Open Dataset Identifier for Open Innovation and Knowledge Management

Marta Ortiz-de-Urbina-Criado[a]*

Alberto Abella-García[b]

Diego García-Luna[c]

[a] Universidad Rey Juan Carlos. Paseo de los Artilleros, s/n, 28032, Madrid, Spain. Ph. D. Associate Professor of Management at Universidad Rey Juan Carlos. Email: marta.ortizdeurbina@urjc.es

[b] FIWARE Foundation, Madrid, Spain. Data Modelling Expert and Technical Evangelist at FIWARE and Ph. D. Researcher of Management. Email: alberto.abella@fiware.org

[c] Universidad Politécnica de Madrid, Madrid, Spain. Student of the Bachelor's Degree in Telecommunications Engineering. Email: diegogarcialuna@gmail.com

*Corresponding author

Abstract

Purpose - This paper highlights the importance of open data and the role that knowledge management and open innovation can play in its identification and use. Open data has great potential to create social and economic value, but its main problem is that it is often not easily reusable. The aim of the paper is to propose a unique identifier for open datasets that would facilitate search and access to them and help to reduce heterogeneity in the publication of data in open data portals.

Design/methodology/approach – Considering a model of the impact process of open data reuse and based on the DOI system, the paper develops a proposal of a unique identifier for open datasets called OpenDatId.

Findings – The paper presents some examples of the application and advantages of OpenDatId. For example, users can easily consult the available content catalogues, search the data in an automated way and examine the content for reuse. It is also possible to find out where this data comes from, solving the problems caused by the increasingly frequent federation of data in open data portals and enabling the creation of additional services based on open data.

Originality – From an integrated perspective of knowledge management and open innovation, the paper presents a new unique identifier for open datasets (OpenDatId) and a new concept for datasets, the FAIR Open Data Datasets (FAIRODts).

Keywords Knowledge management; open innovation; open data; unique identifier; OpenDatId; FAIRODts

Article classification Research paper, technical paper

Journal of Knowledge Management. Accepted Version- 05.02.2023. NoneCC BY-NC. Institutional Repository, Non-Commercial Subject Repository, Author's Homepage, No Embargo. Licence CC BY-NC 4.0 (https://v2.sherpa.ac.uk/id/publication/2852).

Citation: Ortiz-de-Urbina-Criado, M., Abella, A. and García-Luna, D. (2023), "Open data-set identifier for open innovation and knowledge management", Journal of Knowledge Management, Vol. 27 No. 10, pp. 2779-2796. https://doi.org/10.1108/JKM-07-2022-0514

Open Dataset Identifier for Open Innovation and Knowledge Management

Introduction

The business world is going through a process of constant change, which has been accelerated in recent years by events such as the pandemic and the Ukrainian war. Such events are causing a great deal of uncertainty and several economic, social, and political problems. In this context, knowledge (Al-Omoush *et al.*, 2020) and innovation are fundamental in any industry (Lu and Tseng, 2010; Okatan, 2012; Witell *et al.*, 2016; Lee and Trimi, 2018). Companies are promoting the development and application of knowledge management models (Klaila, 2000), that act as a value chain that allows assets such as data to create knowledge to develop innovative products and/or services. The application of these knowledge management models leads us to consider the need to be innovative (Jennex and Durcikova, 2014; Gloet and Samson, 2020; López-Cabarcos *et al.*, 2021) and the importance of data in achieving this.

Innovation can yield advantages for the next generation of products and services, or for improving the existing ones (Hilmersson and Hilmersson, 2021; Taques *et al*., 2021). Innovation means novelty, something qualitatively new, created through learning processes and knowledge (Smith, 2004). Innovation today must be continuous and consistent, but there is a high degree of homogeneity between innovation typologies by sub-sectors (Witell *et al.*, 2016; Taques *et a*l., 2021). In the case of the open data sector, the dimension of innovation is underdeveloped, although there are studies that analyse the open innovation that can be achieved in this sector (Corrales-Garay *et al*., 2020; 2022). On the other hand, nowadays, one of the most important resources for companies, public organisations, governments, and society in general is data. Contemporary society is characterised by the massive generation of data due to the use of technology and the development of Big Data and Machine Learning (Del Vecchio *et al.*, 2018; Lystras *et al*.,

2020; Gupta *et al.*, 2021; Arowolo *et al*., 2022). In this context, this interest in data as assets and a necessary element for knowledge management has given rise to a movement advocating for the opening up of datasets in order to innovate and create value for society. Open data serves as a basis for entrepreneurs to generate new business models designed to generate new digital services through the reuse of open data (Lindman, 2014). However, some authors point out the problems of reusing open data published through open data portals (Abella *et al.*, 2022a). One such problem is the publication of datasets in different formats and under different names. If data is not published properly, it loses its reuse value, and does not allow for service innovation and entrepreneurship (Corrales-Garay *et al.*, 2019a; 2022).

Open data has great potential to create social and economic value, but its main problem is that it is often not easily reusable, due among other things, to the lack of standardization and the difficulties in identifying it. This paper will analyse these problems from an integrated perspective of knowledge management and open innovation. The aim of the paper is to propose a unique identifier for open datasets that would facilitate their search and access to them and help to reduce heterogeneity in the publication of data in open data portals.

The main novelties of this work are, on one hand, the creation of a unique identifier, OpenDatId, because there is not such element available right now. And, on the other hand, the explanation of the mechanism on how the identifier would help the ecosystem of open data agents to identify the datasets and simplifies the access to it. This identifier not only helps to unambiguously find a dataset published on an open data portal, but also can be integrated into metrics such as MELODA 5, developed to measure, through various dimensions, the reusability of the dataset. That is, whether new products/services or businesses could be created from them. Then, OpenDatId can be used as a tool that helps

MELODA to first identify and select datasets and then access their metadata to obtain all the information needed to assess each dimension of the metric. The paper explains how this indicator has been integrated into MELODA 5 and presents an example of its application.

After this introduction we will reflect on open data in the context of knowledge management and open innovation and identify some of the problems for publishing open data. The third section will focus on the identification of open datasets and propose a unique identifier called Open Dataset Identifier (OpenDatId). The fourth section will present some examples of the application of OpenDatId and explain its benefits. Finally, in the conclusions section, the main academic and practical implications, limitations, and future lines of research will be discussed.

Knowledge Management and Open Innovation through Open Data

Knowledge management is a multidisciplinary field that draws on disciplines such as Computer Science, Information Science, Organisational Science, and Cognitive Science (Wiig, 2002; Koulouriotis and Emiris, 2004; Dalkir, 2011). As Corrales-Garay *et al.* (2022) note, the more Information Science and Computer Science-centred view is based on the codification of knowledge elements using technologies (Waltz, 2003).

The change and uncertainty of new business environments are promoting the development and application of knowledge management models (Malhotra, 2000; Di Vaio *et al.*, 2021). The process of knowledge management acts as a value chain that requires raw materials such as data to create or develop products, and involves the acquisition, classification, filtering, and indexing of data (Rautenberg *et al.*, 2017). Knowledge management is a good framework for knowledge to create value by learning, sharing, and codifying data and providing a coordination mechanism to transform data and knowledge into products and services (Bashir and Farooq, 2019).

To develop an open data business model, knowledge management capabilities can be utilised by considering knowledge acquisition, conversion, dissemination, application, and reuse (Corrales-Garay *et al.*, 2022). In this context, knowledge management practices can support innovation in general (Okatan, 2012; Wu and Hu, 2018), open innovation (Ferreira and Teixeira, 2019; Öberg and Alexandery, 2019), and collaborative innovation (An *et al.*, 2014), in particular. The use of open data in services, especially if they are digital, requires the development of innovation practices, service development models, and a collaborative environment (Immonen *et al.*, 2018).

"Open innovation is defined as the systematic realization of knowledge exploration, retention and exploitation within and outside the boundaries of an organization throughout the innovation process" (Lichtenthaler, 2011, 77). Open innovation has changed the innovation paradigm (Chesbrough *et al.*, 2008) by offering more possibilities for commercialising internal and external ideas: inbound - opening up to external ideas and technologies to improve the value of products; outbound - externalising internal resources to refine, exploit and bring them to market; and coupled - a combination of inbound and outbound processes (Gassmann and Enkel, 2004; Enkel *et al.*, 2009).

An interesting aspect is to analyse the flow of knowledge that can be achieved with open innovation. To innovate successfully, it is necessary to internally develop the capabilities to explore knowledge and to reap the benefits of external knowledge acquisition (Vilas Boas Viveiros Lopes and Monteiro de Carvalho, 2018).

Considering a knowledge management framework, Lichtenthaler and Lichtenthaler (2009) propose a theoretical model for open innovation based on three processes: exploration, retention, and exploitation, which require different capacities. Inventive capacity refers to the capacity to generate and exploit knowledge internally; absorptive capacity is the capacity to exploit external knowledge and use it in the best way internally;

transformative capacity is the firm's capacity to maintain acquired knowledge over time; connective capacity refers to the capacity to maintain knowledge in inter-firm relationships; innovative capacity is the firm's capacity to generate innovations from new knowledge; and desorptive capacity refers to the firm's capacity to transmit knowledge to the market (Vilas Boas Viveiros Lopes and Monteiro de Carvalho, 2018).

According to Naqshbandi (2016), the success of the open innovation process requires that firms can explore, transform, and commercialise externally acquired knowledge. In special, Jiménez-Barrinuevo *et al*. (2011) tested a tool to measure absorptive capacity that has four phases: acquisition, assimilation, transformation, and exploitation. Some studies have shown that absorptive capacity helps to have a positive relationship between access to external resources and competitive advantage (Zobel, 2017) or acts as an intermediary between open innovation and performance (Ahn *et al*., 2016; Kokshagina *et al*., 2017). Absorptive capacity can be considered as a driver or a constraint for open innovation (Vilas Boas Viveiros Lopes and Monteiro de Carvalho, 2018).

Under the umbrella of the theoretical foundations discussed above - knowledge management and open innovation - the current environment demands new ways of adapting to the environment based on openness and transparency. The importance of data as raw material and a necessary element for knowledge management and value creation has led to the development of studies that include the open data movement to innovate and create value in society.

The Open Knowledge Foundation (2008) defines open data in the following way: "Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness" (https://opendefinition.org/od/2.1/en/). Open data provides access to internal and external data, mainly from public organisations. Governments and public bodies are releasing their

data and want open data to be used to solve problems and to create and improve products and services.

Open data can be a suitable tool to improve the open innovation capacities of companies by giving access to internal (inventive capacity) and external information that if well codified can be easily assimilated and applicable (absorptive and connective capacities) and to create and commercialise products and services adapted to the current market (transformative, innovative and desorptive capacities). The model in Table 1 proposes a process for analysing innovation and value creation through open data that focuses on the capacities discussed above.

Insert Table 1 here

Data can be published and managed by different public and private organisations (Kampars *et al.*, 2020). Open data has the quality of increasing transparency, accountability, participation, and empowerment of citizens and has economic and social impact by stimulating business innovation (Corrales-Garay *et al.*, 2022). But access to open data does not in itself lead to innovation because data needs to be made available in an appropriate form. For value creation, open data must be accessible and free (Kitsios *et al.*, 2017), so it has to be available in a machine-readable format and without restrictions on permissions to use or distribute it (Sadiq and Indulska, 2017).

In that line, Janssen and Zuiderwijk (2014, p. 707) recommended that "an ecosystem view should be adopted to understand how added value can be created for users by taking advantage of already existing models and social media". But Kitsios *et al.* (2017) found that there is a limited knowledge about the open data ecosystem from the business perspective. Zuiderwijk *et al.* (2014) explain that open data ecosystem could be seen as the combination of different types of ecosystems: 1) government ecosystems -open data is mainly published by the government-; 2) business ecosystems -open data can also be

provided by the private sector-; 3) innovation ecosystems -for the collaborative arrangements between agents-; 4) information ecosystem -interconnection of people, work, value supported by technology-; 5) software ecosystems -a networked community of organisations based in software technology-; and 6) digital ecosystem -interconnected, interrelated and interdependent digital species enabling the service co-innovation and co-creation among members utilising and sharing common assets and knowledge-. In this context, open innovation processes can be applied to foster collaboration between different actors, and for service development and improvement (Bican *et al*., 2017).

Kampars *et al.* (2020) consider the capability- driven development approach, which allows modelling of open data processing ecosystems This approach facilitates knowledge exchange about open data usage among members of the ecosystem and supports configuring information systems for open data processing. In that context, actors in the open data ecosystem -data providers, service providers, application developers, application users, and infrastructure and tool providers- and business models are considered in previous literature (e.g., Immonen *et al*., 2014; Zimmermann and Pucihar, 2015; Kitsios *et al*., 2017). For example, Janssen and Zuiderwijk (2014) identifies six types of open data business models: single-purpose apps; interactive apps; information aggregators; comparison models; open data repositories; and service platforms.

In the open data ecosystem, both the usability of open data and the value that can be derived from its reuse are important considerations. It is therefore important to distinguish between the use and reuse of data. Pasquetto *et al.* (2017) provide a clear explanation of this difference in their work on the "Reuse of Scientific Data". For example, data can be collected for a specific project, and the initial "use" is made by the collector. If the dataset is then used again, either for the same project or for a subsequent one, it would be considered a "use." However, if the dataset is placed in a repository and retrieved by

someone else for use in a different project, it is typically considered a "reuse." In other words, reuse typically involves the use of a dataset by someone other than its creator (Pasquetto *et al*., 2017; 2019). For open data in general, open data reuse refers to the use of data for a purpose other than what was intended by the original producer (Abella *et al*., 2014; 2019b). According to Abella *et al.* (2014), for open data to be reused, four minimum conditions must be met: absence of technical barriers to reuse; possibility of automated access to the information; existence of a legal framework that allows its use; and access to knowledge of the structure of the published information.

Some examples about open data reuse are presented. Services that provide personalized recommendations for sustainable transportation, such as avoiding areas with high pollution levels or alerting users about pollen levels are based on open data on the environment. Most of the mobility applications are based on open data released by the cities regarding traffic and public transportation services. Another group of services that reuse data provide support for healthy lifestyle habits, such as providing information on nutrition and sports facilities. Connemara Programme is an example of an Irish organisation that sells products and services based on open data. It has an open data repository whose data and photographs can be used for business web sites, apps, online brochures, e-books, in research, in business applications, reports presentations… (http://www.connemaraprogramme.com/opendata/). Authors such as Berends *et al.* (2020) provide several more examples of the reuse of open data by organizations, mainly in the private sector, and demonstrate some of the business models that have been developed around their reuse.

Given the great utility that the reuse of open data can have for open innovation, some authors have already proposed theoretical models for to better understand the process of open innovation through open data. Corrales-Garay *et al.* (2019b; 2020) develop a model

that considers the four phases of the impact process of open data reuse proposed by Abella *et al.* (2019a): 1) candidate data; 2) published data; 3) reused data, and 4) impact, and combine it with the types of open innovation -inbound, outbound and coupled (Table 2). Following that model, it is necessary to have reusable data to create products and services with added value for open innovation with open data to have an impact. But if data is not published properly, it loses its reuse value, and does not allow for product or service innovation (Corrales-Garay *et al*., 2019a; 2022).

Insert Table 2 here

Considering the objective of this paper, we will focus on the second phase of the model (Table 2), where published data is analysed. Corrales-Garay *et al.* (2020) explain that previous literature has analysed where and how data can be published, paying special attention to web platforms and open data portals, but that there are several topics of interest that could be new areas of research. Two of those topics are 1) the analysis of the characteristics that an open data portal should have to publish reusable data for innovation, and 2) publishing data in a standardised format that allows for comparison between portals.

The first topic has been developed by Abella *et al.* (2022a), who analyse open data portals that are inefficient, which are called "pretender open data portals" - PODP. They are portals that are not suitable for the professional reuse of their data. These authors propose three minimum criteria that an open data portal must have to publish data that can be reused for professional purposes: 1) have a timely announcement mechanism on the updates; 2) have DMS to allow the automatic publication and update of the data, and 3) have APIs to be access the data in a reusable format.

However, the second topic is unexplored. It is necessary to consider how to publish open data in a standardised format that allows comparison between portals and facilitates direct

and automated access to them. This is the aspect that we are going to develop in the next sections, building on the insight of Steinberg and Brehm (2009, p. 63) who state that "the more we know about a source, in this case, about open data, the better we can reuse it". In that line, there are some initiatives such as the MELODA metric, which was established in 2011, as a reaction to the lack of consistency in the datasets published in open data portals, to analyse the degree of open data reuse (Abella *et al.*, 2014; 2019b). At present, there is still no system that enables the unique identification of open datasets and allows direct and automated access to them.

Identifying open data

Background

Open data portals publish information in a grouped form through so-called datasets. However, some authors state that it is necessary to provide access to them and give them a structure that allows them to be reused (Garriga-Portolà, 2011).

In the context of the scientific community, much progress has been made with proposals for sharing data that come from academic research. So-called Open Science promotes the public accessibility of research data, especially for those datasets that have been funded with public funds. It is in this context that FAIR Data has been defined, i.e., data that is discoverable, accessible, interoperable, and reusable but which, unlike open data, is not always available to everyone (Dunning *et al.*, 2017; Gvishiani *et al.*, 2021). FAIR data principles can be applied for achieving reusability (Hasnain and Rebholz-Schuhmann, 2018; Groth *et al.*, 2020). However, although they are different concepts, open data can be improved by making it discoverable, since the first step towards its reuse is to be able to find it. Metadata is necessary so that data can be read by machines and found by humans (Groth *et al.*, 2020) and for this, a unique and persistent identifier with a correct description is required.

From Open Science, we have a useful example of a unique identifier, the DOI (Digital Object Identifier). This is a permanent link in the form of an alphanumeric code that uniquely identifies electronic content such as a scientific article, a book, or even an image or song. A DOI name can be assigned to any entity - physical, digital, or abstract - and is designed for interoperability, i.e., for use with existing identifier and metadata schemes. In addition, DOI names can also be expressed as URLs - URIs (International DOI Foundation, 2016). The advantage of the DOI over the URL system used in web pages is that the DOI does not change over time and with the information available in its metadata it is easily accessible even if it has been relocated to a different address or web page. The DOI name has two components, the prefix, and the suffix, which together form the DOI name, separated by the "/" character. There is no limitation on the length of a DOI name. The prefix precedes the "/" character and denotes a unique naming authority. The suffix can be an existing identifier, or any unique string chosen by the registrant. A DOI name can be assigned to any entity that can have multiple prefixes, defined by structured metadata, and remains persistent across ownership changes, and unchanged once assigned (International DOI Foundation, 2016).

Proposal for a unique identifier for open data: OpenDatId

Open data can follow this unique identifier system for being published in a way that it can easily be found and read by machines. Therefore, in this section, we propose a unique identifier for open data, called the Open Dataset Identifier (OpenDatId). Based on the DOI system, the Open Dataset Identifier will contain the information of the entity that publishes it, the subject that is published, the date of publication and the version of the data.

To identify the entity (XXXXXXXX), a variable number of characters is allocated that cannot contain any full stops, namely eight digits of a number that is assigned by the

central registry. The code XXXXXXXX identifies the original source (publisher), not the creator. For the topic (TTTT), it has been chosen to use the URI of the API access for the metadata of that dataset. The date is in the format YYMMDD, i.e., year, month, and day, and separated by a T, the time of publication –HHMMSS- (hour, minutes, and seconds) is added (compliant with part of the ISO 8601-1:2019 standard). Finally, the version of the dataset is included after the date, separated by a slash with any group of characters.

For datasets that have already been published where all the information is available, the following adjustments have been made: 1) the entity has to guarantee the uniqueness of its TTTT denomination; 2) for the date, take the releaseDate field and if there is no date metadata, put 8 zeros (00000000); 3) for the version (VersionText), take the version field and in case of an empty value, put a 1. This text can be of any length and value can include any character but '#'.

Thus, we define the Open Dataset Identifier (OpenDatId) as a unique identifier for open datasets containing four fields - entity, text, date, and version - which are indicated with the following structure: XXXXXXXX#TTTT#YYYYMMDDDTHHMMSS.VersionText

OpenDatId utilities and examples of use

OpenDatId utilities

The proposed open dataset identifier shares the functionalities of the DOI system (International DOI Foundation, 2016). It is persistent even if the dataset is moved or reordered; it allows interoperability with other data from other sources because it supports DCAT-AP 2.1.0 metadata, and a small extension makes it possible to share the principal metadata. It enables single data management for multiple output formats regardless of platform. It allows the management of classes of applications and services as well as a dynamic updating of metadata, applications, and services. Furthermore, by integrating these functionalities it endows open data with the quality of being discoverable,

accessible, interoperable, and reusable by providing metadata that can be read by machines and found by humans thanks to the availability of a unique and persistent identifier.

Open datasets with OpenDatId could be considered not only open, but also FAIR data. Therefore, we propose a new concept, "FAIR Open Data Datasets" (FAIRODts), which are defined as open datasets, usually published on open data portals, that, by having a unique identifier, OpenDatId, have the qualities of being discoverable, accessible, interoperable, and reusable.

The implementation of an OpenDatId system could have similar benefits to the DOI system, facilitating the internal management of content and the possibility of the faster and more scalable development of products and services (International DOI Foundation, 2016). Following the advantages proposed in the manual developed by the International DOI Foundation (2016), the first advantage of OpenDatId is that users can see who the publisher is, and they can consult the available catalogue OpenDatId. The second advantage is that you can find what the users are looking for, and directly access the data to be used or reused. The third advantage is that you can easily reach the source by exploring the attached metadata. The fourth is that you can identify the original source in those open data portals federating [1] data from other sources. And the last advantage is that data can be retrieved both manually and automatically through software or applications.

In addition to its functionalities and advantages, OpenDatId makes easier the identification of datasets that are published in pretender open data portals where no API is available. Abella *et al.* (2022a) explain that these portals, PODP, are inefficient and waste resources, so incentives should be sought so that new portals of this type are not developed and/or those that exist are improved. If a unique identification system for open

datasets is established, the use of APIs will be more likely to become widespread, as they would be necessary in order to have an OpenDatId.

Another benefit of OpenDatId is that it helps to automate searching for and downloading open datasets. For example, with the software IDRA (https://idra.readthedocs.io/en/latest/), an open-source package for the federation of open data sources, the use of the identifier would make it possible to filter unique datasets. Trombino *et al*. (2017) developed a study about data and metadata quality for smart cities implementing MELODA 4 (Abella *et al.*, 2018) as a module for IDRA.

Examples of identification and use of OpenDatId

Open Innovation is a multidimensional phenomenon that has been analysed in the literature with multiple examples and case studies (Di Minin *et al*., 2016). Some companies have adopted open innovation for innovative software development to help solve problems. For example, NASA created in collaboration with TopCoder, Harvard Business School and London Business School, innovative software with a mathematical algorithm to determine the optimal content of medical kits for future NASA manned missions. Samsung embraced open innovation through the Samsung Accelerator programme which provides office space, static capital, and product support to entrepreneurs to help them create software and services. In line with these examples, we present an open-source software development to identify the degree of reuse of open data using OpenDatId.

To present some examples of the OpenDatId identifier, the *Reusing Open Data in Spain III* report (Abella *et al.*, 2022b) has been used as a reference. This report identifies 289 valid portals, lists 58.318 available datasets, and samples 300 datasets, of which 280 (93.3%) are valid one. To analyse their degree of reuse, the authors applied MELODA (MEtric for the evaLuation of Open DAta) version 5, identifying the following

dimensions for each dataset (Abella *et al.*, 2019b): legal licence for reuse, technical standard in which the information is presented, mechanisms of access to the information, the data model used, geographic content of the information, update frequency, dissemination, and reputation. Taking this information as a starting point, 20 datasets from the most popular Data Management Systems -CKAN, ODS, Socrata, Arcgis Open data and DKAN- have been selected from different portals that were federated in the datos.gob portal - https://datos.gob.es/es- (Table 3).

Based on datasets in Table 3, an example is presented to demonstrate the actual implementation of OpenDatId (Table 4). To understand the usefulness of open data reuse and to facilitate automation in accessing open datasets, we have developed a proposal for the automation of the information search process to update and improve the IDRA module to adapt it to MELODA 5. An open-source software for implementing this unique identifier has been developed [2].

Insert Tables 3 and 4 here

MELODA (MEtric for the evaLuation of Open DAta) is a metric to measure and score the reusability of open data on eight dimensions with various levels. Once each level has been identified, the sum of the scores obtained in each dimension is made, with a maximum value of 61 points (Abella *et al.*, 2019b). According to these scores, the degree of reusability of open data is interpreted as follows: if the total score is between 8 (no less can be obtained) and 23 points, it can be said to be inadequate; if a score is obtained between 24 and 47 points, it is basic; and if it is between 48 and 62, it is considered advanced. With this information, a ranking of datasets according to their degree of reuse can be obtained (Abella *et al*., 2022b).

But to use MELODA, data has to be collected one by one and analysed. This way of analysing open data is too time-consuming, so it is necessary to streamline this process.

IDRA is a component of the FIWARE platform. IDRA is an open-source software developed by Engineering Ingenería Informática, SpA, within the EU funded project FESTIVAL. It is a web application that allows to federate existing open data management systems (ODMS) from different technologies providing a single access point to search and discover open datasets coming from heterogeneous sources. Trombino *et al*. (2017) explain the creation of an external module with IDRA API, capable of evaluating datasets with the MELODA metric. This module is not present in the official IDRA repository but was used in specific instances of IDRA integrated in other engineering products. However, the version of MELODA used by IDRA is already outdated and does not apply the new quality criteria, which makes it necessary to implement the new metric (MELODA 5), so that it is possible to make a correct assessment in the most up-to-date way possible. Already commented before, a new open-source software has been designed using the JAVA object-oriented programming language to measure the reusability of open data by means of the MELODA 5 metric [2].

To establish a complete automation of the different datasets and to check the reusability of the data using the MELODA 5 metric, the first problem encountered is that the portals encode the datasets differently. This means that they do not use the same tags to identify the same fields, which is a problem for an automation process. It is also impossible to know if the dataset has originally published by this portal or if it is federated from others. These facts mean that it is impossible to know how many datasets publish a portal or to a larger scale how many are published in a country. To solve this problem, OpenDatId has been used, which contains the API where each dataset comes from.

Once the open-source software has been developed, a real application case has been carried out to check that the expected results are indeed achieved. OpenDatID for each dataset has been stored and all identifiers have been entered into the program. The

software has been run and the result obtained is a list of all the datasets analysed with the MELODA 5 value of each one and indicating whether they have an inadequate, basic or advanced degree of reuse (Table 4).

The OpenDatId identifier helps to locate the datasets and thus to access their metadata. These metadata can store the actual values of the MELODA metric dimensions, so that the necessary information can be obtained to assess the degree of reuse of the data, in an automatic way. MELODA 5 analyses, for each dataset, several dimensions and provides a score that allows datasets to be classified into three categories: inadequate, basic or advanced. OpenDatId can be used as a tool that helps MELODA to first identify and select datasets and then access their metadata to obtain all the information needed to measure each dimension of the metric. The identifier does not assign the associated score that measures the degree of reuse, but by identifying the metadata location, rather facilitates access to the information needed to score each dimension of the metric in order to obtain the final score.

Following Table 1, this example shows the usefulness of the unique indicator developed to facilitate open innovation by helping to improve capacities -specially, the inventive, absorptive, and transformative capacities. and by allowing easier and more unambiguous access to open data. It also helps to improve the application of metrics such as MELODA facilitating the use of this data to create or develop apps or business models with open data (innovative and desorptive capacity). Being able to uniquely identify datasets makes the process more transparent, replicable and allows the information to be kept up to date (connective capacity). It also makes it possible to better assess the quality of the open data and to perform quality controls on the products / services developed.

Implementing the identifier

To extend the implementation of OpenDatId, it would be necessary that an entity or a federation of entities to implement, monitor and improve the OpenDatId and their governance regulations for mutual recognition. It is also necessary to have a catalogue with the encoding of the names of the dataset publishers, and their corresponding unique codes. Additionally, this catalogue has to maintain the metadata of the dataset at the time of publication. It also needs to check if the publishing licence is an open licence according to the definition of open data (there are more than 400 licences for this: https://spdx.org/licenses/). This catalogue should be available for consultation by the public. In addition to this, an ordered metadata structure should be in place. This structure should respect the DCAT-AP 2.1.0 terminology and extend it for those attributes not available in this version of the specification (attributes not available at the 2.1.0 version include the data model reference allowed by the dataset, whether the dataset contains geographic coordinates, the list of fields of the datasets, etc.).

Conclusions

This paper highlights the importance of data and the role that perspectives such as knowledge management and open innovation play in its discovery, identification, and use. What would today's world be without data?

OpenDatId is a new identifier created expressly for open datasets to have a unique identifier following the same idea as doi. For its creation, the API has been used because it is the most standardised way to access metadata from data portals, but it is not specifically created for open datasets, so this identifier has been created from it.

The creation of a unique identifier for open datasets is an important step towards making efficient and ethical use of these data, because if we cannot find the data in the first place or be certain as to its provenance, how can we access it efficiently, i.e., in an automated

way? How can we keep track of any updates to? Ultimately, how can we efficiently use and/or reuse it?

In this paper, a unique identifier for open datasets called the Open Dataset Identifier (OpenDatId) has been developed. DCAT-AP 2.1.0 optionally defines the attribute *identifier* as (European Commision, 2021) "The property that contains the main identifier for the dataset, e.g., the URI or other unique identifier in the context of the catalogue". The proposed OpenDatId complies with the requirements of the attribute and would standardise its information.

Academic implications

This paper addresses one of the many challenges related to the increasing availability of data and how to facilitate its access and reuse. Firstly, this paper lies between the research and technical domains and identifies foundations based on knowledge management and open innovation. In that sense, a theoretical model based on the development of capacities for open innovation and value creation is presented (Table 1). Secondly, a framework that integrates knowledge management, open innovation and open data has been developed and the role of OpenDatId for the development of capabilities has been highlighted.

Thirdly, a new concept, FAIR Open Data Datasets (FAIRODts), has been proposed. It reflects the importance of open data, in addition to being freely available, having the qualities of being discoverable, accessible, interoperable, and reusable. This can be achieved if the use of a unique dataset indicator such as OpenDatId is widely implemented according to the defined schema.

Fourthly, a discussion on the requirements and uses of OpenDatId has been carried out. Some examples of the application of the OpenDatId in real datasets published on open data portals have been presented. An a new open-source software for implementing this unique identifier is explaining as example. Finally, the interest of creating an entity for

standardisation of open data has been claimed. Besides this, it would be very useful to create an open data observatory that, among other tasks, would be dedicated to implementing, developing, and monitoring the OpenDatId system.

Practical implications

OpenDatId would enable the creation of a common system for naming open datasets and facilitating their publication, discovery, and reuse. This system would be of interest to all actors in the open data ecosystem (Table 2). It would facilitate the task of publishing and updating open datasets for primary open data sources. For reusers, it would help to automatically search and download data. For end-users, it presents a more transparent data management model and benefits from the creation and development of products and services.

For all actors in the ecosystem, OpenDatId allows for immediate identification of the publisher of the data, access to the API and finding out when it has been updated and which version is being used. Also, these open data could be used by professional reusers to develop services for citizens, or even citizens themselves are implementing services of their own (they account for 20% of open data-based services listed in open data portals) according to Abella et al. (2022b).

Open data have three scopes of impact: 1) government policies and practices, 2) innovation by the combination of technology, business, and government, and 3) businesses, users, and civil society (Zuiderwijk *et al.*, 2014). For those scopes, benefits of OpenDatId for open Innovation are summarised.

The first benefit is creating new or innovating old products and services. OpendatId will allow a unique identification of datasets, including the access to its metadata and therefore enables the creation of new products and services or their update by allowing the finding and reuse of the published data. The second benefit is building a strong community. The

OpendatId creation service described along the article defines, implicitly, a community of practice and allows single points of contact of all the publishers. It enables the creation of a community where other issues regarding open data publication and reuse can be analysed and debated. The third benefit is staying ahead of the competition. OpendatId has a deep network effect. The more portals using it, the larger usefulness for the reusers. And whatever country starts with the system it will allow their datasets to be more used and analysed and therefore it will stay ahead of other countries in terms of open data reuse. The fourth benefit is the costs reduction. OpendatId reduces dramatically the cost of finding and selecting information for open data reusers because it allows to know the unique source of a dataset and to quickly purge redundant data sources. It also reduces the cost for global analysis because it filters unique datasets and therefore federated datasets can be easily purged from collected data sources. Currently, this task must be performed manually on the hundreds or thousands of datasets that open data portals can hold. And the last benefit is the time-to-market acceleration. OpendatId speeds up the discovering of new datasets by identifying when a dataset is new or a republication of another.

Limitations and future research lines

The consideration of open data from knowledge management and open innovation opens very interesting lines of research, but it is still under development. Future studies can analyse issues such as the social and economic impact of reusing open data for open innovation and for creating new knowledge to promote business creation.

Having a unique identifier for datasets is important for facilitating identification and access to them, but there are other aspects related to publishers and their visibility that can help to make open data more reusable and have a greater economic and social impact. Future research can analyse topics such as the dissemination of open data or the reputation

of the publishers and creators of open data portals. On the other hand, it would be interesting to further explore the development and implementation of OpenDatId through other qualitative methodologies such as the Delphi method. There are two interesting research questions that complement the contributions of the present study: What are the social and economic effects of having a unique identifier for open datasets? And how can these effects be measured?

Finally, another challenge is to emphasise the value of their use, following on from the idea that they are the raw material for innovation. The theories and indicators of service innovation (Taques *et al*., 2021) can be a useful complement to the open innovation framework for looking into the third and fourth stages - reuse and impact - (Rajapathirana and Hui, 2018; Rauter *et al.*, 2019) of the model (Table 2) proposed by Corrales-Garay *et al.* (2020). Future work can analyse the opportunities offered by data and the knowledge generated from them to carry out not only open innovations, but also other types of innovation such as innovation in services, processes and organisational and/or commercial aspects (Witell *et al*., 2016; Abdi *et al.*, 2018; Huarng *et al.*, 2021; Taques *et al.*, 2021).

Notes

[1] Data federation occurs when datasets belonging to other portals are published in the catalogue of an open data portal. This publication can be complete with the duplicated dataset or only linking to the original portal (Abella *et al.*, 2022b).

[2] The design and development of the open-source software is explained in a Final Project that obtained the highest grade (10 out of 10). This academic work is not cited to maintain anonymity. The reference will be included in the final version of the manuscript.

Acknowledgements

To maintain the anonymity of the authors, financing sources have been cited on the title page.

References

Abdi, K., Mardani, A., Senin, A.A., Tupenaite, L., Naimaviciene, J., Kanapeckiene, L. and Kutut, V. (2018), "The effect of knowledge management, organizational culture and organizational learning on innovation in automotive industry", *Journal of Business Economics and Management*, Vol. 19 No. 1, pp. 1-19. https://doi.org/10.3846/jbem.2018.1477

Abella, A., Ortiz-de-Urbina-Criado, M. and De-Pablos-Heredero, C. (2014), "Meloda, a metric to assess open data reuse", *Profesional de la información*, Vol. 23 No. 6, pp. 582-588. http://recyt.fecyt.es/index.php/EPI/article/view/34878

Abella, A., Ortiz-de-Urbina-Criado, M. and De-Pablos-Heredero, C. (2018), "Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona", *Profesional de la Información*, Vol. 27 No. 2, pp. 375–382. https://doi.org/10.3145/epi.2018.mar.16

Abella, A., Ortiz-de-Urbina-Criado, M. and De-Pablos-Heredero, C. (2019a), "The process of open data publication and reuse", *Journal of the Association for Information Science and Technology*, Vol. 70 No. 3, pp. 296-300. https://doi.org/10.1002/asi.24116

Abella, A., Ortiz-de-Urbina-Criado, M. and De-Pablos-Heredero, C. (2019b), "Meloda 5: A metric to assess open data reusability", *Profesional de la información*, Vol. 28 No. 6, e280620. https://doi.org/10.3145/epi.2019.nov.20

Abella, A., Ortiz-de-Urbina-Criado, M. and De-Pablos-Heredero, C. (2022a), "Criteria for the identification of ineffective open data portals: Pretender open data portals",

Profesional de la información, Vol. 31 No. 1, e310111. https://doi.org/10.3145/epi.2022.ene.11

Abella, A., Ortiz-de-Urbina-Criado, M., De-Pablos-Heredero, C. and García-Luna, D. (2022b), *Reusing Open Data in Spain III. October 2021*, ESIC Editorial, Madrid. Retrieved 27/03/2022 from https://www.meloda.org/wp-content/uploads/2022/03/open_data_in_Spain_III_2021.pdf

Al-Omoush, K.S., Simón-Moya, V. and Sendra-García, J. (2020), "The impact of social capital and collaborative knowledge creation on e-business proactiveness and organizational agility in responding to the COVID-19 crisis", *Journal of Innovation & Knowledge*, Vol. 5 No. 4, pp. 279-288. https://doi.org/10.1016/j.jik.2020.10.002

An, X., Deng, H., Chao, L. and Bai, W. (2014), "Knowledge management in supporting collaborative innovation community capacity building", *Journal of Knowledge Management*, Vol. 18 No. 3, pp. 574–590. https://doi.org/10.1108/JKM-10-2013-0413

Ahn, J.M., Ju, Y., Moon, T.H., Minshall, T., Probert, D., Sohn, S.Y. and Mortara, L. (2016), "Beyond absorptive capacity in open innovation process: the relationship between openness, capacities and firm performance", Technology Analysis & Strategic Management, Vol. 28. pp. 1009-1028. https://doi.org/10.1080/09537325.2016.1181737

Arowolo, M.O., Ogundokun, R.O., Misra, S., Agboola B.D. and Gupta, B. (2022), "Machine learning-based IoT system for COVID-19 epidemics", Computing, https://doi.org/10.1007/s00607-022-01057-6

Bashir, M. and Farooq, R. (2019), "The synergetic effect of knowledge management and business model innovation on firm competence: a systematic review", International

Journal of Innovation Science, Vol. 11 No. 3, pp. 362-387. https://doi.org/10.1108/IJIS-10-2018-0103

Berends, J., Carrara, W., Engbers, W. and Vollers, H. (2020), "Reusing open data: A study on companies transforming open data into economic and societal value", *Publications Office of the European Union*, Luxembourg: European Union. https://data.europa.eu/doi/10.2830/876679

Bican, P.M., Guderian, C.C. and Ringbeck, A. (2017), Managing knowledge in open innovation processes: An intellectual property perspective, *Journal of Knowledge Management*, Vol. 21 No. 6, pp. 1384-1405. https://doi.org/10.1108/JKM-11-2016-0509

Chesbrough, H., Vanhaverbeke, W. and West, J. (2008), "Open Innovation: Researching a New Paradigm", Oxford University Press, https://EconPapers.repec.org/RePEc:oxp:obooks:9780199226467

Corrales-Garay, D., Mora-Valentín, E.-M. and Ortiz-de-Urbina-Criado, M. (2019a), Open data for open innovation: An analysis of literature characteristics", *Future Internet*, Vol. 11 No. 3, pp. 77. https://doi.org/10.3390/fi11030077

Corrales-Garay, D., Ortiz-de-Urbina-Criado, M. and Mora-Valentín, E.-M. (2019b), "Knowledge areas, themes and future research on open data: A co-word analysis", *Government Information Quarterly*, Vol. 36 No. 1, pp. 77-87. https://doi.org/10.1016/j.giq.2018.10.008

Corrales-Garay, D., Ortiz-de-Urbina-Criado, M. and Mora-Valentín, E.-M. (2020), "A Research Agenda on Open Data Impact Process for Open Innovation", *IEEE Access*, Vol. 8 No. 1, pp. 34696-34705. https://doi.org/10.1109/ACCESS.2020.2974378

Corrales-Garay, D., Ortiz-de-Urbina-Criado, M. and Mora-Valentín, E.-M. (2022), "Understanding open data business models from innovation and knowledge

management perspectives", *Business Process Management Journal,* Vol. 28 No. 2, pp. 532-554. https://doi.org/10.1108/BPMJ-06-2021-0373

Dalkir, K. (2011), *Knowledge Management in Theory and Practice*, The MIT Press, Cambridge. https://doi.org/10.4324/9780080547367

Del Vecchio, P., Di Minin, A., Messeni Petruzzelli, A., Panniello, U. and Pirri, S. (2018), "Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges", *Creativity and Innovation Management,* Vol. 27 No. 1, pp. 6-22. https://doi.org/10.1111/caim.12224

Di Minin, A., De Marco, C.E., Marullo, C., Piccaluga, A., Casprini, E., Mahdad, M. and Paraboschi, A. (2016), "Case Studies on Open Innovation in ICT", Institute for Prospective Technological Studies, *Joint Research Centre Working Papers, JRC100823.* https://www.doi.org/10.2791/433370

Di Vaio, A., Palladino, R., Pezzi, A. and Kalisz, D.E. (2021), "The role of digital innovation in knowledge management systems: A systematic literature review", *Journal of Business Research,* Vol. 123, pp. 220-231. https://doi.org/10.1016/j.jbusres.2020.09.042

Dunning, A., de Smaele, M. and Böhmer, J. (2017), "Are the FAIR Data Principles fair?", *International Journal of Digital Curation,* Vol. 12 No. 2, pp. 177-195. https://doi.org/10.2218/ijdc.v12i2.567

Enkel, E., Gassmann, O. and Chesbrough, H. (2009), "Open R&D and open innovation: Exploring the phenomenon", *R & D Management,* Vol. 39 No. 4, pp. 311-316. https://doi.org/10.1111/j.1467-9310.2009.00570.x

European Commision (2021). *DCAT Application Profile for data portals in Europe Version 2.1.0*. Available at:

https://joinup.ec.europa.eu/sites/default/files/distribution/access_url/2021-12/5bf41792-1a2f-4851-aee2-6ecf43815bc1/dcat-ap_2.1.0.pdf

Ferreira, J.J. and Teixeira, A.A.C. (2019), "Open innovation and knowledge for fostering business ecosystems", *Journal of Innovation & Knowledge,* Vol. 4 No. 4, pp. 253-255. https://doi.org/10.1016/j.jik.2018.10.002

Garriga-Portolà, M. (2011), "¿Datos abiertos? Sí, pero de forma sostenible [Open data? Yes, but in a sustainable way]", *Profesional de la información*, Vol. 20 No. 3, pp. 298-303. https://doi.org/10.3145/epi.2011.may.08

Gassmann, O. and Enkel, E. (2004), "Towards a theory of open innovation: Three core process archetypes", in *R&D Management Conference (RADMA) 2004*. – Lissabon, Portugal. Retrieved from https://www.alexandria.unisg.ch/274/1/Gassmann_Enkel.pdf

Gloet, M. and Samson, D. (2020), Knowledge management and systematic innovation capability", In *Disruptive Technology: Concepts, Methodologies, Tools, and Applications* (pp. 1198-1218). IGI Global, USA.

Groth, P., Cousijn, H., Clark, T. and Goble, C. (2020), "FAIR Data Reuse – the Path through Data Citation", *Data Intelligence*, Vol. 2 No. 1-2, pp. 78–86. https://doi.org/10.1162/dint_a_00030

Gupta, B.B., Gaurav, A. and Peraković, D. (2021), "A Big data and deep learning based approach for DDOS detection in cloud computing environment", in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)* (pp. 287-290), https://doi.org/10.1109/10.1109/GCCE53005.2021.9622091

Gvishiani, A., Dobrovolsky, M. and Rybkina, A. (2021), "Chapter 6 Big Data and FAIR Data for Data Science", Roberts, F.S., & Sheremet, I.A. (eds). *Resilience in the*

Digital Age. Lecture Notes in Computer Science, 12660. Springer, Cham. https://doi.org/10.1007/978-3-030-70370-7_6

Hasnain, A. and Rebholz-Schuhmann, D. (2018), "Assessing FAIR data principles against the 5-Star open data principles", in: Gangemi, A. et al. *The Semantic Web: ESWC 2018 Satellite Events. ESWC 2018. Lecture Notes in Computer Science, 11155*, pp. 469-477. Springer, Cham. https://doi.org/10.1007/978-3-319-98192-5_60

Hilmersson, F.P. and Hilmersson, M. (2021), "Networking to accelerate the pace of SME innovations", *Journal of Innovation & Knowledge*, Vol. 6 No. 1, pp. 43-49. https://doi.org/10.1016/j.jik.2020.10.001

Huarng, K.-H., Guaita-Martínez, J.M. and Yu, T.K.-Y. (2021), "Challenges and opportunities of new research methods in innovation, entrepreneurship, and knowledge", *Journal of Business Research,* Vol. 134, pp. 101-103. https://doi.org/10.1016/j.jbusres.2021.04.058

Immonen, A., Ovaska, E. and Paaso, T. (2018), "Towards certified open data in digital service ecosystems", *Software Quality Journal,* Vol. 26, pp. 1257–1297. https://doi.org/10.1007/s11219-017-9378-2

Immonen, A., Palviainen, M. and Ovaska, E. (2014), "Requirements of an open data based business ecosystem", *IEEE Access,* Vol. 2, pp. 88-103. https://doi.org/10.1109/ACCESS.2014.2302872

International DOI Foundation (2016). *DOI® Handbook*. Last updated: October 17, 2016. Retrieved 27/03/2022 from https://www.doi.org/doi_handbook/TOC.html

Janssen, M. and Zuiderwijk, A. (2014), "Infomediary business models for connecting open data providers and users", *Social Science Computer Review,* Vol. 32 No. 5, pp. 694-711. https://doi.org/10.1177/0894439314525902

Jennex, M. and Durcikova, A. (2014), "Integrating IS Security with Knowledge Management: Are We Doing Enough to Thwart the Persistent Threat?", in 2014 47th Hawaii International Conference on System Sciences, pp. 3452-3459. https://doi.org/10.1109/HICSS.2014.429

Jiménez-Barrinuevo, M.M., García-Morales, V.J. and Molina, L.M. (2011), "Validation of an instrument to measure absorptive capacity", *Technovation*, Vol. 31, pp. 190-202. https://doi.org/10.1016/j.technovation.2010.12.002

Kampars, J., Zdravkovic, J., Stirna, J. and Grabis, J. (2020), "Extending organizational capabilities with open data to support sustainable and dynamic business ecosystems", *Software and Systems Modeling*, Vol. 19 No. 2, pp. 371-398. https://www.doi.org/10.1007/s10270-019-00756-7

Klaila, D. (2000), "Knowledge as a transformation agent", *Journal of Knowledge Management*, Vol. 4 No. 2, pp. 138-144. https://doi.org/10.1108/13673270010372260

Kitsios, F., Papachristos, N. and Kamariotou, M. (2017), "Business models for open data ecosystem: Challenges and motivations for entrepreneurship and innovation", P. Loucopoulos, Y. Manolopoulos, O. Pastor, B. Theodoulidis, & J. Zdravkovic (Eds.), *IEEE 19th conference on business informatics (CBI)*, pp. 398-407. https://doi.org/10.1109/CBI.2017.51

Kokshagina, O., Le Masson, P. and Bories, F. (2017), "Fast-connecting search practices: on the role of open innovation intermediary to accelerate the absorptive capacity", *Technological Forecasting and Social Change*, Vol. 120, pp. 232-239. https://doi.org/10.1016/j.techfore.2017.02.009

Koulouriotis, D.E. and Emiris, D.M. (2004), "Towards a holistic knowledge management model", *Journal of Knowledge Management*, Vol. 8 No. 1, pp. 32-46. https://doi.org/10.1108/13673270410523899

Lee, S.M. and Trimi, S. (2018), "Innovation for creating a smart future", *Journal of Innovation & Knowledge,* Vol. 3 No. 1, pp. 1-8. https://doi.org/10.1016/j.jik.2016.11.001

Lichtenthaler, U. (2011), "Open innovation: past research, current debates, and future directions", *Academy of Management Perspectives* Vol. 25 No. 1, 75–93. http://www.jstor.org/stable/23045037

Lichtenthaler, U. and Lichtenthaler, E. (2009), "A Capability-Based Framework for Open Innovation: Complementing Absorptive Capacity", *Journal of Management Studies*, Vol. 46 No. 8, pp. 1315-1338. https://doi.org/10.1111/j.1467-6486.2009.00854.x

Lindman, J. (2014), "Similarities of open data and open source: Impacts on business", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 9 No. 3, pp. 59-70. https://doi.org/10.4067/S0718-18762014000300006

López-Cabarcos, M.A., Piñeiro-Chousa, J. and Quiñoá-Piñeiro, L. (2021), "An approach to a country's innovation considering cultural, economic, and social conditions", *Economic Research-Ekonomska Istraživanja,* Vol. 34 No. 1, pp. 2747-2766. https://www.doi.org/10.1080/1331677X.2020.1838314

Lu, I.-Y. and Tseng, C.-J. (2010), "A study of the service innovation activities of tourist hotels in Taiwan", *International Journal of Organizational Innovation,* Vol. 3 No. 1, pp. 156-172.

Lytras, M.D., Visvizi, A. and Chui, K.T. (Eds.) (2020), *Big Data Research for Social Sciences and Social Impact*. Book printed edition of the Special Issue Big Data

Research for Social Sciences and Social Impact, Sustainability, Switzerland: MDPI. https://doi.org/10.3390/books978-3-03928-221-0

Malhotra, Y. (2000), "Knowledge management and new organization forms: a framework for business model innovation", *Information Resources Management Journal*, Vol. 13 No. 1, pp. 5-14. https://www.doi.org/10.4018/irmj.2000010101

Naqshbandi, M.M. (2016), "Managerial ties and open innovation: examining the role of absorptive capacity", Management. Decision, Vol. 54, pp. 1-25. https://www.doi.org/10.1108/MD-03-2016-0161

Öberg, C. and Alexander, A.T. (2019), "The openness of open innovation in ecosystems – Integrating innovation and management literature on knowledge linkages", *Journal of Innovation & Knowledge,* Vol. 4 No. 4, pp. 211-218. https://doi.org/10.1016/j.jik.2017.10.005

Okatan, K. (2012), "Managing Knowledge for Innovation and Intra Networking: A Case Study", *Procedia - Social and Behavioral Sciences,* Vol. 62, pp. 59-63. https://doi.org/10.1016/j.sbspro.2012.09.012

Open Knowledge Foundation (2008). *The Open Definition*. Retrieved 31/03/2022 from https://opendefinition.org/od/2.1/en/

Pasquetto, I.V., Borgman, C.L. and Wofford, M. F. (2019), "Uses and reuses of scientific data: The data creators' advantage", *Harvard Data Science Review*, Vol. 1 No. 2, pp. 1-32. https://doi.org/10.1162/99608f92.fc14bf2d

Pasquetto, I.V., Randles, B.M. and Borgman, C.L. (2017), "On the reuse of scientific data", *Data Science Journal*, Vol. 16 No. 8, pp. 1-9. http://doi.org/10.5334/dsj-2017-008

Rajapathirana, R.P.J. and Hui, Y. (2018), "Relationship between innovation capability, innovation type, and firm performance", *Journal of Innovation & Knowledge*, Vol. 3 No. 1, pp. 44-55. https://doi.org/10.1016/j.jik.2017.06.002

Rautenberg, S., Motyl Kaue, S., Cassiana Burda, A., Anderson, S. and Marom de Moura, F. (2017), "Dados abertos conectados e gestão do conhecimento: estudos de caso cientométricos em uma universidade brasileira", *Perspectivas em Ciência da Informação*, Vol. 22 No. 3, pp. 116-142. https://doi.org/10.1590/1981-5344/2885

Rauter, R., Globocnik, D., Perl-Vorbach, E. and Baumgartner, R.J. (2019), "Open innovation and its effects on economic and sustainability innovation performance", *Journal of Innovation & Knowledge*, Vol. 4 No. 4, pp. 226-233. https://doi.org/10.1016/j.jik.2018.03.004

Sadiq, A. and Indulska, M. (2017), "Open data: Quality over quantity", *International Journal of Information Management*, Vol.37 No. 3, 150-154. https://doi.org/10.1016/j.ijinfomgt.2017.01.003

Smith, K. (2004), "Measuring innovation", J. Fagerberg, D.C. Mowery, & R.R. Nelson (Eds.), *The Oxford handbook of innovation*, University Press, Oxford, pp. 148–170.

Steinberg, M. and Brehm, J. (2009), "Towards utilizing open data for interactive knowledge transfer", in *Proceedings of 9th International Conference on Mobile, Hybrid, and On-line Learning, Cancun*, Mexico, pp. 61-66. Retrieved 27/03/2022 from https://ieeexplore.ieee.org/abstract/document/4782604

Taques, F.H., López, M.G., Basso, L.F. and Areal, N. (2021), "Indicators used to measure service innovation and manufacturing innovation", *Journal of Innovation & Knowledge*, Vol. 6 No. 1, pp. 11-26. https://doi.org/10.1016/j.jik.2019.12.001

Trombino, D., La Cascia, M. and Aiello, G. (2017), *Data and Metadata Quality Measurement Software in a Smart City Context*, Thesis, Scuola Politecnica, Academic Year 2016/2017, Università Degli Studi Di Palermo (Italia).

Vilas Boas Viveiros Lopes, A.P. and Monteiro de Carvalho, M. (2018), "Evolution of the open innovation paradigm: Towards a contingent conceptual model", *Technological Forecasting & Social Change*, Vol. 132, 284-298. https://doi.org/10.1016/j.techfore.2018.02.014

Waltz, E. (2003), *Knowledge Management in the Intelligence Enterprise*, Artech House, Boston, MA.

Wiig, K.M. (1997), "Integration intellectual capital and knowledge management", *Long Range Planning*, Vol. 30 No. 3, pp. 399-405. https://doi.org/10.1016/S0024-6301(97)90256-9

Witell, L., Snyder, H., Gustafsson, A., Fombelle, P. and Kristensson, P. (2016), "Defining service innovation: A review and synthesis", *Journal of Business Research*, Vol. 69 No. 8, pp. 2863-2872. https://doi.org/10.1016/j.jbusres.2015.12.055

Wu, I.-L. and Hu, Y.-P. (2018), "Open innovation based knowledge management implementation: a mediating role of knowledge management design", *Journal of Knowledge Management*, Vol. 22 No. 8, pp. 1736-1756. https://doi.org/10.1108/JKM-06-2016-0238

Zimmermann, H. and Pucihar, A. (2015), "Open innovation, open data and new business models", in Petr, D., Gerhard, C., & Vaclav, O. (Ed.s), Schriftenreihe Informatik, 44, Universitatsverlag Rudolf Trauner, Linz, pp. 449-458. https://doi.org/10.2139/ssrn.2660692

Zobel, A.-K. (2017), "Benefiting from open innovation: a multidimensional model of absorptive capacity", Journal of Product Innovation Management, Vol. 34, pp. 269-

288. https://doi.org/10.1111/jpim.12361

Zuiderwijk, A., Janssen, M. and Davis, C. (2014), “Innovation with open data: Essential elements of open data ecosystems”, *Information Polity*, Vol. 19 No. 1-2, pp. 17-33. https://doi.org/10.3233/IP-140329

Tables

Table 1.

Capabilities for open innovation and value creation through open data.

Inputs	Process	Results
Knowledge	**Open innovation model for knowledge management**	**Value creation**
Open data ecosystems	• Inventive capacity (generate and exploit open data internally) • Absorptive capacity (exploit external knowledge and use internally) • Transformative capacity (maintain acquired knowledge over time) • Connective capacity (maintain knowledge in inter-firm relationships) • Innovative (generate innovations from open data) • Desorptive capacity (pass the knowledge to the market)	• Products / services (digital, apps…) • Open data business models

Source: own elaboration

Table 2.

Open data impact process for open innovation model.

Open innovation type / reusers categories	**Open data impact process**			
	Phase 1: Candidate data	**Phase 2: Published data**	**Phase 3: Reused data**	**Phase 4: Impact**
Type of open innovation	Outbound	Outbound	Inbound Coupled	Outbound Inbound Coupled
Agent type	Primary open data source	Primary open data source	Direct End users	Primary Direct reusers End users

Source: Corrales-Garay et al. (2020) adapted from Corrales-Garay et al. (2019b).

Table 3.

Sample of datasets.

Catalogue	Dataset' URL
CKAN	https://opendata-ajuntament.barcelona.cat/data/es/dataset/est-cadastre-carrecs-tipus-propietari
CKAN	https://datosabiertos.rivasciudad.es/dataset/calidad-del-aire
CKAN	https://datosabiertos.ayto-arganda.es/dataset/contratos-mayores-4-trimestre-2016
CKAN	https://datos.alcobendas.org/dataset/subvenciones-2016-asociadiones-de-salud-consumo-integracion-social-y-mayores
ODS	https://gijon.opendatasoft.com/explore/dataset/contratos-menores-adjudicados/table/
ODS	https://gijon.opendatasoft.com/explore/dataset/presupuesto-de-gastos-de-la-fundacion-municipal-de-cultura-educacion-y-universi3/table/
ODS	https://dipcas.opendatasoft.com/explore/dataset/planeamiento-urbanistico/table/
ODS	https://dipcas.opendatasoft.com/explore/dataset/resultados-electorales-provincia-de-castellon/information/
Socrata	https://analisi.transparenciacatalunya.cat/Sector-P-blic/Pressupostos-dels-ens-locals-Dades-b-siques/kv4y-3ks8
Socrata	https://analisi.transparenciacatalunya.cat/Medi-Rural-Pesca/Sistema-d-informaci-geogr-fica-de-parcel-les-agr-c/uq9g-cc59
Socrata	https://opendata.l-h.cat/Urbanisme-i-infraestructures/Guia-oficial-de-noms-de-carrers/mxs6-mjeq
Socrata	https://opendata.l-h.cat/Cultura-i-oci/Agenda-de-la-ciutat/qtv3-9x52
Arcgis Open data	https://www.opendatalapalma.es/datasets/vegetacion-afectada-incendio-07082016/explore?location=28.562271%2C-17.844553%2C13.12
Arcgis Open data	https://opendata.dadesobertesmanlleu.cat/datasets/SITUAM::projectes-demprenedoria/explore?location=41.959443%2C2.217957%2C12.07
Arcgis Open data	https://www.opendatalapalma.es/datasets/parte-de-incendios-2014/explore
Arcgis Open data	https://opendata.dadesobertesmanlleu.cat/datasets/SITUAM::punts-daccidentalitat-atropellaments/explore?location=42.003384%2C2.277053%2C15.83
DKAN	https://dadesobertes.diba.cat/datasets/exposicions
DKAN	https://dadesobertes.diba.cat/datasets/agenda-dels-carrecs-electes-de-la-diputacio
DKAN	https://dadesobertes.diba.cat/datasets/agenda-general-de-la-diputacio
DKAN	https://dadesobertes.diba.cat/datasets/perfil-del-compromis-ambiental-dels-membres-de-la-xarxa-de-ciutats-i-pobles-cap-a-la

Source: own elaboration.

Table 4. Examples of OpenDatId and results obtained from open-source software.

OpenDatId	MELODA 5 Values
00000005#https://opendata-ajuntament.barcelona.cat/data/api/3/action/package_show?id=bc2c4827-3e71-4492-b248-82234ee03b84#20210618T1149.v1	29. BASIC
00000097#https://datosabiertos.rivasciudad.es/api/3/action/package_show?id=6f1bbf39-d5ca-4459-8430-ffb9e03f5738#20191030T1201.v1	32. BASIC
00000026#https://gijon.opendatasoft.com/api/v2/catalog/datasets/contratos-menores-adjudicados?timezone=UTC#20220316T0005.v1	29. BASIC
00000026#https://gijon.opendatasoft.com/api/v2/catalog/datasets/presupuesto-de-gastos-de-la-fundacion-municipal-de-cultura-educacion-y-universi3?timezone=UTC#20180115T1358.v1	29. BASIC
00000043#https://dipcas.opendatasoft.com/api/v2/catalog/datasets/planeamiento-urbanistico?timezone=UTC#20181231T0000.v1	25. BASIC
00000043#https://dipcas.opendatasoft.com/api/v2/catalog/datasets/resultados-electorales-provincia-de-castellon?timezone=UTC#20190712T0000.v1	25. BASIC
00000025#https://analisi.transparenciacatalunya.cat/api/views/metadata/v1/kv4y-3ks8#20220316T0000.v1	27. BASIC
00000025#https://analisi.transparenciacatalunya.cat/api/views/metadata/v1/uq9g-cc59#20220112T0000.v1	30. BASIC
00000115#https://opendata.l-h.cat/api/views/metadata/v1/mxs6-mjeq#20170614T0000.v1	25. BASIC
00000115#https://opendata.l-h.cat/api/views/metadata/v1/qtv3-9x52#20170614T0000.v1	30. BASIC
00000041#https://services.arcgis.com/hkQNLKNeDVYBjvFE/arcgis/rest/services/Perimetro_incendio_veg_2016/FeatureServer/0?f=pjson#20190912T0000.v1	30. BASIC
00000099#https://services.arcgis.com/WNsrZDHEJ88NE4N8/arcgis/rest/services/Projectes_Emprenedoria/FeatureServer/0?f=pjson#20220118T0000.v1	30. BASIC
00000041#https://services.arcgis.com/hkQNLKNeDVYBjvFE/arcgis/rest/services/Parte_de_incendios_2014/FeatureServer/0?f=pjson#20150218T0000.v1	25. BASIC
00000099#https://services.arcgis.com/WNsrZDHEJ88NE4N8/arcgis/rest/services/Punts_accidentalitat_atropellaments/FeatureServer/0?f=pjson#20180917T0000.v1	30. BASIC
00000033#https://dadesobertes.diba.cat/api/3/action/package_show?id=f279b7bf-9f81-44e9-a552-bf4c528d2c6b#20140528T1241.v1	25. BASIC
00000033#https://dadesobertes.diba.cat/api/3/action/package_show?id=9e55f0cc-841c-496c-ab6c-23aba9a79c5e#20210718T0855.v1	30. BASIC
00000033#https://dadesobertes.diba.cat/api/3/action/package_show?id=2cdb8a5e-394e-4f4a-bbaf-00ab9d4d20fb#20200327T2138.v1	30. BASIC
00000033#https://dadesobertes.diba.cat/api/3/action/package_show?id=4bbe2741-8a35-4f87-84c8-bd84d2d5f506#20200306T1321.v1	22. INADECUATE
00000088#https://datosabiertos.ayto-arganda.es/api/3/action/package_show?id=9f5f42d8-138c-461d-97f1-e040c368abe8#20180325T1541.v1	25. BASIC
00000029#https://datos.alcobendas.org/api/3/action/package_show?id=fd6f4abd-4aca-43e3-94ff-16f7d151cc9d#20170119T1158.v1	25. BASIC

Source: own elaboration.