Tracking of Artificial Intelligence Adoption in ProgrammableWeb Directory

Blagoj Delipetrev and **Uroš Kostić** and **Lorenzino Vaccari** and **Francesco Pignatelli** [1]

Abstract. This research presents an analysis of AI adoption based on the analysis of the ProgrammableWeb directory. ProgrammableWeb.com is known as the Web's most popular open Application programming interfaces (APIs) repository. Ever since it was founded in 2005, it has been chronicling the daily evolution of the global API, and it is the most relied-on directory when it comes to discovering and searching for APIs. With owner's permission, we downloaded the ProgrammableWeb directory dataset, containing 21609 APIs, 1641 libraries, 15329 SDKs, 7629 mashups, and 10989 sample source codes. These resources were analysed to have an indication/measure on the AI adoption related to the API domain. Overall, the analysis showed increasing artificial intelligence adoption trends from 2005 onwards.

1 INTRODUCTION

Digital technologies and artificial intelligence are changing every dimension of our life, including our interaction with public government authorities and services. This digital transformation has an important impact on citizens, the private sector and the same public administration units acting in different policy areas. Within the public sector, the digital transformation designates the process of conversion or substitution of analogue public administration operations to their digital counterparts. Information and Communication Technologies (ICTs) provide tools for faster and more efficient processing of data within public administration units, where more efficient public services can result in significant cost savings or development of new kinds of services for the same cost.

As ICT allows the faster and more efficient processing of data, increased efficiency may lead to cost savings or development of new kinds of intelligent services. It can be seen that APIs, machine to machine interfaces that facilitate software exchanging data and services, have become a foundational technological component of modern digital architectures, affecting every sector of the global economy. Although not novel as a technological solution, their capability to enable machine-to-machine communication makes them instrumental for a functional digital fabric for both the private and the public sectors. Moreover, they can be considered key enablers of the accelerated evolution of governments and their agencies, from analogue (manual, paper) operations and workflows to operations and workflows being implemented by digital artefacts, making them faster, more flexible and interoperable.

A major reason for recent interest in APIs is their role as the foundation of digital ecosystems. APIs make it easy for individuals to write programs that communicate with online services and shared databases and have been used by many companies since almost twenty years ago to empower their business. Although seemingly a very simple infrastructure, APIs are a fundamental component which has been used by big companies such as Amazon, Google, Facebook, and Twitter, for making powerful digital platforms available to third party application developers. The values of APIs are not only established by its creators and users, but also are a critical component that lets other parties to connect among heterogeneous systems that can reuse them in ways that were not forecast and can potentially create digital innovations.

Artificial intelligence (AI) is on path to become a general purpose technology because of its broad spectrum of application in nearly every industry. The broad spectrum of AI applications are found in computer vision, speech recognition, natural language understanding, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection, games, etc. Monitoring AI evolution and its impact in different industries becomes a major challenge because of the AI wide spread and rapid advancements.

The AI Watch project monitors European Union's industrial, technological and research capacity in AI, AI-related policy initiatives in the Member States, uptake and technical developments of AI, and AI impact. AI Watch has a European focus within the global landscape. In the context of AI Watch, the EC works in coordination with Member States. AI Watch results and analyses are published on the AI Watch Portal[2].

A number of relevant studies are in progress in the frame of the European Location Interoperability Solutions for e-Government action (ELISE)[3], part of the ISA^2 Programme. This Action supports Better Regulation and Digital Single Market Strategy goals, including specific actions of the e-Government Action Plan[4] and the European Interoperability Framework[5], which are reinforced by the Tallinn Declaration[6] vision and the Communications on Building the Data Economy[7] and on Artificial Intelligence for Europe[8]. In particular, ELISE studies explore the role of location information in digital government and the technologies involved in delivering innovative public services. It is worth mentioning, the "Digital Government Benchmark – API study"[9] provides an early-stage analysis of Web APIs as en-

[1] European Commission, Joint Research Centre (JRC), Italy, email: uros.kostic@ext.ec.europa.eu

[2] https://ec.europa.eu/knowledge4policy/ai-watch_en
[3] https://ec.europa.eu/isa2/actions/elise_en
[4] https://ec.europa.eu/digital-single-market/en/european-egovernment-action-plan-2016-2020
[5] https://ec.europa.eu/isa2/eif_en
[6] http://ec.europa.eu/newsroom/document.cfm?doc_id=47559
[7] https://eur-lex.europa.eu/content/news/building_EU_data_economy.html
[8] https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe
[9] https://joinup.ec.europa.eu/collection/elise-european-location-

ablers for the digital transformation of government. A multiple-case study comparative analysis has been applied to selected cases, with a particular focus on geospatial API. Another report under finalisation is a "Landscape of Artificial Intelligence (AI) in the public sector", aiming at assessing 100+ cases to examine the impact and the added value of AI. In selected cases, the linkages between AI and the data governance is explored.

Since there is no single knowledge stream from which to monitor and assess AI uptake, there is a need to merge several data streams based on a unified analytical framework starting from scientific discovery, to product engineering, marketing and business development, and topic coverage by the media. One of the data streams is the ProgrammableWeb[10] (PW) which we used to track the adoption of AI.

Another important reason to use APIs to monitor AI uptake is that the rapid increase in data volumes makes it unsustainable to transfer big data to centralised computing platforms, or even clouds. This problem will become more acute in the coming age of the Internet of Things, with billions of data collecting sensors connected to each other and to the Web (the Web of Things). The emerging pattern is therefore to shift more and more processing, including the AI-powered one, to, or close to, sensing devices (the so-called edge computing) [1]. As an example, the European Space Agency will soon be launching its first Φ-sat, a satellite with an AI chip on board to pre-process the data and reduce significantly the volume of data transmission, which may show the trend that, in the near future, data and processing will be performed more frequently together with or close to the source and later made accessible via APIs.

To analyse the PW directory, we scraped it (with permission from the editors) and stored the data into a database. We performed a full text search in the *description* and the *abstract* fields of the API, Library, Mashup, Software Development Kit (SDK), and SampleSourceCode collections in the PW database using AI related keywords. Several Python scripts were prepared to query the database, analyse the data and plot the results. In the following, we refer to the entries in the API, Library, Mashup, SDK, and SampleSourceCode collections as "resources".

2 THE ProgrammableWeb.com AND Programmable DIRECTORY

As the world's leading source of news and information about Internet-based APIs, ProgrammableWeb.com is known as the Web's de-facto journal of the API economy. Ever since it was founded in 2005, the ProgrammableWeb has been chronicling the daily evolution of the global API economy while amassing the Web's most relied-on directory when it comes to discovering and searching for APIs to use in Web and mobile applications. The ProgrammableWeb is where you can keep up with what's new and interesting in a world where the Web is a programmable platform. When we say "the Web is a platform," we are referring to how Web-based and mobile applications are enabled by Internet-based APIs. For example, the way in which the developers of many location-aware applications are able to incorporate Google Maps into their software/applications with just a few lines of code (using the Google Maps API).

The ProgrammableWeb directories are not only about APIs, but also about cataloguing SDKs, libraries, frameworks, Web apps and applications that consume APIs (primarily mashups and mobile apps), and pointers to sample source code found elsewhere on the Web. Since the users can upload the information on voluntary basis, this directory does not provide a complete list of all APIs around the world. Nevertheless, this is the largest directory, therefore, it is reasonable to assume that it can be used to provide insights on global API trends.

Since ProgrammableWeb is predominantly concerned with the Web as a programmable platform, the API directory accommodates Web and other APIs that enable developers to build Web and mobile apps. These include typical RESTful APIs, SOAP XML RPC APIs, Javascript APIs found in browsers and even device specific APIs as long as they are mashable into Web apps.

3 THE ProgrammableWeb DATABASE

We obtained the information from the ProgrammableWeb directory via a collaboration with the owners of the directory, the ProgrammableWeb.com and Mulesoft and store it into a MongoDB database securely maintained at the JRC premises.

The directory's contains 21609 APIs, 1641 libraries, 15329 SDKs, 7629 mashups, and 10989 sample source codes (see Fig.1)

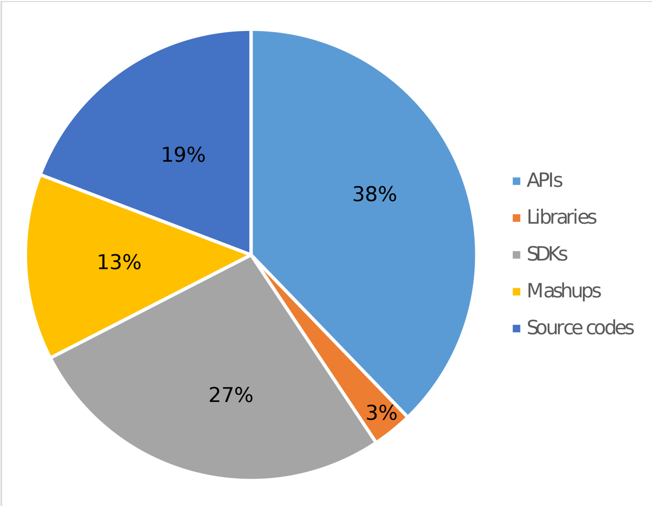

Figure 1. The number of ProgrammableWeb resources

The high level structure of the database is presented in the following sections.

3.1 The structure of the database

The MongoDB database (DB) collections strictly follow the structure of the PW directory, i.e. each "tab" in the directory entry for an API is mapped to a collection in the DB (e.g. the API directory is mapped to the API collection). For example, an API in the directory may contain the following tabs: SDKs, Articles, How To, Source Code, Libraries, Developers, Followers, Changelog. Each of these tabs is mapped to a collection in the DB and the web links between these tabs have been implemented as relations in the DB. For example, under each web page tab, there is a list of resources linked to the API, e.g. the tab "SDKs" contains a list of SDKs which are related to the API.

Each resource in the list under a tab leads to a web page for this particular resource. This web page contains fields with information about the resource, and these fields are mapped to keys in the collection. For example, in a list of SDKs each SDK resource links to

interoperability-solutions-e-government/document/report-digital-government-benchmark-api-study
[10] https://www.programmableweb.com

a web page which may contain the following fields: Related APIs, Related Platform/Languages, Categories, Added, SDK Provider, etc. and each of these fields is mapped to the collection keys.

Figure 2 gives the conceptual overview of the DB structure, showing relations between the main collections gathered to build the database. Not all collections are shown in this diagram and the key names are omitted.

One of the most interesting collections is the one related to categories of the PW database. These categories were used to classify in the same way the main collections of the PW directory, including APIs and Mashups.

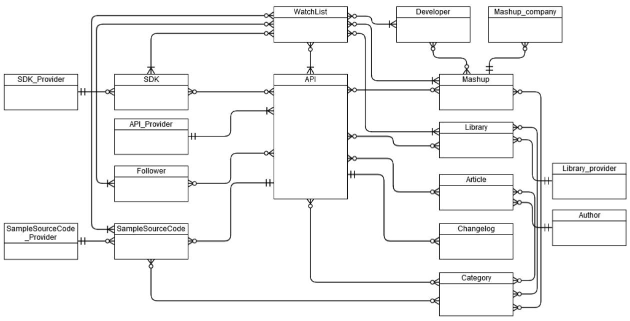

Figure 2. Conceptual view of the database

From Fig. 2, we can identify the following main collections and their relationships:

- API – Application Programming Interface, an interface or communication protocol between different parts of a computer program,
- Library – a collection of non-volatile resources used by computer programs, often for software development,
- SDK – Software Development Kit, a collection of software development tools in one installable package,
- Mashup – a web page or web application that uses content from more than one source to create a single new service displayed in a single graphical interface.

Each collection contains information from the PW, e.g. the API collection from the database contains a list of documents, where each document from the database maps to an API entry in the PW directory and contains information scraped for that particular API. Similarly is true for the Library, SDK, and Mashup collections. The relation between all four main collections is such that a document in the Library, SDK and Mashup collections is linked to an API document if it is listed under this API in the PW directory, e.g. the Twilio C++ REST SDK entry is listed under the Twilio API entry in the PW directory, therefore, the Twilio C++ REST SDK document in the SDK collection is also linked to the Twilio API document in the API collection in the DB. The link is done through the "_id" keys in both collections.

There are no direct links between Library, SDK, and Mashup collections – all the linking goes through the API collection.

3.1.1 Additional and auxiliary collections

The additional collections provide the information on:

- WatchList – a collection of developers' watch lists,
- SampleSourceCode – examples of source code using the API,
- Developer – a collection of developers registered on PW website, using the resources (APIs, Libraries, Mashups etc.),
- Mashup_company – a collection of companies providing the mashups,
- *_provider – a collection of providers for the APIs, SDKs, sample source codes and libraries,
- Follower – a collection of followers; a follower is a user who follows the updates about APIs, SDKs, developers, etc.,
- Article – a collection of articles about the APIs, SDKs, libraries, etc. This collection includes also HowTo's, which are a treated as a separate class in the PW directory!
- ChangeLog – log of all changes to Programmable web assets (e.g. APIs, Articles, etc.),
- Author – a collection of authors of articles,
- Category – a collection of all possible categories for all main collections.

These collections contain the information scraped from the PW using the same "mapping principle" as described above, for the main collections.

3.1.2 Overview of the main collections

The API collection can be considered as the primary, or the top collection, while the SDK, Library, and the Mashup collections as the secondary. These secondary collections are related to the API collection (e.g. a Mashup which uses a certain API is related to this API).

The "category" collection provides a list of categories, which briefly describe the basic applicability of the resource. The API collection keeps this information in the "Primary_category" key, which links to the appropriate document's "_id" and "name" keys in the "category" collection. Each API has only one primary category. In addition, each API can have multiple secondary categories, which provide additional description of the resource.

The SDK, Library, and Mashup collections are also described briefly by a category, however, in this case, there is no distinction between the primary and secondary categories and multiple categories can be assigned to each document.

4 ARTIFICIAL INTELLIGENCE ADOPTION IN THE ProgrammableWeb DIRECTORY

For the purpose of this research, we performed a full text search in the *description* and the *abstract* fields of the API, Library, Mashup, SDK, and SampleSourceCode (SSK) collections in the PW database using AI related keywords. The AI related keywords were taken from the AI Watch "Defining Artificial Intelligence" technical report which contains an operational definition and taxonomy of AI.[11] The results of the AI adoption analysis in the PW directory are the following:

- As shown in Fig.3 the number of all resources in PW is continuously increasing since its inception in 2005. The biggest growth of new resources was in 2017 with almost $10,000$ new entries. Among different resources, the more dominant ones are APIs, SDKs and SSKs, while Mashups and libraries are not well represented.

[11] The report is still in progress of publishing.

- The number of resources containing AI keywords has grown steadily from 2015 onwards as shown in Fig.4. APIs was a major AI adopter compared to other resources, especially until 2014. Since 2015, AI was also adopted by SDKs and SSKs. The trends on total number of resources and AI adoption among resources are quite similar as shown in Fig.3 and Fig.4.
- The percentage of AI adopted newly created resources per year confirms that the ratio of AI adoption has been increasing until 2014, and after that it remains approximately unchanged, as shown in Fig.5.
- The frequency of AI keywords in the resources in all years show that the most used keywords are: machine learning, internet of things, artificial intelligence, sentiment analysis, natural language processing, and classification[12] as shown in Fig.6.
- The results of grouping the results from Fig.6 into AI subdomains: the most represented domain is "AI Applications, Platforms, Software services", which is followed by "Machine Learning" and "Natural Language Processing", shown in Fig.7. In these domains, the APIs are the most widely used type of resource.
- Grouping all the different types of resources (i.e. APIs, SDKs, etc.) to see how the frequency of subdomains changes with time is presented in Fig.8.

There is a trend of AI adoption in all the resources from 2005 onwards.

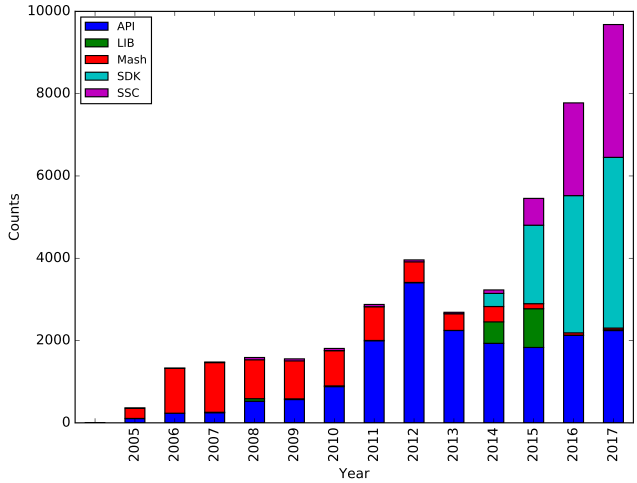

Figure 3. Number of all resources added per year

5 CONCLUSIONS

In this document, we briefly presented the set of data gathered from an important source, namely, the ProgrammableWeb directory. From this directory, the evidence on the adoption of web APIs for many purposes and in many fields can be extracted and analysed, with PW being the primary community resource for amateurs and professionals in the industry. This resource gathers public API end-points in a comprehensive directory with information that is self-reported by developers.

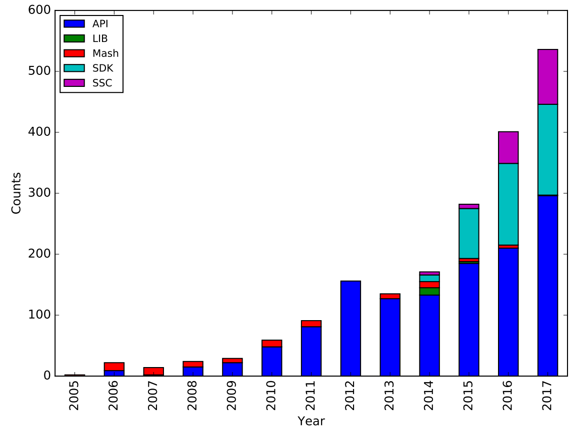

Figure 4. Number of resources containing AI keywords added per year

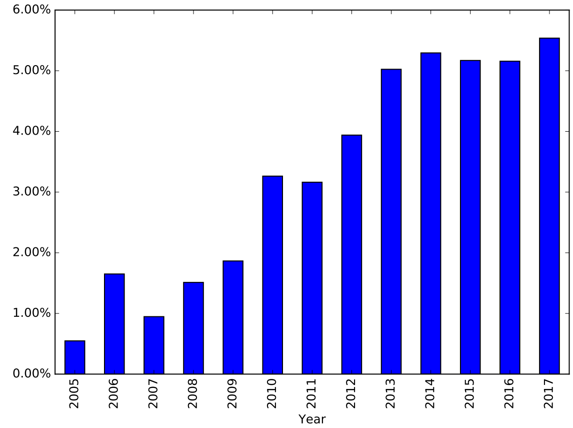

Figure 5. Percentage of AI related, newly created resources added per year

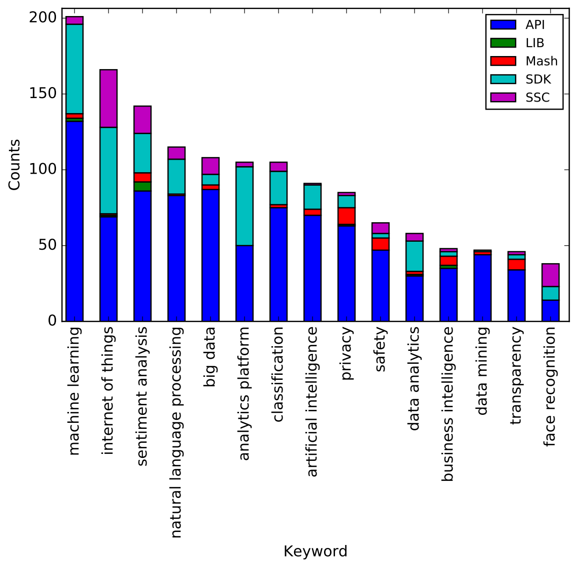

Figure 6. Frequency of the first 15 most commonly found keywords in the resources in all years

[12] Actually, the most frequent keyword was "security" with about 700 appearances. However, we excluded it from the analysis because this keyword, if used independently, encompasses also many non-AI related resources, dealing with non-AI aspects of security.

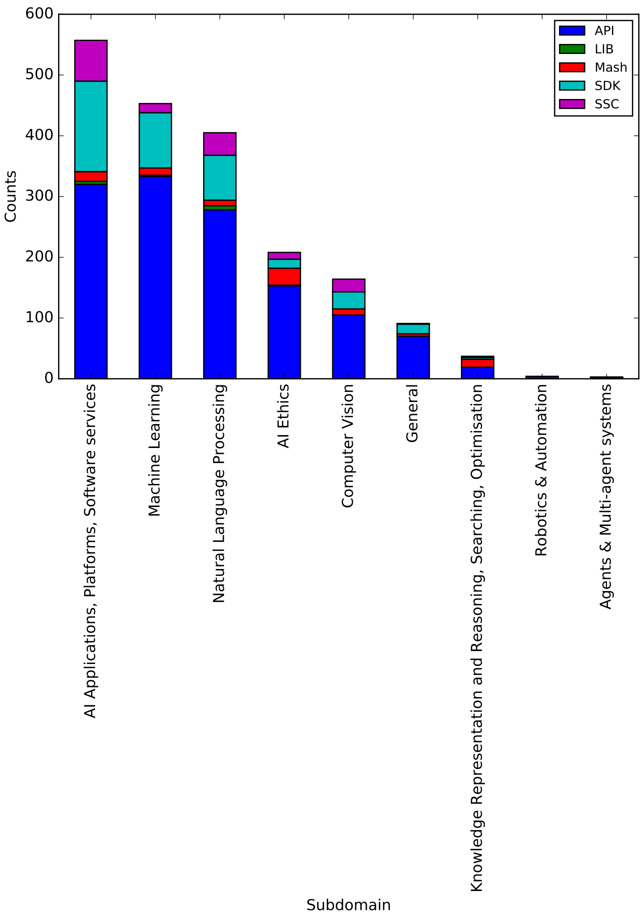

Figure 7. Frequency of AI subdomains in the resources in all years

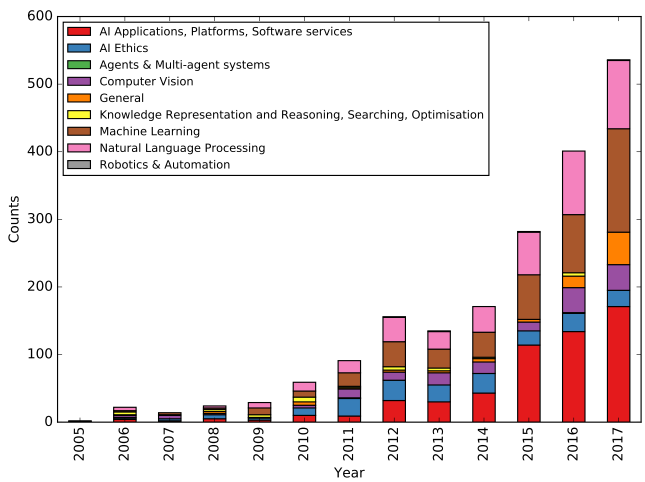

Figure 8. Frequency of new resources per year, grouped by subdomain

After a brief introduction of what PW is, we illustrated the structure of the set of collections, stored in a physical database we built by scraping the ProgrammableWeb web site, in particular focussing on the main classes (i.e. API, SDK, Mashups and Libraries).

In the last part of the document, we presented analysis of the use of APIs, SDK, Mashups and libraries in AI adoption. There is a general increase in number of AI related reosurces per year, showing the increasing trend of AI adoption across all resources.

Even if some limitations have to be considered, such as the fact that the PW directory is built on a voluntary basis (and so it does not represent the complete state of the art of APIs and related subjects), we think that the database could represent a valid support for many analyses.

ACKNOWLEDGMENTS

This research has been prepared as part of the study 'APIs4DGov – digital government APIs: the road to value-added open API-driven services' and will contribute to the study's final outputs in early 2020.

Performed by the Joint Research Centre (JRC) of the European Commission, in collaboration with the Directorate-General (DG) for Communications Networks, Content and Technology, the study is being conducted in the context of the European Commission's Digital Single Market strategy and aims to improve the understanding of the current use of APIs in digital government and their added value, as well as to assess the feasibility of establishing a European API framework for digital government.

This work was made possible due to contributions by several people, including many colleagues of the unit B6 – Digital economy of the Joint Research Centre. We would like in particular to thank Giuditta De Prato, Montserrat Lopez Cobo, Robin Smith and Monica Posada for their insightful comments.

We would like to thank our colleague, Francesco Pignatelli, that leads the ELISE action of the ISA^2 programme, under which this work has been done.

We would also like to thank David Berlind, editor in Chief of ProgrammableWeb for his constant support, comments, feedback and collaboration in giving us the possibility to access the ProgrammableWeb set of data. We, of course, thank the entire ProgrammableWeb company and team.

REFERENCES

[1] M. Craglia, A. Annoni, P. Benczur, P. Bertoldi, B. Delipetrev, G. De Prato, C. Feijoo, E. Fernandez Macias, E. Gomez, M. Iglesias, H. Junklewitz, M. López Cobo, B. Martens, S. Nascimento, S. Nativi, A. Polvora, I. Sanchez, S. Tolan, I. Tuomi, and L. Vesnic Alujevic, *Artificial Intelligence – A European Perspective*, chapter 11, Publications Office, Luxembourg, 2018. Report EUR 29425 EN.