1995)
31. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336–343. Institute of Electrical and Electronics Engineers (1996)
32. Sonnenwald, D.H., Wildemuth, B., Harmon, G.L.: A research method to investigate information seeking using the concept of information horizons: an example from a study of lower socio-economic students’ information seeking behavior. *New Rev. Inf. Behav. Res.* **2**, 65–86 (2001)
33. *The Cambridge Dictionary: Help*. Cambridge University Press (2019). <https://dictionary.cambridge.org/dictionary/english/help>. Accessed 1 Oct 2019

34. Virkar, S., Alexopoulos, C., Stavropoulou, S., Tsekeridou, S., Novak, A.S.: User-centric decision support system design in legal informatics: a typology of users. In: Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance, pp. 711–722. Association for Computing Machinery (2020)
35. Weerakkody, V., Irani, Z., Kapoor, K., Sivaraman, U., Dwivedi, Y.K.: Open data and its usability: an empirical view from the citizen's perspective. *Inf. Syst. Front.* **19**(2), 285–300 (2017)