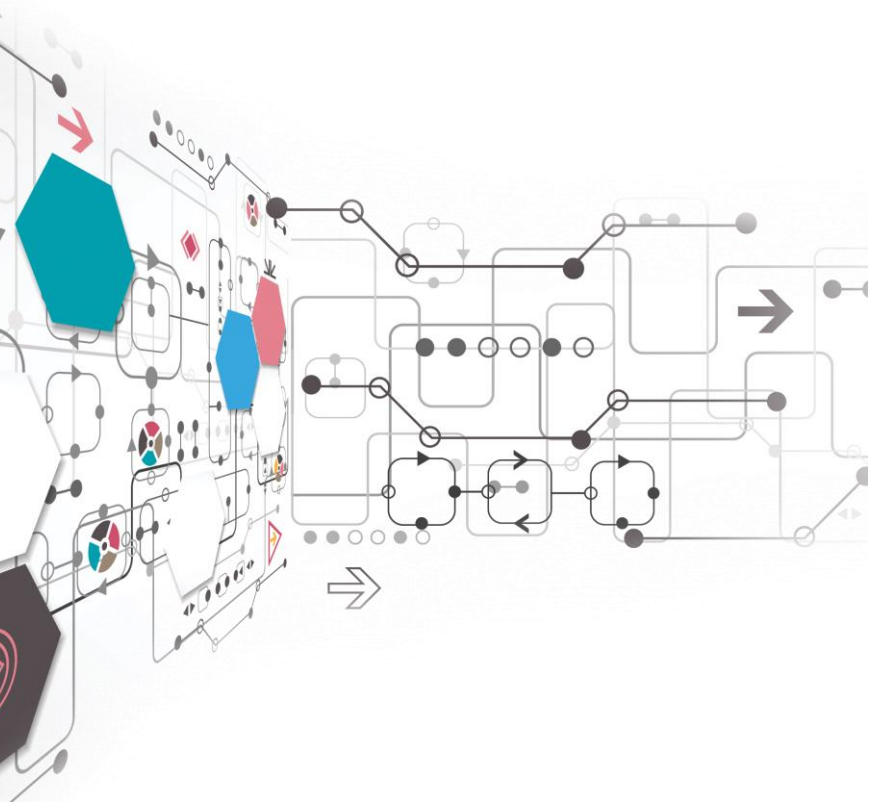

ABSTRACT This document, as part of WP1 (Management and Coordination), presents the workplan for development of common open data language which will serve as an input for a whole TODO project.

Author
University of Zagreb, Faculty of Law

DELIVERABLE REPORT

W1 – D1.1 – Workplan for Development of Common Open Data Language

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857592-TODO

Delivery Title:	**Workplan for Development of Common Open Data Language**
Delivery Number:	**D1.1.**
Lead beneficiary:	**Faculty of Law (LAW)**
Type:	**Report**
Work Package Title:	**Management and Coordination**
Work Package Number:	**WP1**
Dissemination level:	**Public**
Due Date:	**December 31, 2019**

STAKEHOLDERS INCLUDED: Faculty of Law (LAW)

INPUT DOCUMENTS: websites, journals and books, as well as publicly available documents, such as handbooks, lists, legislation, etc.

OUTPUT DOCUMENTS/MATERIALS: Plan/Report

Approved by: Dražen Tutić, WP1 Leader

Date of approval: December 31, 2019

Contents

1. Introduction

This work plan for the development of open data language is a result of the Twinning Open Data Operational project[1] Work Package 1 Management and Coordination, deliverable *D1.1 Workplan for the Development of Open Data Language.* This deliverable includes the initial glossary of common open data language and plan for its development, and will be the input for the whole project, including the requirements for the Open Data Training Programme and joint activities. The document will be published on the TODO project website http://todo-project.eu/

The document contains an initial glossary of open data related words and expressions which will be used in the course of the implementation of the TODO project (Annex II).

The second part of this deliverable is updated status on a survey of open data research at UNIZG. An initial survey done at the time of writing the project proposal is updated and extended.

The third part of this deliverable reports on kick-off meetings with its outcomes and directions that were discussed at this initial project meeting.

2. Workplan for development of open data language

2.1. Rationale

Developing common open data language for the TODO project is of major importance, it serves for setting the common culture, effective communication, sense of identity, consistency, and productivity. The open data ecosystem includes a wide range of stakeholders dealing with different domains. For the purpose of TODO, open data language has three major functions. First, to establish a common language between partners, second, to establish a common language of the project towards external stakeholders, and third to investigate and further refine open data language on an international level as well as to further develop open data language on the national level (Croatian language).

Common open data language should be used for the following project deliverables:

- D2.1 Learning Training Programme
- D2.2 Online Training Material
- D2.3 Summer School Training Material
- D4.4 Open Science Collaboration Platform
- D5.6 MOOC on Best Practices
- D6.2 Sustainable Cooperation Structure
- D6.4 Incorporation of Open Data in University Curricula

[1] https://cordis.europa.eu/project/rcn/224077/factsheet/en

2.2. Formalization of open data language

Open data language will be defined as a glossary of terms and definitions. This deliverable gives the initial glossary. The selection of terms and expressions relies on publicly accessible websites and guides on open data, open data policies and relevant EU legislation. The glossary also contains the list of key web sources.

The items are listed in alphabetical order with each item being referred to the source with the most used sources presented with abbreviations in the following way:

EUOD – EU Publications Office, EU Open Data; The Basics for EU Open Data Providers

ODH - Open Knowledge Foundation, Open Data Handbook

USOD - US WH Open Data Project

OD Directive - EU Open Data Directive 2019

This initial glossary will be transformed into a collaborative online platform (wiki) on the TODO Open Science Collaboration Platform (M12) which will facilitate its maintenance, development and public access.

2.3. Updating of glossary

The glossary of common open data language will be updated on a regular basis, with periodical surveys conducted biannually (M3, M9, M15, M21, M27, and M33) among all partners on the need on an adaptation of the existing terms and the inclusion of the new terms.

Translation to Croatian will be done in M24. This translation will be presented to national stakeholders ensuring project relevance and impact on the national level.

2.4. Dissemination of open data language

Terminology is one of the key components of any discipline and domain. Open data terminology is no exception, and for that reason is the subject of continuous development. Creating standardized domain language is a demanding task that requires special research activities. For that reason, the Consortium will undertake an effort to publish journal articles (estimation is at least 2 articles) and conference presentations (estimation is at least 4 presentations) on the topic of open data terminology. Further, through the TODO Open Science Collaboration Platform, we will provide public access and feedback to the created body of open data language.

3. Status of open data research at UNIZG

While open data research in Croatia is advancing and promising, it is not sufficiently mature to support or to steer the open data agenda. The results of the pre-project SWOT analysis (Table 1) of open data research at the UNIZG resulted with TODO agenda which aims to improve this situation by establishing a sustainable and internationally leading open data research environment capable of addressing key challenges in open data research through the development and application of an **interdisciplinary and multi-domain research approach** on open data.

Table 1. SWOT analysis of open data research at UNIZG

Strengths	• Involvement of researchers from many different disciplines related to open data agenda: law, organization studies, geomatics, ICT, transport, etcetera • Long tradition and international recognition in many scientific and artistic fields • Significant experience of national and international scientific projects • A large number of researchers in scientific and teaching positions • Existence of internationally recognized researchers and research groups • Researchers cover all scientific and artistic fields • Well-developed ICT infrastructure in most of the faculties
Weaknesses	• A large number of relatively small and fragmented research units • Limited international collaboration with and connection to the international open data research community • Different levels of achievements and excellence at faculties regarding open data • Low level of cooperation between faculties regarding open data • A low number of publications on open data in international peer-reviewed journals • Limited awareness of the value of open research data • The departure of young and prospective researchers to international institutions, limiting the possibility to open new research topics
Opportunities	• Implementation of open data agenda in Croatia requires support from research • Strong desire for education among open data professionals • New generations of researchers and practitioners require updated training and education curricula • National, regional and international universities interested in cooperation with open data projects • Enterprises interested in cooperation with the university on technological projects • Raising attention in the academic community to open data developments in the Balkan (and similar countries and regions).
Threats	• Lack of national research funding in Croatia • Scientific open data community builds on established research institutions and networks • Low attractiveness of research positions compared to enterprise and international institutions participating in open data agenda • Rather low interest among enterprises to collaborate on open data projects

The SWOT analysis was used to identify the current position of open data research in UNIZG. Lack of strategic specialization and fragmentation into a large number of small, almost one-member teams, are among the weakest points of the scientific structure at UNIZG.

A wish of TODO is not only to create user uptake of open data in Croatia, which certainly is one of the wider impacts, but it is more focused to **create a research environment that will inspire the existing body of knowledge, and explore novel ways for researching open data societal and scientific challenges**.

3.1. Open data research infrastructure at UNIZG

3.1.1. Organization of the University of Zagreb

The University of Zagreb (UNIZG) (http://www.unizg.hr/) established in 1669, is one of the oldest and largest universities in South-Eastern Europe. It consists of 34 faculties and academies and the Centre for Croatian Studies. With its comprehensive programs and over 72,000 students, the UNIZG is the strongest teaching institution in Croatia. At the UNIZG 191 undergraduate, 174 graduate, 72 doctoral (PhD) and 165 specialist postgraduate programs are offered. Students at the UNIZG (40% female) represent 48,5% of all students in Croatia. About 11.500 students enroll in the 1st year of study. Annually, some 5.500 students reach BA/BSc degrees, 2.300 MA/MSc degrees, and 540 PhD degrees. Teaching staff counts 7.915 persons, producing 51% of the total yearly scientific productivity in Croatia.

3.1.2. The University Computing Centre (SRCE) of the University of Zagreb

The University Computing Centre (SRCE) of the University of Zagreb has a long tradition in the area of information and communication technologies. It was founded in 1971 within the University of Zagreb, the only Croatian university at the time, with the purpose to enhance the implementation of information technologies in the academic community as well as in Croatia in general.

Today, SRCE is the main computing center and the architect of the e-infrastructure, covering both the University of Zagreb and the whole research and high education system. Furthermore, SRCE is the competence center for information and communication technologies as well as the center for education and support in the area of ICT application.

On September 10, 2019, SRCE has become the Croatian national RDA node. The Research Data Alliance (RDA) is a global organization focused on developing the social and technical conditions for open sharing and re-use of data and for fostering data-driven innovation. In 2013, RDAs were jointly launched by the European Commission (EC), the National Science Foundation (NSF), and the US Government's National Institute for Standards and Technology (NIST) and the Australian Government's Ministry of Innovation.

The tasks of the national RDA node are (https://www.srce.unizg.hr/):

- support research data at the national level by promoting the RDA vision and adopting RDA recommendations and results

- act as a liaison between national researchers, the global RDA community, and national science funding bodies
- contribute to defining European policies related to research data
- contribute to defining the RDA strategy and participating in RDA processes.

SRCE established DABAR (Digital Academic Archives and Repositories) (https://dabar.srce.hr/en) as the key component of the data layer of Croatian e-infrastructure, a system that enables all higher education and/or research institutions to easily establish and maintain reliable and interoperable institutional repositories. It also enables the research community to establish thematic repositories and archives. Thus, DABAR is also an ICT infrastructure that can be used for open research data.

SRCE established Hrčak (https://hrcak.srce.hr/?lang=en) as a central portal that gathers the **Croatian scientific and professional journals** which offer **open access** to their works (or at least to bibliographic information and abstracts of their papers). Hrčak has been validated and registered as a data source on the OpenAIRE portal, making articles published on Hrčak also available on that portal.

SRCE was hosting Open Data Day events in Croatia and is a partner on project EOSC-hub (https://www.eosc-portal.eu/) and together with the consortium of 100 partners from more than 50 countries will develop the vision of the Hub as the integration and management system of the European Open Science Cloud.

3.1.3. Current Open Data and Open Science Projects at the University of Zagreb

Besides the TODO project, there is currently only one project at which institutions of the University of Zagreb participates.

The National Initiatives for Open Science in Europe (NI4OS-Europe) project began on September 1, 2019, as part of the European Union's Horizon 2020 (Horizon 2020) Research and Innovation Program 2014-2020. It is part of an effort to establish the EOSC (European Open Science Cloud), in which SRCE participates as one of 22 consortium partners led by Ethniko Diktyo Erevnas Technologias AE (GRNET) from Greece.

The aim of the project is to support the coordination, convergence, and federation of national and / or thematic initiatives for open EOSC research data and services by developing appropriate common tools and mechanisms. On the one hand, the objectives are to support the gradual alignment of policies and practices of EOSC-relevant national and / or thematic initiatives with EOSC standards and, on the other, to allow open access to non-commercial services through the EOSC portal.

3.1.4. Conferences

Pubmet (http://pubmet.unizd.hr/pubmet2019/) annual conference organized by the University of Zadar since 2014 strives to present and discuss the plurality of approaches to the scholarly publishing and assessment (metrics) and new trends in the scholarly publishing area.

Since 2016, co-organizer of the conference is the University of Zagreb and since 2018, co-organization is done by the Faculty of Food Technology and Biotechnology of the University of Zagreb.

Open Data Day Croatia (https://odd.codeforcroatia.org/about/) is an annual conference organized by Code for Croatia, a non-government initiative. The participation of the University of Zagreb at that conference is low and not recognized. In previous years, hosting that was provided by SRCE of the University of Zagreb.

3.1.5. Publications in high impact journals

In D1.4. Evolution of the publications in high impact journals in the relevant research fields survey of publications from October 2016 to September 2019 published by researchers at 6 faculties of UNIZG that are partners in TODO was done. Publications are categorized into single discipline vs. interdisciplinary and to open data development vs. applications of open data. Thay way, it will be possible to measure the impact of the project on raised interdisciplinarity during project time, which is an innovative approach in this project. Most of the selected publications were written by researchers from the same faculty, or with international collaboration. This indicates that the collaboration of researchers on open dana topics from different faculties of UNIZG is non-existent or very low.

UNIZG Faculty	Single discpline		Interdisciplinary	
	Open data development	Applications of open data	Open data development	Applications of open data
GEOD	2	9	2	5
FER	2	2	1	8
FOI	0	5	3	2
LAW	1	1	1	1
TRANS	1	2	1	1
AGRI	1	10	0	6

A complete list of selected publications is given in **Annex I**.

A similar analysis is planned for all faculties of the University of Zagreb (M15), but in order to fulfill this task, Consortium will try to involve researchers from other disciplines and faculties by inviting them to join project activities and provide feedback on open data research in their disciplines. An alternative way, which also will be tested, is to search and analyze the Croatian Scientific Bibliography (CROSBI) which in the moment of writing this deliverable (Dec 2019) counts 201190 journal articles, 103257 theses, and 97071 conference proceeding papers.

4. Kick-off meeting

On November 7-8, 2019, the TODO project team met in Croatia on the Faculty of Geodesy (Zagreb) and held the kick-off meeting. In two days a series of constructive and productive meetings were held and set guidelines for further project activities in the coming months.

Final agenda of kick-off meeting, Zagreb, 7-8 November 2019

Nov 7, 2019 - Internal project meeting - Faculty of Geodesy, Kačićeva 26, Zagreb
Vijećnica AGG, ground floor

13:30 - 14:00	*Registration*
14:00 - 14:10	Welcome - **D. Medak** (Vice-dean for Science and Int. Collaboration, GEOD)
14:10 - 14:30	Project structure and organization - **D. Tutić**
14:30 - 15:00	WP2 Capacity Building - **F. Welle Donker & TG 2.x Leaders**
15:00 - 15:30	WP3 ESRs Training and Research - **I. Bosnić & TG 3.x Leaders**
15:30 - 16:00	*Group photo and refreshment*
16:00 - 16:30	WP4 Collaboration and Knowledge Sharing - **M. Vujić & TG 4.x Leaders**
16:30 - 17:00	WP5 Dissemination and Outreach - **N. Žajdela Hrustek & TG 5.x Leaders**
17:00 - 17:30	WP6 Sustainability - **C. Alexopoulos & TG 6.x Leaders**
17:30 - 18:00	WP1 Management and Coordination - **D. Tutić & TG 1.x Leaders**
18:00 - 18:30	Meeting ESRs - **I. Bosnić & ESRs**
19:00 -	*Social Dinner -* Pivnica Budweiser, Adžijina 16, Zagreb

Nov 8, 2019 - Open Participation - University of Zagreb, Trg Republike Hrvatske 14, Zagreb
Aula of Rectorate, ground floor, first on the left

09:30 - 10:00	*Registration*
10:00 - 10:30	Welcome - **M. Judaš** (UNIZG vice-rector), **A. Đapo** (GEOD dean), **T.J. Mlinarić** (TRANS dean), I**. Franić** (Assistant Minister, Ministry of Science and Education)
10:30 - 11:00	TODO - Project Presentation - **B. van Loenen**, **D. Tutić, C. Alexopoulos**
11:00 - 11:30	TODO - Expert Perspective - **D. Bevandić**
11:30 - 11:45	Open Data in the City of Rijeka - **T. Perše**
11:45 - 12:00	Open Data and Open Research - SRCE - I**. Marić**
12:00 - 12:15	Open Data Portals - Omega Software - **S. Gavrilov**
12:15 - 12:30	Open Data in State Geodetic Administration - **S. Zekušić**
12:30 - 12:45	SPIDER OpenSDI Erasmus+ project - **U. Klein**
12:45 - 13:00	*Closing and group photo*
13:00 - 14:00	*Coffee and refreshement*

First day – internal project meeting

Second day – open participation and outreach activities

Annex I – List of publications in high impact journals of UNIZG partners from October 2016 to September 2019

Faculty of Geodesy (GEOD)

Luka Rumora, Mario Miler & Damir Medak (2019) Contemporary comparative assessment of atmospheric correction influence on radiometric indices between Sentinel-2A and Landsat 8 imagery, Geocarto International, DOI: 10.1080/10106049.2019.1590465, SINGLE DISCIPLINE, APPLICATION

C., Verbeke, Mays, M. L., Temmer, M., Bingham, S., Steenburgh, R., Dumbović, M., et al. (2019). Benchmarking CME arrival time and impact: Progress on metadata, metrics, and events. *Space Weather*, 17, 6– 26. https://doi.org/10.1029/2018SW002046, SINGLE DISCIPLINE, APPLICATION

Mateo Gašparović, Luka Rumora, Mario Miler, and Damir Medak "Effect of fusing Sentinel-2 and WorldView-4 imagery on the various vegetation indices," *Journal of Applied Remote Sensing* 13(3), 036503 (30 July 2019). https://doi.org/10.1117/1.JRS.13.036503, SINGLE DISCIPLINE, APPLICATION

Pilaš, Ivan; Gašparović, Mateo; Đodan, Martina; Balenović, Ivan; Dugački, Ivica: Mogućnosti korištenja optičkih satelitskih snimaka srednje i visoke rezolucije (Landsat 8, RapidEye) u vizualizaciji i detekciji promjena šumskog pokrova nakon vjetroizvala // *Geodetski list,* 73 (96) (2019), 3; 261-276 (recenziran, članak, znanstveni) (https://www.bib.irb.hr/1014564), SINGLE DISCIPLINE, APPLICATION

Brkić, M. (2019) Monitoring geomagnetic information in the territory of Croatia, GEOFIZIKA VOL. 36 2019, DOI: 10.15233/gfz.2019.36.3 (http://dx.doi.org/10.15233/gfz.2019.36.3), SINGLE DISCIPLINE, APPLICATION

Gašparović, I.; Gašparović, M. Determining Optimal Solar Power Plant Locations Based on Remote Sensing and GIS Methods: A Case Study from Croatia. *Remote Sens.* 2019, *11*, 1481. (http://dx.doi.org/10.3390/rs11121481), SINGLE DISCIPLINE, APPLICATION

Jogun, T., Lukić, A. i Gašparović, M. (2019). Simulation model of land cover changes in a post-socialist peripheral rural area: Požega-Slavonia County, Croatia. *Hrvatski geografski glasnik, 81* (1), 31-59. https://doi.org/10.21861/HGG.2019.81.01.02 , SINGLE DISCIPLINE, APPLICATION

Gašparović, Iva; Gašparović, Mateo; Medak, Damir; Zrinjski, Mladen. Analiza prostornih podataka o solarnom potencijalu za Hrvatsku // *Geodetski list,* 73 (96) (2019), 1; 25-44 (međunarodna recenzija, članak, znanstveni) (https://www.bib.irb.hr/995829), SINGLE DISCIPLINE, APPLICATION

Slaven Marasović, Joep Crompvoets & Vesna Poslončec-Petrić (2019) State and development of local spatial data infrastructures in Croatia, Journal of Spatial Science, 64:3, 405-422, DOI: 10.1080/14498596.2018.1429331, INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

V. Flego, M. Roić. Land tenure registration on the marine areas in Croatia, Ocean & Coastal Management, Volume 166, 2018, Pages 72-81, ISSN 0964-5691, https://doi.org/10.1016/j.ocecoaman.2018.03.008. (http://www.sciencedirect.com/science/article/pii/S0964569117305847), SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

Tutić, D.; Štanfel, M.; Horvat, M.T. Multi-Criteria Land Evaluation of Suitability for the Sport of Foot Orienteering: A Case Study of Croatia and Slovenia. *ISPRS Int. J. Geo-Inf.* 2018, *7*, 227. (http://dx.doi.org/10.3390/ijgi7060227), INTERDISCIPLINARY, APPLICATION

Iva Gašparović, Mateo Gašparović, Damir Medak: Determining and analysing solar irradiation based on freely available data: A case study from Croatia, Environmental Development, Volume 26, 2018, Pages 55-67, ISSN 2211-4645, https://doi.org/10.1016/j.envdev.2018.04.001., INTERDISCIPLINARY, APPLICATION

Tomić, H.; Mastelić Ivić, S.; Roić, M. Land Consolidation Suitability Ranking of Cadastral Municipalities: Information-Based Decision-Making Using Multi-Criteria Analyses of Official Registers' Data. *ISPRS Int. J. Geo-Inf.* 2018, *7*, 87. (http://dx.doi.org/10.3390/ijgi7030087), INTERDISCIPLINARY, APPLICATION

Tavra, M., Duplančić Leder, T. i Cetl, V. (2018). Stakeholders Needs Requisite Analysis: Towards Croatian Marine Spatial Data Infrastructure Establishment. *Tehnički vjesnik, 25* (Supplement 1), 176-182. https://doi.org/10.17559/TV-20160607222834, INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Jovanović, Nataša; Župan, Robert: Analiza stanja vegetacije prije i nakon šumskih požara pomoću satelitskih snimaka Sentinel-2 na području Dalmacije *// Geodetski list : glasilo Hrvatskoga geodetskog društva,* 71 (94) (2017), 3; 233-248 (https://www.bib.irb.hr/909090), SINGLE DISCIPLINE, APPLICATION

Abramic, A.; Kotsev, A.; Cetl, V.; Kephalopoulos, S.; Paviotti, M. A Spatial Data Infrastructure for Environmental Noise Data in Europe. *Int. J. Environ. Res. Public Health* 2017, *14*, 726. (http://dx.doi.org/10.3390/ijerph14070726), SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

Dražen Tutić, Tomislav Jogun, Ana Kuveždić Divjak & Martina Triplat Horvat (2017): World political map from OpenStreetMap data, Journal of Maps, 13:1, 67-73, DOI: 10.1080/17445647.2017.1323683, INTERDISCIPLINARY, APPLICATION

Faculty of Electrical Engineering and Computing (FER)

M. Piškorec, T. Šmuc and M. Šikić: "Disentangling Sources of Influence in Online Social Networks," in *IEEE Access*, vol. 7, pp. 131692-131704, 2019. doi: 10.1109/ACCESS.2019.2940762, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8835042&isnumber=8600701, INTERDISCIPLINARY, APPLICATION

Marina Ptiček, Boris Vrdoljak & Marko Gulić (2019): The potential of semantic paradigm in warehousing of big data, Automatika, 60:4, 393-403, DOI: 10.1080/00051144.2019.1630582 , INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Petar Zečević, Colin T. Slater, Mario Jurić, Andrew J. Connolly, Sven Lončarić, Eric C. Bellm, V. Zach Golkhou, and Krzysztof Suberlak AXS: A Framework for Fast Astronomical Data Processing Based on Apache Spark, The Astronomical Journal, Volume 158, Number 1, (http://dx.doi.org/10.3847/1538-3881/ab2384), INTERDISCIPLINARY, APPLICATION

Goran Rumin and Igor Mekterović LOD Construction Through Supervised Web Relation Extraction and Crowd Validation, Journal of Web Engineering, Vol: 18, Issue: Combined Issue 1, 2 & 3, January 2019, Article No: 7, Page: 229-256, https://doi.org/10.13052/jwe1540-9589.18137 , SINGLE DISCIPLINE, APPLICATION

L. Humski, D. Pintar and M. Vranić, "Analysis of Facebook Interaction as Basis for Synthetic Expanded Social Graph Generation," in *IEEE Access*, vol. 7, pp. 6622-6636, 2019. doi: 10.1109/ACCESS.2018.2886468, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8573573&isnumber=8600701, SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

Sabolić, Dubravko Can decarbonization policy results be detected by simplistic analysis of macro-level statistical data? *// Technology in society,* 53 (2018), 5; 103-109 (https://www.bib.irb.hr/954765), INTERDISCIPLINARY, APPLICATION

Pevec, D, Babic, J, Kayser, MA, Carvalho, A, Ghiassi-Farrokhfal, Y, Podobnik, V. A data-driven statistical approach for extending electric vehicle charging infrastructure. *Int J Energy Res*. 2018; 42: 3102– 3120. https://doi.org/10.1002/er.3978, INTERDISCIPLINARY, APPLICATION

Piškorec M., Antulov-Fantulin N., Miholić I., Šmuc T., Šikić M. (2018) Modeling Peer and External Influence in Online Social Networks: Case of 2013 Referendum in Croatia. In: Cherifi C., Cherifi H., Karsai M., Musolesi M. (eds) Complex Networks & Their Applications VI. COMPLEX NETWORKS 2017. Studies in Computational Intelligence, vol 689. Springer, Cham First Online 27 November 2017, DOI https://doi.org/10.1007/978-3-319-72150-7_82, Publisher Name Springer, Cham Print ISBN 978-3-319-72149-1, Online ISBN 978-3-319-72150-7, eBook Packages Engineering, INTERDISCIPLINARY, APPLICATION

Bagić Babac, M. and Podobnik, V. (2018), "What social media activities reveal about election results? The use of Facebook during the 2015 general election campaign in Croatia", *Information Technology & People*, Vol. 31 No. 2, pp. 327-347. https://doi.org/10.1108/ITP-08-2016-0200, INTERDISCIPLINARY, APPLICATION

I. P. Žarko *et al*., "The symbIoTe Solution for Semantic and Syntactic Interoperability of Cloud-based IoT Platforms," *2019 Global IoT Summit (GIoTS)*, Aarhus, Denmark, 2019, pp. 1-6. doi: 10.1109/GIOTS.2019.8766420, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8766420&isnumber=8766345, SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

G. Antolić and L. Brkić, "Recommender system based on the analysis of publicly available data," *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2017, pp. 1379-1384. doi: 10.23919/MIPRO.2017.7973637, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7973637&isnumber=7973374, INTERDISCIPLINARY, APPLICATION

M. Marjanović, S. Grubeša and I. P. Žarko, "Air and noise pollution monitoring in the city of Zagreb by using mobile crowdsensing," *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Split, 2017, pp. 1-5. doi: 10.23919/SOFTCOM.2017.8115502, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8115502&isnumber=8115491, INTERDISCIPLINARY, APPLICATION

Luka Humski , (Student Member, IEEE), Damir Pintar, (Member, IEEE), and Mihaela Vranić, (Member, IEEE) Analysis of Facebook Interaction as Basis for Synthetic Expanded Social Graph Generation, IEEE Access , date of publication December 12, 2018, date of current version January 23, 2019, Digital Object Identifier 10.1109/ACCESS.2018.2886468 (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8573573), INTERDISCIPLINARY, APPLICATION

Faculty of Organization and Informatics (FOI)

Maja Gligora Marković, Sandra Debeljak, Nikola Kadoić: Preparing Students for the Era of the General Data Protection Regulation (GDPR) TEM Journal. Volume 8, Issue 1, Pages 150-156, ISSN 2217-8309, DOI: 10.18421/TEM81-21, February 2019. (http://dx.doi.org/10.18421/TEM81-21), INTERDISCIPLINARY, APPLICATION

Andročec, Darko; Vrček, Neven: Ontology-Based Resolution of Cloud Data Lock-in Problem *// Computing and informatics,* **37** (2018), 5; 1231-1257 doi:10.4149/cai 2018 5 1231, (https://www.bib.irb.hr/970655), INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Andročec, Darko; Novak, Matija; Oreški, Dijana: Using Semantic Web for Internet of Things Interoperability: A Systematic Review *// International journal on semantic web and information systems,* 14 (2018), 4; 147-171 doi:10.4018/ijswis.2018100108, (https://www.bib.irb.hr/949959), INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Darko Andročec: Analysis of Sci-Hub downloads of computer science papers, Acta Universitatis Sapientiae, Informatica, Volume 9: Issue 1, Aug 2017, 83–96, (https://doi.org/10.1515/ausi-2017-0006), DISCIPLINARY, APPLICATION

Markus Schatten, Bogdan Okreša Đurić: Social Networks in „The Mana World“ – an Analysis of Social Ties in an Open Source MMORPG, International Journal of Multimedia and Ubiquitous Engineering, Vol. 11, No. 3 (2016), pp 257-272, (http://dx.doi.org/10.14257/ijmue.2016.11.3.25), DISCIPLINARY, APPLICATION

Markus Schatten, Jurica Ševa, Igor Tomičić: A roadmap for scalable agent organizations in the Internet of Everything, Journal of Systems and Software, Volume 115, 2016, Pages 31-41, ISSN 0164-212, https://doi.org/10.1016/j.jss.2016.01.022. (http://www.sciencedirect.com/science/article/pii/S0164121216000170), INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Dina Korent, Ksenija Vuković & Ruža Brčić (2015): Entrepreneurial activity and regional development, Economic Research-Ekonomska Istraživanja, 28:1, 939-958, DOI: 10.1080/1331677X.2015.1084237, DISCIPLINARY, APPLICATION

Dijana Oreski, Stjepan Oreski, Bozidar Klicek: Effects of dataset characteristics on the performance of feature selection techniques, Applied Soft Computing, Volume 52, 2017, Pages 109-119, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2016.12.023. (http://www.sciencedirect.com/science/article/pii/S156849461630641X), INTERDISCIPLINARY, APPLICATION

Maršić, Kristina; Oreški, Dijana: Estimation and comparison of underground economy in Croatia and European Union countries: Fuzzy logic approach *// Journal of information and organizational sciences,* 40 (2016), 1; 83-104, (https://www.bib.irb.hr/824487), DISCIPLINARY, APPLICATION

Faculty of Law (LAW)

Romana Matanovac Vučković, Ivana Kanceljak: Does the Right to Use Digital Content Affect Our Digital Inheritance?, EU and comparative law issues and challenges series, Vol 3 (2019), (http://dx.doi.org/10.25234/eclic/9029), SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

Manojlović Toman, Romea; Lalić Novak, Goranka: The (Lack of) Demand for Performance Information by the Croatian Parliament. Transylvanian Review of Administrative Sciences, [S.l.], p. 100-115, Oct. 2019. ISSN 1842-2845. Available at: <http://rtsa.ro/tras/index.php/tras/article/view/608>. doi:http://dx.doi.org/10.24193/tras.58E.7., SINGLE DISCIPLINE, APPLICATION

Katulić, Tihomir; Katulić, Anita: GDPR and the Reuse of Personal Data in Scientific Research *// MIPRO 2018 41st International Convention Proceedings* / Skala, Karolj (ur.). Rijeka: MIPRO, 2018. str. 1514-1519 ,(https://www.bib.irb.hr/982445), INTERDISCIPLINARY, APPLICATION

N. Gumzej: "Law and technology in data processing: Risk-based approach in EU data protection law and implementation challenges in Croatia," *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2017, pp. 1424-1430. doi: 10.23919/MIPRO.2017.7973645, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7973645&isnumber=7973374, INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Faculty of Transport and Traffic Sciences (TRANS)

Koltovska Nečoska, D., Ivanjko, E. and Pavleski, D. (2018): “Creating Infrastructure for Urban Mobility: Case Study of Skopje”, *Promet - Traffic & Transportation*, 30(4), pp. 429-443. doi: 10.7307/ptt.v30i4.2675., (http://dx.doi.org/10.7307/ptt.v30i4.2675), SINGLE DISCIPLINE, APPLICATION

G. Vojković and M. Milenković, "GDPR in access control and time and attendance systems using biometric data," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2018, pp. 1138-1142. doi: 10.23919/MIPRO.2018.8400207, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8400207&isnumber=8399814, SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

G. Vojkovic, "Will the GDPR slow down development of smart cities?," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2018, pp. 1295-1297. doi: 10.23919/MIPRO.2018.8400234 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8400234&isnumber=8399814, INTERDISCIPLINARY, OPEN DATA DEVELOPMENT

Erdelić, Tomislav; Carić, Tonči; Erdelić, Martina; Tišljarić, Leo, Electric vehicle routing problem with single or multiple recharges // Transportation Research Procedia, 40 (2019), 217-224 doi:10.1016/j.trpro.2019.07.033, SINGLE DISCIPLINE, APPLICATION

Erdelić, Tomislav; Carić, Tonči, A: Survey on the Electric Vehicle Routing Problem: Variants and Solution Approaches // Journal of Advanced Transportation, 2019 (2019), 1-48 doi:10.1155/2019/5075671, SINGLE DISCIPLINE, APPLICATION

Semanjski, Ivana; Gautama, Sidharta; Ahas, Rein; Witlox, Frank: Spatial context mining approach for transport mode recognition from mobile sensed big data // Computers environment and urban systems, 66 (2017), 38-52 doi:10.1016/j.compenvurbsys.2017.07.004, INTERDISCIPLINARY, APPLICATION

Faculty of Agriculture (AGRI)

Andlar, G., Šrajer, F., Trojanović, A.: Discovering Cultural Landscape In Croatia: History And Classification Of Croatian Adriatic Enclosed Landscape, *ANNALES · Ser. hist. sociol.* · 28 · 2018 · 4, DOI 10.19233/ASHS.2018.46, (http://dx.doi.org/10.19233/ASHS.2018.46), INTERDISCIPLINARY, APPLICATION

Joy Michael K., Foote Kyleisha J., McNie Pierce, Piria Marina (2018): Decline in New Zealand's freshwater fish fauna: effect of land use. *Marine and Freshwater Research 70*, 114-124. (https://doi.org/10.1071/MF18028), SINGLE DISCIPLINE, APPLICATION

Ribic, B, Pezo, L, Sincic, D, Loncar, B, Voca, N.: Predictive model for municipal waste generation using artificial neural networks—Case study City of Zagreb, Croatia. *Int J Energy Res*. 2019; 43: 5701– 5713. (https://doi.org/10.1002/er.4632), INTERDISCIPLINARY, APPLICATION

Piria, Marina; Simonović, Predrag; Zanella, Davor; Ćaleta, Marko; Šprem, Nikica; Paunović, Momir; Tomljanović, Tea; Gavrilović, Ana; Pecina, Marija; Špelić, Ivan et al.: Long-term analysis of fish assemblage structure in the middle section of the Sava River – The impact of pollution, flood protection and dam construction *// Science of the total environment,* 651 (2018), 1; 143-153, (https://www.bib.irb.hr/971518), SINGLE DISCIPLINE, APPLICATION

Philippe Négrel, Benedetto De Vivo, Clemens Reimann, Anna Ladenberger, Domenico Cicchella, Stefano Albanese, Manfred Birke, Walter De Vos, Enrico Dinelli, Annamaria Lima, Patrick J. O'Connor, Ignace Salpeteur, Timo Tarvainen, M. Andersson, R. Baritz, M.J. Batista, A. Bel-lan, A. Demetriades, M. Ďuriš, A. Dusza-Dobek, O.A. Eggen, M. Eklund, V. Ernstsen, P. Filzmoser, D.M.A. Flight, S. Forrester, M. Fuchs, U. Fügedi, A. Gilucis, M. Gosar, V. Gregorauskiene, W. De Groot, A. Gulan, J. Halamić, E. Haslinger, P. Hayoz, R. Hoffmann, J. Hoogewerff, H. Hrvatovic, S. Husnjak, L. Janik, G. Jordan, M. Kaminari, J. Kirby, V. Klos, F. Krone, P. Kwecko, L. Kuti, J. Locutura, P. Lucivjansky, A. Mann, D. Mackovych, M. McLaughlin, B.I. Malyuk, R. Maquil, R.G. Meuli, G. Mol, K. Oorts,

R.T. Ottesen, A. Pasieczna, V. Petersell, S. Pfleiderer, M. Poňavič, C. Prazeres, U. Rauch, S. Radusinović, M. Sadeghi, R. Scanlon, A. Schedl, A. Scheib, I. Schoeters, E. Sellersjö, I. Slaninka, J.M. Soriano-Disla, A. Šorša, R. Srvkota, T. Stafilov, V. Trendavilov, P. Valera, V. Verougstraete, D. Vidojević, Z. Zomeni: U-Th signatures of agricultural soil at the European continental scale (GEMAS): Distribution, weathering patterns and processes controlling their concentrations, Science of The Total Environment, Volumes 622–623, 2018, Pages 1277-1293, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2017.12.005. (http://www.sciencedirect.com/science/article/pii/S0048969717334241), SINGLE DISCIPLINE, APPLICATION

Philippe Négrel, Anna Ladenberger, Clemens Reimann, Manfred Birke, Martiya Sadeghi: Distribution of Rb, Ga and Cs in agricultural land soils at European continental scale (GEMAS): Implications for weathering conditions and provenance, Chemical Geology, Volume 479, 2018, Pages 188-203, ISSN 0009-2541, https://doi.org/10.1016/j.chemgeo.2018.01.009. (http://www.sciencedirect.com/science/article/pii/S0009254118300214), SINGLE DISCIPLINE, APPLICATION

Jörg Matschullat, Clemens Reimann, Manfred Birke, Debora dos Santos Carvalho, S. Albanese, M. Anderson, R. Baritz, M.J. Batista, A. Bel-lan, D. Cicchella, A. Demetriades, B. De Vivo, W. De Vos, E. Dinelli, M. Ďuriš, A. Dusza-Dobek, O.A. Eggen, M. Eklund, V. Ernsten, K. Fabian, P. Filzmoser, D.M.A. Flight, S. Forrester, U. Fügedi, A. Gilucis, M. Gosar, V. Gregorauskiene, W. De Groot, A. Gulan, J. Halamić, E. Haslinger, P. Hayoz, J. Hoogewerff, H. Hrvatovic, S. Husnjak, F. Jähne-Klingberg, L. Janik, G. Jordan, M. Kaminari, J. Kirby, V. Klos, P. Kwećko, L. Kuti, A. Ladenberger, A. Lima, J. Locutura, P. Lucivjansky, A. Mann, D. Mackovych, M. McLaughlin, B.I. Malyuk, R. Maquil, R.G. Meuli, G. Mol, P. Négrel, P. O'Connor, K. Oorts, R.T. Ottesen, A. Pasnieczna, V. Petersell, S. Pfleiderer, M. Poňavič, C. Prazeres, S. Radusinović, U. Rauch, M. Sadeghi, I. Salpeteur, R. Scanlon, A. Schedl, A. Scheib, I. Schoeters, P. Šefčik, E. Sellersjö, I. Slaninka, J.M. Soriano-Disla, A. Šorša, R. Svrkota, T. Stafilov, T. Tarvainen, V. Tendavilov, P. Valera, V. Verougstraete, D. Vidojević, A. Zissimos, Z. Zomeni: GEMAS: CNS concentrations and C/N ratios in European agricultural soil, Science of The Total Environment, Volume 627, 2018, Pages 975-984, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2018.01.214. (http://www.sciencedirect.com/science/article/pii/S0048969718302559), SINGLE DISCIPLINE, APPLICATION

Igor Bogunovic, Sebastiano Trevisani, Paulo Pereira, Vesna Vukadinovic: Mapping soil organic matter in the Baranja region (Croatia): Geological and anthropic forcing parameters, Science of The Total Environment, Volume 643, 2018, Pages 335-345, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2018.06.193. (http://www.sciencedirect.com/science/article/pii/S0048969718322691), INTERDISCIPLINARY, APPLICATION

Šestak I., Mesić M., Zgorelec Ž., Perčin A., Stupnišek I. (2018): Visible and near infrared reflectance spectroscopy for field-scale assessment of Stagnosols properties. Plant Soil Environ., 64: 276-282. , (http://dx.doi.org/10.17221/220/2018-PSE), SINGLE DISCIPLINE, APPLICATION

Dragičević, Nevena; Karleuša, Barbara; Ožanić, Nevenka; Kisić, Ivica: Effect of Source-Varying Input Data on Erosion Potential Model Performance // *Geocarto International,* 34 (2018), 10; 1109-1122 doi:10.1080/10106049.2018.1474273, (https://www.bib.irb.hr/937238), SINGLE DISCIPLINE, OPEN DATA DEVELOPMENT

Bozzi, Riccardo; Škrlep, Martin; Lenoir, Herveline; Lebret, Benedicte; Garcia Casco, Juan Mari; Petig, Mathias; Charneca, Rui; Paixim, Hugo; Karolyi, Danijel; Radović, Čedomir et al:. Survey of demographic and phenotypic data of local pig breeds of TREASURE project // *Archivos de Zootecnia,* 68 (2018), Suplement 1; 1-4 doi:10.21071/az.v67iSupplement.3193, (https://www.bib.irb.hr/923535), INTERDISCIPLINARY, APPLICATION

Banadinović, Maja; Džidić Alen; Simčič, Mojca, Šalamon Dragica: Geographic Patterns of Genetic Variation in Indigenous Eastern Adriatic Sheep Breeds // *ACS. Agriculturae conspectus scintificus,* 82 (2017), 3; 281-285, (https://www.bib.irb.hr/936334), INTERDISCIPLINARY, APPLICATION

R. Kulmatov, M. Groll, A. Rasulov, I. Soliev, M. Romic: Status quo and present challenges of the sustainable use and management of water and land resources in Central Asian irrigation zones - The example of the Navoi region (Uzbekistan), Quaternary International, Volume 464, Part B, 2018, Pages 396-410, ISSN 1040-6182, https://doi.org/10.1016/ j.quaint. 2017.11.043. (http://www.sciencedirect.com/science/article/pii/S1040618216312095), SINGLE DISCIPLINE, APPLICATION

Biško, A., Savić, Z., Jukić, L., Leko, M., Radunić, M. and Miloloža, D. (2017). Sweet cherry cultivation in the Republic of Croatia. Acta Hortic. 1161, 185-192, DOI: 10.17660/ActaHortic.2017.1161.31 (https://doi.org/10.17660/ActaHortic.2017.1161.31), SINGLE DISCIPLINE, APPLICATION

Željka MESIĆ, Miroslav BOŽIĆ, Marija CERJAK: The impact of geographical indications on the competitiveness of traditional agri-food products, Journal of Central European Agriculture, 2017, 18(1), p. 1-14, DOI: 10.5513/JCEA01/18.1.1859, (http://dx.doi.org/DOI:%2010.5513/JCEA01/18.1.1859), INTERDISCIPLINARY, APPLICATION

Gotlin Čuljak, Tanja; Pernar, Renata; Juran, Ivan; Ančić, Mario; Bažok, Renata: Impact of oilseed rape crop management systems on the spatial distribution of Brassicogethes aeneus (Fabricius 1775): implications for Integrated pest management // *Crop protection,* 89 (2016), 129-138, (https://www.bib.irb.hr/803207), SINGLE DISCIPLINE, APPLICATION

Annex II - Glossary of open data language v1.0.

31 December 2019

The selection of terms and expressions relies on the publicly accessible websites and guides on open data, open data policies and relevant EU legislation. The glossary also contains the list of key web sources.

The items are listed in alphabetical order with each item being referred to the source with the most used sources presented with abbreviations in the following way:

EUOD – EU Publications Office, EU Open Data; The Basics for EU Open Data Providers

ODH - Open Knowledge Foundation, Open Data Handbook

USOD - US WH Open Data Project

OD Directive - EU Open Data Directive 2019

Glossary

A

Access to Information / Freedom of Information

A legal requirement for public bodies to provide data held by them to citizens on request as well as proactively, unless a specific exemption applies, e.g. the data is confidential for the reasons of national security, privacy, market competition or similar. Information obtained under access to information law is not automatically considered open data, unless it is delivered in a machine-readable format and under an open licence. In many of the EU countries the right of access to information (documents) is considered to be a constitutional rank right which is protected by the independent redress mechanism and/or courts. In Anglo-Saxon countries the term is freedom of information (USA, UK etc.).

ADMS (Asset Description Metadata Schema)

The asset description metadata schema is a vocabulary to describe interoperability assets making it possible for ICT developers to explore and search for interoperability assets. ADMS allows public administrations, businesses, standardisation bodies and academia to:

- describe semantic assets in a common way so that they can be seamlessly cross-queried and discovered by information and communications technology (ICT) developers from a single access point;
- search, identify, retrieve and compare semantic assets to be reused, avoiding duplication and expensive design work through a single point of access;
- keep their own system for documenting and storing semantic assets;
- improve indexing and visibility of their own assets;
- link semantic assets to one another in cross-border and cross-sector settings.

Source: https://joinup.ec.europa.eu/asset/adms/description, cited by EU OD.

Anonymisation

Processing data that includes personal information so that individuals can no longer be identified in the resulting data. Anonymisation enables data to be published without breaching data protection principles. The principal techniques are aggregation and de-identification. Care must be taken to avoid data leakage that would result in individuals' privacy being compromised. *Source: ODH.*

Under EU rules anonymisation means the process of changing documents into anonymous documents which do not relate to an identified or identifiable natural person, or the process of rendering personal data anonymous in such a manner that the data subject is not or no longer identifiable. *Source: OD Directive*

API (Application Programming Interface)

An application programming interface, which is a set of definitions of the ways one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between higher-level and lower-level software. *Source: USOD*

A way computer programmes talk to one another. Can be understood in terms of how a programmer sends instructions between programmes. For data, this is usually a way provided by the data publisher for programs or apps to read data directly over the web. The app sends the API a query asking for the specific data it needs, e.g. the time of the next bus leaving a particular stop. This allows the app to use the data without downloading the whole dataset, saving bandwidth and ensuring that the data used is the most up-to-date available. *Source: ODH.*

API Analytics

Rate limiting will be part of any API platform, without some sort of usage log and analytics showing developers where they stand, the rate limits will cause nothing but frustration. Clearly show developers where they are at with daily, weekly or monthly API usage and provide proper relief valves allowing them to scale their usage properly. *Source: US OD*

API Documentation

Quality API documentation is the gateway to a successful API. API documentation needs to be complete, yet simple–a very difficult balance to achieve. This balance takes work and will take the work of more than one individual on an API development team to make happen.

API documentation can be written by developers of the API, but additional edits should be made by developers who were not responsible for deploying the API. As a developer, it's easy to overlook parameters and other details that developers have made assumptions about. *Source: US OD*

App / Application

A piece of software (short for 'application'), especially one designed to run on the web or on mobile phones and similar platforms. Apps can make network connections to large databases and thus be a powerful way of consuming open data, which may be real-time, personalised, and (using a mobile phone's GPS) location-specific information. Crowdsourcing apps can also be used to build or improve datasets. *Source: ODH.*

Application Library

Complete, functioning applications built on an API is the end goal of any API owner. Make sure and showcase all applications that are built on an API using an application showcase or directory. App showcases are a great way to showcase not just applications built by the API owner, but also showcase the successful integrations of ecosystem partners and individual developers. *Source: US OD*

Application Profile

A specification that re-uses terms from one or more base standards, adding more specificity by identifying mandatory, recommended and optional elements to be used for a particular application, as well as recommendations for controlled vocabularies to be used. *Source: https://data.europa.eu/euodp/en/developerscorner*

Asset list

Contains information on all databases public body holds and maintains with key information (metadata), including published and unpublished datasets. It is an obligation of the public sector bodies to publish asset lists on their websites. See also → *Information Asset Register*.

Attribution

Acknowledging the source of data when using or re-publishing it. A data licence permitting the data to be used may include a requirement to attribute the source. Data subject to this restriction may still be considered open data according to the Open Definition. *Source: ODH.*

Attribution Licence (CC BY)

A Creative Commons Licence that lets others distribute, remix, adapt, and build upon data, even commercially, as long as they credit the source for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials. *Source: Creative Commons*

Attribution-ShareAlike (CC BY-SA)

A Creative Commons Licence that lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit the author and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on the source will carry the same license, so any derivatives will also allow commercial use. For example, this is the license used by Wikipedia. *Source: Creative Commons*

Attribution-NoDerivs (CC BY-ND)

A Creative Commons Licence that lets others reuse the work for any purpose, including commercially; however, it cannot be shared with others in adapted form, and credit must be provided to the author. *Source: Creative Commons.*

Attribution-NonCommerical (CC BY-NC)

A Creative Commons Licence that lets others remix, adapt, and build upon the data non-commercially, and although their new works must also acknowledge the source and be non-commercial, they don't have to license their derivative works on the same terms. *Source: Creative Commons.*

Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)

A Creative Commons Licence that lets others remix, adapt, and build upon data non-commercially, as long as they credit the source and license their new creations under the identical terms. *Source: Creative Commons.*

Attribution-NonCommercial- NoDerivs (CC BY-NC-ND)

A Creative Commons Licence that is the most restrictive of the CC six main licenses, only allowing others to download data and share them with others as long as they credit the source, but they can't change them in any way or use them commercially. *Source: Creative Commons.*

B

Bandwidth

The rate at which data can be transferred between computers. As bandwidth is limited, apps aim to download only the minimum amount of data needed to fulfil a user's request. *Source: ODH.*

Big Data

A collection of data so large that it cannot be stored, transmitted or processed by traditional means. The increasing availability of and need to process such datasets (for example, huge collections of weather or other scientific data) has led to the development of specialised computer technologies, architectures and programming languages. *Source: ODH.*

Bit Torrent

BitTorrent is a protocol for distributing the bandwidth for transferring very large files between the computers which are participating in the transfer. Rather than downloading a file from a specific source, BitTorrent allows peers to download from each other. *Source: ODH.*

Bulk / Bulk download

A download containing files from multiple collections that can be retrieved at once. Data is available in bulk if the entire dataset can be downloaded easily and efficiently to a user's own system. Conversely it is non-bulk if one is limited to getting small parts of the dataset, for example, are you restricted to a few elements of the data at a time and therefore require thousands or millions of requests to get the entire dataset. The provision of bulk access is a requirement of open data. *Source: ODH.*

C

Catalog

A catalog is a collection of datasets or web services. *Source: US OD*

Civic Hacking

Building tools and communities, usually online, that address particular civic or social problems. Examples could be tools that help users meet like-minded people locally based on particular interests, report broken infrastructure to their local council, or collaborate to clear litter from their neighbourhood. Local-level open data is particularly useful for civic hacking projects. *Source: ODH*

CKAN

An open-source software platform for creating data portals, built and maintained by Open Knowledge. CKAN is used as the official data-publishing platform of around 20 national governments and powers many more local, community, scientific and other data portals. Notable features are configurable metadata, user-friendly web interface for publishers and data users, data preview, organisation-based authorisation levels, and APIs giving access to all features as well as data access. *Source: ODH*

A data management system that makes data accessible by providing tools to streamline publishing, sharing, finding and using data. CKAN is aimed at data publishers (national and regional governments, companies and organisations) working to make their data open and available. *Source: http://ckan.org/*

Cloud

Data stored 'in the cloud' is handled by a hosting company, relieving the data owner of the need to manage its physical storage. Instead of being stored on a single machine, it may be stored across or moved between multiple machines in different locations, but the data owner and users do not need to know the details. The hosting company is responsible for keeping it available and accessible via the internet. *Source: ODH*

Code Library

Working code samples in all the top programming languages are common place in the most successful APIs. Documentation will describe in a general way, how to use an API, but code samples will speak in the specific language of developers. *Source: US OD*

Connectivity

Connectivity relates to the ability for communities to connect to the Internet, especially the World Wide Web. *Source: ODH*

Content API

A web service that provides dynamic access to the page content of a website, includes the title, body, and body elements of individual pages. Such an API often but not always functions atop a Content Management System. *Source: US OD*

Conversion

The process of automatically reading data in one file format and emitting the same data in a different format, thus making the data accessible to a wider range of applications. *Source: ODH*

Copyright

A legal right over intellectual property (e.g. a book) belonging to the creator of the work. While individual data (facts) cannot be copyright, a database will in general be covered by copyright protecting the selection and arrangement of data within it. Within the European Union separate 'database rights' protect a database where there was a substantial effort in 'obtaining' the data. A copyright holder may use a licence to grant other people rights in the protected material, perhaps subject to specified restrictions. *Source: ODH*

CORDIS

The European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their results. Source: http://cordis.europa.eu/home_en.html

Cost recovery

The principle of setting a price for a resource, e.g. data, aiming to recover the cost of collecting the data, as distinct from marginal cost. In the EU it is allowed only exceptionally, as determined by the PSI/OD Directive.

Creative Commons

A non-profit organisation founded in 2001 that promotes re-usable content by publishing a number of standard licences, some of them open (though others include a non-commercial clause), that can be used to release content for re-use, together with clear explanations of their meaning. Website: https://creativecommons.org/ *Source: ODH*

Creative Commons Licences

A set of open standard licences determined by the Creative Commons organisation. See → https://creativecommons.org/licenses/

Crowdsourcing

A model in which individuals and organisations obtain goods and services (ideas, money) from a large relatively open and often rapidly evolving group of internet users. It divides work between participants to achieve cumulative result. The term was coined 2006 (crowd plus sourcing), although it existed as a model before digital age. *Source: https://en.wikipedia.org/wiki/Crowdsourcing*

Dividing the work of collecting a substantial amount of data into small tasks that can be undertaken by volunteers. *Source: ODH*

CSV

'Comma-separated values' is a standard format for spreadsheet data. Data is represented in a plain text file, with each data row on a new line and commas separating the values on each row. As a very simple open format it is easy to consume and is widely used for publishing open data. *Source: ODH*

It is often used to exchange data between differently similar applications. The CSV file format is useable by KSpread, OpenOffice Calc and Microsoft Excel spreadsheet applications. Many other applications support CSV to import or export data. *Source: http://edoceo.com/utilitas/csv-file-format, cited by EU OD*

It is a computer data file used for implementing the tried and true organizational tool, the Comma Separated List. The CSV file is used for the digital storage of data structured in a table of lists form. Each line in the CSV file corresponds to a row in the table. Within a line, fields are separated by commas, and each field belongs to one table column. CSV files are often used for moving tabular data between two different computer programs (like moving between a database program and a spreadsheet program). *Source: US OD*

CSW

Catalog Service for the Web (CSW) is an API used by geospatial systems to provide metadata in open standards, including in the FGDC-endorsed ISO 19115 schema. The CSW-provided metadata can be mapped into the Project Open Data metadata schema. *Source: US OD*

D

Data

Data may be thought of as unprocessed atomic statements of fact. It very often refers to systematic collections of numerical information in tables of numbers such as spreadsheets or databases. When data is structured and presented so as to be useful and relevant for a particular purpose, it becomes information available for human apprehension. *Source: ODH*

A value or set of values representing a specific concept or concepts. Data become “information” when analysed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary depending on its context. Data includes *all* data. It includes, but is not limited to, 1) geospatial data 2) unstructured data, 3) structured data, etc. *Source: US OD*

Data Access Protocol

A system that allows outsiders to be granted access to databases without overloading either system. *Source: ODH*

Data Asset

A collection of data elements or datasets that make sense to group together. Each community of interest identifies the Data Assets specific to supporting the needs of their respective mission or business functions. Notably, a Data Asset is a deliberately abstract concept. A given Data Asset may represent an entire database consisting of multiple distinct entity classes, or may represent a single entity class. *Source: US OD*

Data page

A hub for data discovery which provides a common location that lists and links to an organization’s datasets. Such a hub is often located at www.example.com/data. *Source: US OD*

Data cleaning

Processing a dataset to make it easier to consume. This may involve fixing inconsistencies and errors, removing non-machine-readable elements such as formatting, using standard labels for row and column headings, ensuring that numbers, dates, and other quantities are represented appropriately, conversion to a suitable file format, reconciliation of labels with another dataset being used (see data integration), etc. *Source: ODH*

Data collection

Datasets are created by collecting data in different ways: from manual or automatic measurements (e.g. weather data), surveys (census data), records of decisions (budget data) or ongoing transactions (spending data), aggregation of many records (crime data), mathematical modelling (population projections), etc. *Source: ODH*

Data crawling

A crawler is a programme that visits web sites and reads their pages and other information in order to create entries for a search engine index. The major search engines on the web all have such a programme, which is also known as a 'spider' or a 'bot'. *Source: http://searchsoa.techtarget.com/definition/crawler; cited by EU OD*

When extracting data from the web, the term 'crawling' is often also referred to as 'data scraping' or 'harvesting'. There is a difference between these terms: crawling refers to dealing with large datasets where one can develop his or her own crawlers (or bots), which crawl to the deepest parts of the web pages. Data scraping, on the other hand, refers to retrieving information from any source (not necessarily the Web). *Source: https://www.promptcloud.com/blog/data-scraping-vs-data-crawling/; cited by EU OD*

Data dump

A large amount of data transferred from one system or location to another. *Source: EU OD*

Data integration

Almost any interesting use of data will combine data from different sources. To do this it is necessary to ensure that the different datasets are compatible: they must use the same names for the same objects, the same units or co-ordinates, etc. If the data quality is good this process of data integration may be straightforward but if not it is likely to be arduous. A key aim of linked data is to make data integration fully or nearly fully automatic. Non-open data is a barrier to data integration, as obtaining the data and establishing the necessary permission to use it is time-consuming and must be done afresh for each dataset. *Source: ODH*

Data journalism

The ability to work with data is an increasingly important part of a journalist's armoury. Skills needed to research and tell a good data-based story include finding relevant data, data cleaning, exploring or mining the data to understand what story it is telling, and creating good visualisations. *Source: ODH*

Data leakage

If personal data has been imperfectly anonymised, it may be possible by piecing it together (perhaps with data available from other sources) to reconstruct the identity of some data subjects together with personal data about them. The personal data, which should not have been published (see data protection), may be said to have 'leaked' from the 'anonymised' data. Other kinds of confidential data can also be subject to leakage by, for example, poor data security measures. See de-identification. *Source: ODH*

Data management

The policies, procedures, and technical choices used to handle data through its entire lifecycle from data collection to storage, preservation and use. A data management policy should take account of the needs of data quality, availability, data protection, data preservation, etc. *Source: ODH*

Data Mining

The practice of examining large pre-existing databases in order to generate new information. 'For example, one Midwest grocery chain used the data mining capacity of Oracle software to analyse local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer display closer to the diaper display. And, they could make sure beer and diapers were sold at full price on Thursdays.' *Source: http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm , cited by EU OD*

Data parsing

Breaking a data block into smaller chunks by following a set of rules so that it can be more easily interpreted, managed or transmitted by a computer. *Source: http://www.businessdictionary.com/definition/parsing.html cited by EU OD*

Data portal

A web platform for publishing data. The aim of a data portal is to provide a data catalogue, making data not only available but discoverable for data users, while offering a convenient publishing workflow for publishing organisations. Typical features are web interfaces for publishing and for searching and browsing the catalogue, machine interfaces (APIs) to enable automatic publishing from other systems, and data preview and visualisation. *Source: ODH*

Data protection by design and by default

Mandated by the General Data Protection Regulation, DPbDD is a core obligation of data controllers and data processors to ensure effective implementation of data protection principles and data subjects' rights and freedoms. The controllers are required to implement appropriate technical and organisational measures and necessary safeguards and are obliged to demonstrate the effectiveness of implemented measures. *Sources: EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Adopted on 13 November 2019*

Data preservation

An act of conserving and maintaining both the safety and integrity of data. Preservation is done through formal activities that are governed by policies, regulations and strategies directed towards protecting and prolonging the existence and authenticity of data and its metadata. The main goal of data preservation is to protect data from being lost or destroyed and to contribute to the reuse and progression of the data. Long-term preservation of datasets is a challenge owing to uncertainty about the future of file formats, computer architectures, storage media and network connectivity. Projects that put particular stress on data preservation take a variety of approaches to solving these problems. *Sources: ODH and http://ifdo.org/preservation/*

Data quality

A measure of the useableness of data. An ideal dataset is accurate, complete, timely in publication, consistent in its naming of items and its handling of e.g. missing data, and directly machine-readable (see data cleaning), conforms to standards of nomenclature in the field, and is published with sufficient metadata that users can easily understand, for example, who it is published by and the meaning of the variables in the dataset. *Source: ODH*

Data subject

Identified or identifiable natural person, a natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. *Source: General Data Protection Regulation, Article 4.1.1.*

Data wrangler

A person converting data into a usable form so that they can be easily used with automated or semi-automated tools. Data wrangling may include further data cleaning. *Sources: ODH*

Database

Any organised collection of data may be considered a database. In this sense the word is synonymous with dataset. It is a collection of data stored according to a schema and manipulated according to the rules set out in one Data Modelling Facility. *Sources: ODH and USOD*

Another meaning relates to a software system for processing and managing data, including features to extend or update, transform and query the data. Examples are the open source PostgreSQL, and the proprietary Microsoft Access. *Source: ODH*

Database rights

A right to prevent or restrict others from extracting and reusing content from a database. In the EU it is regulated by a special piece of legislation – Directive 96/9/EC on the legal protection of databases.

Dataset

A collection of related sets of data that is composed of separate elements but that can be manipulated as a unit and accessed or downloaded in one or more formats. *Source: EU OD*

A collection of data, published or curated by a single source, and available for access or download in one or more formats. *Source: https://data.europa.eu/euodp/en/developerscorner*

Any organised collection of data. 'Dataset' is a flexible term and may refer to an entire database, a spreadsheet or other data file, or a related collection of data resources. *Source: ODH*

The most basic representation of a dataset is data elements presented in tabular form. Each column represents a particular variable. Each row corresponds to a given value of that column's variable. A dataset may also present information in a variety of non-tabular formats, such as an extensible mark-up language (XML) file, a geospatial data file, or an image file, etc. *Source: US OD*

DCAT

Data catalogue vocabulary. RDF vocabulary for interoperability of data catalogues. *Source: http://www.w3.org/TR/vocab-dcat ; cited by EU OD*

DCAT App

DCAT application profile. A common vocabulary for describing datasets hosted in data portals in Europe, based on the DCAT. *Source: : https://joinup.ec.europa.eu/asset/dcat_application_profile/description, cited by EU OD*

DCMI (Dublin Core Meatadata)

Dublin core metadata initiative, an open organisation supporting innovation in metadata design and best practices across the metadata ecology. *Source: http://dublincore.org/*

Deep Learning

Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts by using multiple layers to progressively extract higher level features from the raw input. Because the computer gathers knowledge from experience, there is no need for a human computer operator formally to specify knowledge needed by the computer. *Source: Deng, L.; Yu, D. (2014). "Deep Learning: Methods and Applications" . Foundations and Trends in Signal Processing. 7 (3–4): 1–199.*

De-Identification

A form of anonymisation where personal records are kept intact but specific identifying information, such as names, are replaced with anonymous identifiers. Compared to aggregation, de-identification carries a greater risk of data leakage: for example, if prison records include a prisoner's criminal record and medical history, the prisoner could in many cases be identified even without their name by their criminal record, giving unauthorised access to their medical history. In other cases, this risk is absent, or the value of the un-aggregated data is so great that it is worth making de-identified data available subject to carefully designed safeguards. *Source: ODH*

DET

The Digital Europa Thesaurus (DET) is a multilingual thesaurus covering the main subject matters of the European Commission's public communications. It has been designed to describe and index web content from across the European Commission so that this content can be retrieved, aggregated, and managed. The thesaurus is maintained by DG COMM. *Source: https://op.europa.eu/en/web/eu-vocabularies/det*

Dimension

An ordinary table or spreadsheet can easily represent two data dimensions: each data point has a row and a column. Plenty of real-world data has more dimensions, however: for example, a dataset of Earth surface temperature varying with position and time (two co-ordinates are required to specify the position on earth, e.g. latitude and longitude, and one to specify the time). *Source: ODH*

Discoverable

It is not enough for open data to be published if potential users cannot find it, or even do not know that it exists. Rather than simply publishing data haphazardly on websites, governments and other large data publishers can help make their datasets discoverable by indexing them in catalogues or data portals. *Source: ODH*

DOI

Digital Object Identifier, an identifier for a digital object (such as a document or dataset) that is assigned by a central registry and is therefore guaranteed to be a globally unique identifier: no two digital objects in the world will have the same DOI. *Source: ODH*

Dynamic data

Documents in digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence. Data generated by sensors are typically considered to be dynamic data. *Source: OD Directive.*

E

ELI (European legislation identifier)

European legislation identifier, allowing to uniquely identify and access national and European legislation online and to guarantee easier access, exchange and reuse of legislation for public authorities, professional users, academics and citizens. ELI paves the way for a semantic web of legal gazettes and official journals. *Source: https://en.wikipedia.org/wiki/European_Legislation_Identifier cited by EU OD*

Endpoint

An association between a binding and a network address, specified by a URI, that may be used to communicate with an instance of a service. An end point indicates a specific location for accessing a service using a specific protocol and data format. *Source: US OD*

Error Response Code

Errors are an inevitable part of API integration, and providing not only a robust set of clear and meaningful API error response codes, but a clear listing of these codes for developers to follow and learn from is essential.

API errors are directly related to frustration during developer integration, the more friendly and meaningful they are, the greater the chance a developer will move forward after encountering an error. Put a lot of consideration into your error responses and the documentation that educates developers. *Source: US OD*

EU Open Data Directive

The *Directive on open data and the re-use of public sector information* 2019/1024/EU is a recent legislative piece on open data which replaces the former PSI Directive of 2003 and 2013. The Directive included the term 'open data' in its title, and broadened the scope of application onto public enterprises. It introduced the concept of high-value datasets, insisting on the use of APIs and dynamic data. The OD Directive extends also to reserch data abeit in different regime. The OD Directive came into force in July 2019, and has to be transposed by the member states until July 2021.

EU PSI Directive

The *Directive on the re-use of public sector information*, 2003/98/EC, 2013/37/EU is the first EU legislation on the reuse of public sector information which established a requirement to ensure that public sector publishes its open data in machine readable format, and that it handles the requests of users, with a possibility of a redress before an independent authority and/or the court. It restricted the possibilities for

discrimination among users and for the granting the exclusive rights. It also requires of all member states to establish 'practical arrangements' for publication of open data, what in practice led to the development of the member states' data portals and the European open data portal that connects them. The Directive will be replaced by the EU Open Data Directive which came into force in July 2019, and has to be transposed by the member states until July 2021.

European Data portal

An open data portal which displays data from the EU member states national open data portals, as well as the portals of the EEA countries and some candidate and neighbouring countries. It should not be confused with the EU open data portal which contains datasets from the EU institutions, agencies and services. *See:* *https://www.europeandataportal.eu/*

EuroVoc

EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU, the European Parliament in particular. It contains terms in 23 EU languages (Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish), plus in three languages of countries which are candidates for EU accession: Macedonian, Albanian and Serbian. *Source:* *https://op.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc*

Exclusive arrangements

The legal situation in which one user has an exclusive right to re-use a dataset, excluding others from such a re-use. The OD Directive in general prohibits the of exclusive arrangements making the re-use of documents open to all potential actors in the market, even if one or more market actors already exploit added-value products based on those documents. The exclusive rights can be granted in only exceptional situation, such as the necessity to grant the right for the provision of a service in general interest.

Expert system

In artificial intelligence, an expert system is a computer system that emulates the decision-making ability of a human expert and represents one of the current applicable results of machine learning and a possible step towards development of Artificial Intelligence. *Source*: Jackson, Peter (1998), Introduction To Expert Systems (3 ed.), Addison Wesley, p. 2

F

File format

The description of how a file is represented on a computer disk. The format usually corresponds to the last part of the file name ('extension'), e.g. a file in CSV format might be called schools-list.csv. The file format refers to the internal format of the file, not how it is displayed to users. E.g. CSV and XLS files are structured very differently on disk, but may look similar or identical when opened in a spreadsheet program such as Excel. *Source: ODH*

Five stars of open data

A rating system for open data proposed by Tim Berners-Lee, founder of the World Wide Web. To score the maximum five stars, data must (1) be available on the Web under an open licence, (2) be in the form of structured data, (3) be in a non-proprietary file format, (4) use URIs as its identifiers (see also RDF), (5) include links to other data sources (see linked data). To score 3 stars, it must satisfy all of (1)-(3), etc. See also: https://5stardata.info/en/; *Source: ODH*

FOAF

'Friend of a friend' is a machine-readable descriptive vocabulary of persons, their activities and their relations to other people and objects. FOAF allows groups of people to describe social networks without the need for a centralised database. *Source: https://en.wikipedia.org/wiki/FOAF_%28ontology%29 , cited by EU OD*

Formal open standard

As defined by the OD Directive, 'formal open standard' means a standard which has been laid down in written form, detailing specifications for the requirements on how to ensure software interoperability. *Source: OD Directive*

G

GDPR

Key EU legislation on the personal data protection and privacy. The General Data Protection Regulation (EU)2016/679 is in force since May 2018 in the EU and EEA countries. It aims to give the individuals control over their personal data, requires all data controllers to establish procedures and safeguards, and is very strict towards breaches committed by private companies.

GIS

Geographical Information System, any computer system designed to read, display, analyse and manipulate geodata. *Source: ODH*

GNU General Public Licence (GPL)

The GNU General Public License is a free, "*copyleft"* license for software and other kinds of works in a sense that it protects the freedoms – the rights - of users of thus licensed software instead of focusing on the moral and material rights of authors within the recognized and almost universally accepted framework of copyright and neighbouring rights. *Source: Katulić, T.:" Opportunities and pitfalls of GPL software licencing agreement from the perspective of the software developer", Central European Conference on Information and Intelligent Systems, CECIIS 2013.*

GPS

The Global Positioning System, a satellite-based system which provides exact location information to any equipment with a suitable receiver (including modern smartphones). GPS is invaluable to many location-based apps, providing users with e.g. route-finding information or weather forecasts based on their current location. GPS is also a striking example of successful open data, as it is maintained by the US government and provided free of charge to anyone with a GPS receiver. *Source: ODH*

GeoJSON

A dialect of JSON with specialised features for describing geodata, and hence a popular interchange format for geodata. *Source: ODH*

Geodata

Any dataset where data points include a location, e.g. as latitude and longitude or another standard encoding. Maps, transport routes, environmental data, cadastral data, and many other kinds of data can be published as geodata. *Source: ODH*

GitHub

GitHub is a social coding platform allowing developers to publicly or privately build code repositories and interact with other developers around these repositories– providing the ability to download or fork a repository, as well as contribute back, resulting in a collaborative environment for software development. *Source: US OD*

H

Hackathon

An event, usually over one or two days, where developers, subject experts and others come together to create apps, visualisations and prototypes that aim to address problems in a particular domain, usually making heavy use of data. Hackathons focusing on a particular collection of data are a possible form of community engagement by data publishers. The hackathon is a popular format in the open source community. *Source: ODH.*

An event in which computer programmers and others in the field of software development, like graphic designers, interface designers, project managers and computational philologists, collaborate intensively on software projects. Occasionally, there is a hardware component as well. Hackathons typically last between a day and a week in length. Some hackathons are intended simply for educational or social purposes, although in many cases the goal is to create usable software. Hackathons tend to have a specific focus, which can include the programming language used, the operating system, an application, an API, the subject, or the demographic group of the programmers. In other cases, there is no restriction on the type of software being created. *Source: USOD*

High-value datasets

Document the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets; Under the OD Directive such datasets should be in general free of charge, machine readable, provided via APIs and as bulk download where relevant. *Source: OD Directive.*

Host

A company that stores a customer's data on its own (the host's) computers and makes it available over the internet. A hosted service is one that runs and stores data on the service-provider's computers and is accessed over the network. See also SaaS. *Source: ODH.*

Human Readable

Data in a format that can be conveniently read by a human. Some human-readable formats, such as PDF, are not machine-readable as they are not structured data, i.e. the representation of the data on disk does not represent the actual relationships present in the data. *Source: ODH.*

I

IMMC Core Metadata

The minimum set of metadata elements, the so-called IMMC core metadata, that is to be used in the data exchange. IMMC Core Metadata, within the context of the Interinstitutional Metadata Maintenance Committee (IMMC), is defined as: the minimum set of metadata elements relative to the legal decision-making process, to be used in the data exchange between the institutions involved and the Publications Office. *Source: http://publications.europa.eu/mdr/core-metadata/ cited by EU OD*

Identifier

The name of an object or concept in a database. An identifier may be the object's actual name (e.g. 'London' or 'W1 1AA', a London postcode), or a word describing the concept ('population'), or an arbitrary identifier such as 'XY123' that makes sense only in the context of the particular dataset. Careful choice of identifiers using relevant standards can facilitate data integration. *Source: ODH.*

Information

A structured collection of data presented in a form that people can understand and process. Information is converted into knowledge when it is contextualised with the rest of a person's knowledge and world model. *Source: ODH.*

Any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audio-visual forms. *Source: US OD*

Information Asset Register (IAR)

IARs are registers specifically set up to capture and organise meta-data about the vast quantities of information held by government departments and agencies. A comprehensive IAR includes databases, old sets of files, recent electronic files, collections of statistics, research and so forth.

IARs can be developed in different ways. Government departments can develop their own IARs and these can be linked to national IARs. IARs can include information which is held by public bodies but which has not yet been – and maybe will not be – proactively published. Hence, they allow members of the public to identify information which exists and which can be requested. It is important that IARs are complete as possible because otherwise possible re-users could be discouraged from finding or requesting the dataset.

It is essential that the metadata in the IARs should be comprehensive so that search engines can function effectively. In the spirit of open government data, public bodies should make their IARs available to the general public as raw data under an open license so that civic hackers can make use of the data, for example by building search engines and user interfaces. The EU PSI Directive and the EU OD Directive require of Member states to establish tools that help re-users to find documents available for re-use, such as asset lists. *Source: ODH.*

Information Life Cycle

The stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition. *Source: US OD*

Information Society

The information society is one in which information is the defining feature, unlike the industrial society where steam power and fossil fuels were distinguishing elements. Information society is a concept that responds to the expansion and ubiquity of information. The term has been in use since the 1970s, but has gained in popularity and is now widely used in social and political policy. *Source: Information Society https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100003718*

Information System

A discrete set of information resources organized for the collection, processing, maintenance, transmission, and dissemination of information, in accordance with defined procedures, whether automated or manual. Source: US OD

Information System Life Cycle

The phases through which an information system passes, typically characterized as initiation, development, operation, and termination. *Source: US OD*

INSPIRE Directive

The Directive establishing and Infrastructure for Spatial Information in the European Community (INSPIRE Directive) enacted in 2007 (2007/2/EC) established an infrastructure for spatial information in Europe to support Community environmental policies and policies or activities which may have an impact on the environment. It is based on the infrastructure operated by the Member States of the EU and it addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. To ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and transboundary context, the Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas (Metadata, Data Specifications, Network Services, Data and Service Sharing and Monitoring and Reporting). These IRs are adopted as Commission Decisions or Regulations. *Source: https://inspire.ec.europa.eu/inspire-directive/2*

Intellectual property rights / IP rights

The term intellectual property rights in the context of open data refers primarily to copyright and related rights, including *sui generis* forms of protection. The OD Directive does not apply to documents covered by industrial property rights, such as patents and registered designs and trademarks. The Directive neither affects the existence or ownership of intellectual property rights of public sector bodies, nor does it limit the exercise of these rights in any way beyond the boundaries set by this Directive. The obligations imposed in accordance with the OD Directive apply only insofar as they are compatible with international agreements on the protection of intellectual property rights, in particular the Berne Convention for the Protection of Literary and Artistic Works (Berne Convention), the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) and the WIPO Copyright Treaty (WCT). Public sector bodies are expected to exercise their copyright in a way that facilitates re-use. *Source: OD Directive*

Internet

A worldwide network of interconnected computer networks that use the Internet protocol suite (TCP/IP) to facilitate data transmission and exchange among several billion devices, which are logically linked together by a globally unique address space. *Source: ODH*

Interoperability

The ability of systems to exchange information and use the exchanged information. *Source: EU OD*

ISA2 Programme

Interoperable solutions for European public administrations Programme is a European Commission-funded programme aiming at facilitating transactions among cross-border and/or cross-sector public administrations in Europe and supporting development of digital solutions that enable public administrations, businesses and citizens in Europe to benefit from interoperable cross-border and cross-sector public services. ISA2 is established in 2015 to run from 2016 to 2020, as the follow-up programme to ISA 2010-2015. *Source: https://ec.europa.eu/isa2/isa2_en*

J

JSON

JavaScript Object Notation, a simple but powerful format for data. It can describe complex data structures, is highly machine-readable as well as reasonably human-readable, and is independent of platform and programming language, and is therefore a popular format for data interchange between programs and systems. *Source: ODH.*

JavaScript object notation is an open standard format that uses human readable text to transmit data objects consisting of attribute–value pairs. It is the most common data format used for asynchronous browser/server communication (AJAJ). *Source: EUOD.*

JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language. *Source: USOD*

JSONP

JSONP or “JSON with padding” is a JSON extension wherein the name of a callback function is specified as an input argument of the underlying JSON call itself. JSONP makes use of runtime script tag injection. *Source: USOD*

K

KML

Keyhole Markup Language, an XML-based open format for geodata. KML was devised for Keyhole Earth Viewer, later acquired by Google and renamed Google Earth, but has been an international standard of the Open Geospatial Consortium since 2008. *Source: ODH.*

L

Licence

A legal instrument by which a copyright holder may grant rights over the protected work. Data and content is open if it is subject to an explicitly-applied licence that conforms to the Open Definition. A range of standard open licences are available, such as the Creative Commons CC-BY licence, which requires only attribution. *Source: ODH.*

Licence mixing

If Project X publishes content, and wants to include content from Project Y, it is necessary that Y's licence permits at least the same range of re-uses as X's licence. For example, content published under a non-commercial licence cannot be included in Wikipedia, since Wikipedia's open licence includes rights for commercial re-use which cannot be granted for the non-commercial data, an example of a failure of licences to mix well. *Source: ODH.*

Linked data

A form of data representation where every identifier is an http://… URI, using standard lists (see vocabulary) of identifiers where possible, and where datasets include links to reference datasets of the same objects. A key aim is to make data integration automatic, even for large datasets. Linked data is usually represented using RDF. See also five stars of open data; triple store. Source: ODH

Central for the concept of the semantic web; linked data assigns a web address, similar to a website address, to each piece of data, enabling connection of data through the web. It builds upon standard web technologies such as HTTP and URI, but rather than using them to serve web pages for human readers it extends them to share information in a way that can be read automatically by computers. The connections between linked data can grow without limitations. Linked data is particularly useful for analysing different types of data from various datasets, for example government data. *Source: EUOD*

Linked data principles

Linked data principles provide a common API for data on the web that is more convenient than many separately and differently designed APIs published by individual data suppliers. Tim Berners-Lee, inventor of the web and initiator of the linked data project, proposed the following principles upon which linked data is based. Use URIs to name things. Use HTTP URIs so that things can be referred to and looked up (dereferenced) by people and user agents. When someone looks up a URI, provide useful information using open web standards such as RDF or Sparql. Include links to other related things using their URIs when publishing on the web. *Source: W3C — http://www.w3.org/TR/ld-glossary/#linked-data-principles*

M

Machine learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Source: Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

Machine-Readable (File or Data or Format)

Data in a data format that can be easily automatically read and processed (or identified, recognized and extracted) by a computer, without human intervention, including individual statements of fact, and their internal structure, while ensuring no semantic meaning is lost. Examples of file formats are CSV, JSON, XML, etc. Machine-readable data must be structured data.

Non-digital material (printed or hand-written documents) is by its non-digital nature not machine-readable, as well as digital material in certain file formats, as a PDF document containing tables of data, but they are (Compare →) *human-readable*. The equivalent tables in a format such as a spreadsheet would be machine readable.

As another example scans (photographs) of text are not machine-readable (but are human readable!) but the equivalent text in a format such as a simple ASCII text file can machine readable and processable.

The appropriate machine-readable format may vary by type of data - so, for example, machine readable formats for geographic data may differ from those for tabular data. *Sources: ODH and US OD and OD Directive*

There are two types of machine-readable data: human-readable data that is marked up so that it can also be understood by computers, for example microformats or RDFa; data formats intended principally for computers, for example RDF, XML and JSON. *Source: EU OD*

Mashup

The combination of multiple datasets from multiple sources to create a new service or visualisation or new information. *Source: EU OD*

Many eyes principle

If something is visible to many people then, collectively, they are more likely to find errors in it. Publishing open data can therefore be a way to improve its accuracy and data quality, especially where a good interface for reporting errors is provided. See → crowdsourcing. *Source: ODH*

Marginal cost

The additional cost incurred by supplying a single copy of a resource, e.g. data. For data to be open according to the Open Definition, it must be charged for at no more than marginal cost. Where data is available for download over the internet the marginal cost will usually be zero. There may be a small marginal cost in exceptional cases, e.g. if for reasons of size the data needs to be put on a disk and posted. The OD Directive considers costs incurred for the reproduction, provision and dissemination of documents as well as for anonymisation of personal data and measures taken to protect commercially confidential information as being marginal costs. *Sources: ODH, OD Directive*

Metadata

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data.

It consists of information about a dataset such as its title and description, method of collection, author or publisher, area and time period covered, licence, date and frequency of release, etc. It is essential to publish data with adequate metadata to aid both discoverability and usability of the data. *Source: ODH, EU OD and NISO — http://www.niso.org/publications/press/UnderstandingMetadata.pdf*

MDR (Metadata Registry)

The Metadata Registry is an important interoperability and standardisation tool. It registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different EU institutions. *Source: EU OD and http://publications.europa.eu/mdr/index.html*

N

NGO

Non-governmental organisation. NGOs are voluntary, non-profit organisations focussing on charitable work, community-building, campaigning, research, etc, making up a vital part of civil society. *Source: ODH*

No rights reserved (CC 0)

Used to denote that the holder of copyright or database rights waive all their interest in their data worldwide. May not be applicable in all legal systems. *Source: Creative Commons.*

Non-commercial

A restriction, as part of a licence, that content cannot be freely re-used for 'commercial' purposes. Content or data subject to a non-commercial restriction is not open, according to the Open Definition. Such a restriction reduces economic value and causes problems with licence mixing, as well as often ruling out more than is intended (for example, it is often unclear whether educational uses are 'commercial'). The intent of a non-commercial clause may be better captured by a share-alike requirement. See also → Licences. *Source: ODH*

Non-discrimination principle

A principle of the OD Directive and PSI Directive requiring that any applicable conditions for the re-use of documents are non-discriminatory for comparable categories of re-use, including for cross-border re-use, and limiting the possibilities of exclusive rights.

O

ODRA

Open Data Readiness Assessment, a framework created by the World Bank for assessing the opportunities, obstacles and next steps to be taken in a country (especially a developing country) considering publishing government data as open data. *Source: ODH*

ODbL

Open Database Licence, an attempt to create an open licence for data which covers the 'database rights' as well as copyright itself. It does this by imposing contractual obligations on the data re-user. Unfortunately contract law is fundamentally different from copyright law, since copyright is inherent in a work and binds all downstream users of the work, whereas a contract only binds the parties to the contract and has no force on a later re-user of re-published data. The ODbL remains useful nevertheless, and other attempts are being made to create open licences specifically for data. *Source: ODH*

OGP

The Open Government Partnership, a partnership of national governments launched in 2011 with the aim of promoting open government in the member countries and collaborating on multi-lateral agreements and best practices. It covers also open data.

Ontology

A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms). *Source: W3C — http://www.w3.org/TR/ld-glossary/#ontology as cited by EU OD*

Open access

The principle that access to the published papers and other results of research, especially publicly-funded research, should be freely available to all. This contrasts with the traditional model where research is published in journals which charge subscription fees to readers. Besides benefits similar to the benefits of open data, proponents suggest that it is immoral to withhold potentially life-saving and valuable research from some readers who may be able to use or build on it. Open-access journals now exist and the interest of research funders is giving them some traction, especially in the sciences. *Source: ODH*

Open data

Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike. Specifically, open data is defined by the Open Definition and requires that the data be (1) Legally open: that is, available under an open (data) license that permits anyone freely to access, reuse and redistribute, and (2) Technically open: that is, that the data be available for no more than the cost of reproduction and in machine-readable and bulk form. *Source: ODH*

Open definition

The Open Definition, first released by Open Knowledge in 2005, sets out under what conditions data and content is open. Both legal and technical compatibility is vital, and the Open Definition ensures that openly-licensed data can be combined successfully, avoiding a proliferation of licences and terms of use for open data leading to complexity and incompatibility. Today it is the main international standard for open data and open data licences, with an advisory council of senior open data practitioners and can be found at opendefinition.org. The Open Definition has influenced and steered other communities of practice in the open movement, including open access to publicly-funded research, open hardware, and more, as well as governments' approach towards licences. *Source: ODH*

Open data portal

A Web-based system that contains a data catalogue with descriptions of datasets and provides services enabling discovery and re-use of the datasets. See also → Data portal. *Source: https://data.europa.eu/euodp/en/developerscorner*

Open Development

Open development seeks to bring the philosophy of the open movement to international development. It promotes open government, transparency of aid flows, engagement of beneficiaries in the design and implementation of development projects, and availability and use of open development data. *Source: ODH*

Open format

A file format that is platform-independent and made available to the public without any restriction that impedes the re-use of documents. *Source: OD Directive*

A file format with no restrictions, monetary or otherwise, placed upon its use and can be fully processed with at least one free/libre/open-source software tool. Patents are a common source of restrictions that make a format proprietary. Often, but not necessarily, the structure of an open format is set out in agreed standards, overseen and published by a non-commercial expert body. A file in an open format enjoys the guarantee that it can be correctly read by a range of different software programs or used to pass information between them. *Source: ODH*

Open government

Open government, in line with the open movement generally, seeks to make the workings of governments transparent, accountable, and responsive to citizens. It includes the ideals of democracy, due process, citizen participation and open government data. A thorough-going approach to open government would also seek to enable citizen participation in, for example, the drafting and revising of legislation and budget-setting. *Source: ODH*

Open government data

Data collected, produced or paid for by the public bodies and made freely available for reuse for any purpose. *Source: EU OD*

Open licence

Standardised public licences available online which allow data and content to be freely accessed, used, modified and shared by anyone for any purpose, and which rely on open data formats. The EU Member States are expected to encourage the use of open licences that should eventually become common practice across the Union. *Source: OD Directive*

Open movement

The open movement seeks to work towards solutions of many of the world's most pressing problems in a spirit of transparency, collaboration, re-use and free access. It encompasses open data, open government, open development, open science and much more. Participatory processes, sharing of knowledge and outputs and open source software are among its key tools. The specific definition of "open" as applied to data, knowledge and content, is set out by the Open Definition. *Source: ODH*

Open science

The practice of science in accordance with open principles, including open access publishing, publication of and collaboration around research data as open data together with associated source code, and use and development of open source data processing tools. *Source: ODH*

Open source

Software for which the source code is available under an open licence. Not only can the software be used for free, but users with the necessary technical skills can inspect the source code, modify it and run their own versions of the code, helping to fix bugs, develop new features, etc. Some large open source software projects have thousands of volunteer contributors. The Open Definition was heavily based on the earlier Open Source Definition, which sets out the conditions under which software can be considered open source. *Source: ODH*

Open Source Software

Computer software that is available in source code form: the source code and certain other rights normally reserved for copyright holders are provided under an open-source license that permits users to study, change, improve and at times also to distribute the software. Open source software is very often developed in a public, collaborative manner. Open source software is the most prominent example of open source development and often compared to (technically defined) user-generated content or (legally defined) open content movements. *Source: US OD*

Open Standards

Generally understood as technical standards that are free from licensing restrictions. Can also be interpreted to mean standards that are developed in a vendor-neutral manner. *Sources: http://schoolofdata.org/handbook/appendix/glossary cited by EU OD; ODH*

A standard developed or adopted by voluntary consensus standards bodies, both domestic and international. These standards include provisions requiring that owners of relevant intellectual property have agreed to make that intellectual property available on a non-discriminatory, royalty-free or reasonable royalty basis to all interested parties. *Source: US OD*

P

PDF

Portable Document Format, a file format for representing the layout and appearance of documents on a page independent of the layout software, computer operating system, etc. It is a file format used to present and exchange documents independently of software, hardware or operating systems. An open standard maintained by the International Organisation for Standardisation. Originally a proprietary format of Adobe Systems, PDF has been an open format since 2008. Data in PDF files is not machine-readable; see structured data. *Source: ODH and* *https://acrobat.adobe.com/be/en/products/about-adobe-pdf.html*

Predictive Modelling

Use of statistics to predict outcomes. *Source: Geisser, S.: "Predictive Inference: An Introduction", Chapman & Hall, 1993.*

Privacy

The right of individuals to a private life includes a right not to have personal information about themselves made public. A right to privacy is recognised by the Universal Declaration of Human Rights and the European Convention on Human Rights, and in the EU and the Member States of the EU it is recognised by the Charter of Fundamental Rights and is regulated by special regulation. See → GDPR. *Source: ODH*

Proprietary

(i) Proprietary software is owned by a company which restricts the ways in which it can be used. Users normally need to pay to use the software, cannot read or modify the source code, and cannot copy the software or re-sell it as part of their own product. Common examples include Microsoft Excel and Adobe Acrobat. Non-proprietary software is usually open source.

(ii) A proprietary file format is one that a company owns and controls. Data in this format may need proprietary software to be read reliably. Unlike an open format, the description of the format may be confidential or unpublished, and can be changed by the company at any time. Proprietary software usually reads and saves data in its own proprietary format. For example, different versions of Microsoft Excel use the proprietary XLS and XLSX formats. *Source: ODH*

Pseudonymization

Pseudonymization is a data management and de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. Pseudonymization is suggested as one of the technical measures that can help with compliance with the European Union's General Data Protection Regulation and its obligations for data controllers to ensure secure processing of personal data. *Source: Rec. 26, Article 4.1.5. of the General Data Protection Regulation L 119/1, 4.5.2016*

Public domain

Content to which copyright does not apply, for example because it has expired or it never existed, is free for any kind of use by anyone and is said to be in the public domain. CC0, one of the licences of Creative Commons, is a 'public domain dedication' which attempts so far as possible to renounce all rights in the work and place it in the public domain. *Source: ODH*

The Public Domain means data is free for use by anyone for any purpose without restriction under copyright law. Public domain is the purest form of open/free, since no one owns or controls the material in any way. For the official documents in most of the European countries it is the most convenient licence*. Source: Creative Commons.*

Public sector body

Public sector body' means the State, regional or local authorities, bodies governed by public law or associations formed by one or more such authorities or one or more such bodies governed by public law. *Source: OD Directive*

Public sector Information / Government data

Data that is collected, produced, reproduced, processed, disseminated, or controlled by the public sector bodies in many areas of their activity while accomplishing their institutional tasks. The work of government involves collecting huge amounts of data, much of which is not confidential (economic data, demographic data, spending data, crime data, transport data, etc). The value of much of this data can be greatly enhanced by releasing it as open data, freeing it for re-use by business, research, civil society, data journalists, etc. *Sources: ODH, US OD*

Public undertaking

Any undertaking active in the areas set out in point (b) of Article 1(1) of the OD Directive over which the public sector bodies may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it. A dominant influence on the part of the public sector bodies shall be presumed in any of the following cases in which those bodies, directly or indirectly: (a) hold the majority of the undertakings' subscribed capital (b) control the majority of the votes attaching to shares issued by the undertaking; (c) can appoint more than half of the undertaking's administrative, management or supervisory body. *Source: OD Directive*

Publisher

Anyone who distributes and makes available data or other content. Data publishers include government departments and agencies, research establishments, NGOs, media organisations, commercial companies, individuals, etc. *Source: ODH*

Q

Query

A type of question accepted by a database about the data it holds. A complex query may ask the database to select records according to some criteria, aggregate certain quantities across those records, etc. Many databases accept queries in the specialised language SQL or dialects of it. A web API allows an app to send queries to a database over the web. Compared with downloading and processing the data, this reduces both the computation load on the app and the bandwidth needed. *Source: ODH*

R

Raw data

The original data, in machine-readable form, underlying any application, visualisation, published research or interpretation, etc. An expression that refers to data in its original state that has not been processed, aggregated or manipulated in any other way. It is also defined as 'primary'. *Sources: ODH, EU OD*

RDF

A family of international standards for data interchange on the web. RDF is based on the idea of identifying things using web identifiers or HTTP URIs and describing resources in terms of simple properties and property values. *Source: W3C — http://www.w3.org/TR/ld-glossary/#rdf*

Resource Description Framework, the native way of describing linked data. RDF is not exactly a data format; rather, there are a few equivalent formats in which RDF can be expressed, including an XML-based format. RDF data takes the form of 'triples' (each atomic piece of data has three parts, namely a subject, predicate and object), and can be stored in a specialised database called a triple store. *Source: ODH*

A family of specifications for a metadata model. The RDF family of specifications is maintained by the World Wide Web Consortium (W3C). The RDF metadata model is based upon the idea of making statements about resources in the form of a subject-predicate-object expression…and is a major component in what is proposed by the W3C's Semantic Web activity: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and utilize metadata about the vast resources of the Web, in turn enabling users to deal with those resources with greater efficiency and certainty. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity. *Source: US OD*

RDFa Resource description framework in attributes

A W3C recommendation that adds a set of attribute-level extensions to HTML, XHTML and various XML-based document types for embedding rich metadata within web documents. *Source: https://en.wikipedia.org/wiki/RDFa*

Re-use

A use of public sector information usually by third persons for purposes other than initial purposes within the public tasks for which data were originally collected, produced, or disseminated. It is rare that data gathered for a particular purpose does not have other possible uses. Once gathered, for whatever reason, data can be re-used again and again, in ways that were never envisaged when it was collected, provided only that the data-holder makes it available under an open licence to enable such re-use. Under the EU OD Directive, data can be re-used for both commercial and non-commercial purposes. An exchange of documents between public sector bodies is not considered to be a re-use of public sector information. *Sources: ODH, OD Directive*

Real time

Data (such as the current location of trains on a network) which is being constantly updated, where a query needs to be against the latest version of the data. *Source: ODH*

Reasonable return on investment

A percentage of the overall charge, in addition to that needed to recover the eligible costs, not exceeding 5 percentage points above the fixed interest rate of the ECB. Used to calculate costs above marginal costs by the EU Open data Directive. *Source: OD Directive*

Request for the re-use of public sector information

A possibility for the user to require the disclosure of the dataset by the public bodies in a machine-readable format suitable for re-use. It is a formal procedure that usually requires the issuing of the formal decision. The decision of the public body is scrutinised by the independent authority or the court.

Research data

Documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results. *Source: OD Directive*

Traditionally the data was kept by researchers and only final research outputs, such as papers analysing the data, would be published. Open science holds that the data should be published, both to increase verifiability of the work and to enable it to be used in other research. The full spirit of open science collaboration demands data publication early in the project, but research culture will need to change appreciably before this becomes widespread. Research data management (RDM) is an emerging discipline that seeks best practices in handling this. Research data has been covered by the EU Directive regulating open data and the reuse of public sector information since 2019. *Source: ODH*

Resource

CKAN uses this term to denote one of the individual data objects (a file such as a spreadsheet, or an API) in a dataset. *Source: ODH*

The physical representation of a dataset. Each resource can be a file of any kind, a link to a file elsewhere on the web or a link to an API. For example, if the data is being supplied in multiple formats or split into different areas or time periods, each file is a different 'resource' that should be described individually. *Source: http://www.w3.org/*

RSS

A family of web feed formats (often dubbed Really Simple Syndication) used to publish frequently updated works — such as blog entries, news headlines, audio, and video — in a standardized format. An RSS document (which is called a "feed," "web feed," or "channel") includes full or summarized text, plus metadata such as publishing dates and authorship. *Source: US OD*

S

SaaS

Software as a Service, i.e. a software program that runs, not on the user's machine, but on the machines of a hosting company, which the user accesses over the web. The host takes care of associated data storage, and normally charges for the use of the service or monetises its client base in other ways. *Source: ODH*

Schema

An XML schema defines the structure of an XML document. An XML schema defines things such as which data elements and attributes can appear in a document; how the data elements relate to one another; whether an element is empty or can include text; which types of data are allowed for specific data elements and attributes; and what the default and fixed values are for elements and attributes. A schema is also a description of the data represented within a database. The format of the description varies but includes a table layout for a relational database or an entity-relationship diagram. It is method for specifying constraints on XML documents. *Source: US OD*

Scraping

Extracting data from a non-machine-readable source, such as a website or a PDF document, and creating structured data from the result. Screen-scraping a dataset requires dedicated programming and is expensive in programmer time, so is generally done only after all other attempts to get the data in structured form have failed. Legal questions may arise about whether the scraping breaches the source website's copyright or terms of service. *Source: ODH*

The process of extracting data in machine-readable formats of non-pure data sources, for example webpages or PDF documents. Often prefixed with the source (web scraping, PDF scraping). *Sources: http://en.wikipedia.org/wiki/Data_scraping, http://schoolofdata.org/handbook/appendix/glossary*

SDMX

Statistical data and metadata exchange, an international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries. *Source: https://en.wikipedia.org/wiki/SDMX, cited by EU OD*

Semantic Web

An evolution or part of the World Wide Web that consists of machine-readable data in RDF and an ability to query that information in standard ways (e.g. via Sparql). *Source: W3C — http://www.w3.org/TR/ld-glossary/#semantic-web cited by EU OD*

SEO

Search engine optimisation, a series of techniques which improve the visibility of a website in search engines result page (SERP) via the natural or unpaid ('organic' or 'algorithmic') search results. The goal of such optimisation is to rank as high as possible for a certain search query. *Source: https://en.wikipedia.org/wiki/Search_engine_optimization, cited by EU OD*

Server

A computer on the internet, usually manged by a hosting company, that responds to requests from a user, e.g. for web pages, downloaded files or to access features in a SaaS package being run on the server. *Source: ODH*

Shapefile

A popular file format for geodata, maintained and published by Esri, a manufacturer of GIS software. A Shapefile actually consists of several related files. Though the format is technically proprietary, Esri publishes a full specification standard and Shapefiles can be read by a wide range of software, so function somewhat like an open standard in practice. *Source: ODH*

Share-alike Licence

A license that requires users of a work to provide the content under the same or similar conditions as the original. *Source: ODH*

SOAP

SOAP (Simple Object Access Protocol) is a message-based protocol based on XML for accessing services on the Web. It employs XML syntax to send text commands across the Internet using HTTP. SOAP is similar in purpose to the DCOM and CORBA distributed object systems, but is more lightweight and less programming-intensive. Because of its simple exchange mechanism, SOAP can also be used to implement a messaging system. *Source: US OD*

Solr

An open source enterprise search platform. Its major features include fulltext search, hit highlighting, faceted search, real-time indexing, dynamic clustering, database integration and rich document (e.g. Word, PDF) handling. *Source: https://en.wikipedia.org/wiki/Apache_Solr*

Source code

The files of computer code written by programmers that are used to produce a piece of software. The source code is usually converted or 'compiled' into a form that the user's computer can execute. The user therefore never sees the original source code, unless it is published as open source. *Source: ODH*

SPARQL

A query language similar to SQL, used for queries to a linked-data triple store. Source: ODH

Sparql protocol and RDF query language (Sparql) defines a query language for RDF data, analogous to the structured query language for relational databases. *Source: W3C — http://www.w3.org/TR/ld-glossary/#sparql*

SPARQL Endpoint

A service that accepts Sparql queries and returns answers to them as Sparql result sets. It is best practice for dataset providers to give the URL of their Sparql endpoint to allow access to their data programmatically or through a web interface. *Source: W3C — http://www.w3.org/TR/ld-glossary/#sparql-endpoint*

Spreadsheet

A table of data and calculations that can be processed interactively with a specialised spreadsheet program such as Microsoft Excel or OpenOffice Calc. *Source: ODH*

SQL

Structured Query Language, a standard language used for interrogating many types of database. See → query. *Source: ODH*

Standard

A published specification for, e.g., the structure of a particular file format, recommended nomenclature to use in a particular domain, a common set of metadata fields, etc. Conforming to relevant standards greatly increases the value of published data by improving machine readability and easing data integration. *Source: ODH*

Standard licence

A set of predefined re-use conditions in a digital format, preferably compatible with standardised public licences available online. *Source: OD Directive*

Structured data

All data has some structure, but 'structured data' refers to data where the structural relation between elements is explicit in the way the data is stored on a computer disk. XML and JSON are common formats that allow many types of structure to be represented. The internal representation of, for example, word-processing documents or PDF documents reflects the positioning of entities on the page, not their logical structure, which is correspondingly difficult or impossible to extract automatically. Source: ODH

Data that resides in fixed fields within a record or file. Relational databases and spreadsheets are examples of structured data. Although data in XML files is not fixed in locations like traditional database records, it is nevertheless structured, because the data is tagged and can be accurately identified. *Source: PC Magazine encyclopaedia — http://www.pcmag.com/encyclopedia/term/52162/ , cited by EU OD*

T

Tab-separated values

Tab-separated values (TSV) are a very common form of text file format for sharing tabular data. The format is extremely simple and highly machine-readable. *Source: ODH*

Transparency

A principle of the public administration and democratic political system. Governments and other public sector bodies are said to be transparent when their workings and decision-making processes are well-understood, properly documented and open to scrutiny. Transparency is one of the aspects of open government. An increase in transparency is one of the benefits of open data. *Source: ODH*

Transport data

Public transport routes, timetables and real time data are valuable but difficult candidates for open data. Even when they are published, data from different transit authorities and companies may not be available in compatible formats, making it difficult for third parties to provide integrated transport information. Many transport authorities distribute public transport data using the General Transit Feed Specification (GTFS) which is maintained by Google. Work on standardisation and more open data is ongoing in the sector. *Source: ODH*

Triple store

The 'triples' of RDF data can be stored in a specialised database, called a triple store, against which queries can be made in the query language SPARQL. *Source: ODH*

A triplestore is a purpose-built database for the storage and retrieval of triples through semantic queries. A triple is a data entity composed of subject–predicate–object, like 'Bob is 35' or 'Bob knows Fred'. Much like a relational database, information is stored in a triplestore and retrieved via a query language. Unlike a relational database, a triplestore is optimised for the storage and retrieval of triples. In addition to queries, triples can usually be imported/exported using RDF and other formats. *Source: http://en.wikipedia.org/wiki/Triplestore*

TSV

A simple text format for a database table. Each record in the table is one line of the text file. Each field value of a record is separated from the next by a tab stop character. It is a form of the more general delimiter-separated values format. *Source: US OD*

U

Unconference

A meeting, similar to a conference, but with no agenda fixed in advance. Using various established techniques, participants jointly agree on the day what sessions will run. Some more traditional conference sessions with invited speakers may also be included. A popular format among the tech community, an unconference can be combined with or run alongside a hackathon based on open data. It is a possible method of community engagement by data publishers. *Source: ODH*

Unique identifier (UID)

An identifier for an object which is guaranteed to be different from identifiers of all other objects in a collection. Within a database, every object will have a UID that is unique within the database. A UID assigned by a central registry (such as an ISBN for books, or a DOI for data) will be unique for all objects for which it is assigned. The http://… identifiers of linked data provide a technique for guaranteeing UIDs without a central authority. *Source: ODH*

Unstructured Data

Data that is more free-form, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchical sequence, nor does it follow any relational rules. Unstructured data refers to masses of (usually) computerized information which do not have a data structure which is easily readable by a machine. Examples of unstructured data may include audio, video and unstructured text such as the body of an email or word processor document. Data mining techniques are used to find patterns in, or otherwise interpret, this information. Merrill Lynch estimates that more than 85 percent of all business information exists as unstructured data – commonly appearing in e-mails, memos, notes from call centers and support operations, news, user groups, chats, reports, letters, surveys, white papers, marketing material, research, presentations, and Web pages ("The Problem with Unstructured Data.") *Source: US OD*

URI / URL

Uniform Resource Identifier / Uniform Resource Locator. A URL is the http://… web address of some page or resource. When a URL is used in linked data as the identifier for some object, it is not strictly a locator for the object (e.g. http://dbpedia.org/page/Paris is the location of a document about Paris, but not of Paris itself), so in this context it is referred to as a URI. *Source: ODH*

URI - Uniform Resource Identifier. A string that uniquely identifies virtually anything, including a physical building or more abstract concepts such as colours. It may or may not be resolvable on the web. *Source: W3C — http://www.w3.org/TR/ld-glossary/#uniform-resource-identifier*

URL - Uniform resource locator. A global identifier commonly called a web address. A URL is resolvable on the web. All HTTP URLs are URIs; however, not all URIs are URLs. *Source: W3C — http://www.w3.org/TR/ld-glossary/#uniform-resource-locator*

V

Visualisation

A visual representation of data is often the most compelling way of communicating the data, bringing out its key features, correlations and outliers. Though many tools exist, creating a visualisation for a dataset is not an automatic process, but requires careful attention to the meaning of the variables, the relations between them and the stories inherent in the data, to design a visual representation that lets the message of the data shine through. *Source: ODH*

Vocabulary

A standard specifying the identifiers to be used for a particular collection of objects. Using standard vocabularies where they exist is key to enabling data integration. Linked data is rich in vocabularies in different topic areas. Source: ODH

A collection of terms for a particular purpose. Vocabularies can range from simple, such as the widely used RDF schema, FOAF and Dublin core metadata element set, to complex vocabularies with thousands of terms, such as those used in healthcare to describe symptoms, diseases and treatments. Vocabularies play a very important role in linked data, specifically to help with data integration. The use of this term overlaps with that of 'ontology'. *Source: W3C — http://www.w3.org/TR/ld-glossary/#vocabulary*

W

Web

The World Wide Web, the vast collection of interlinked and linkable documents and services accessible via ‘web browsers’ over the Internet. *Source: ODH*

Web 1.0

The first generation of the World Wide Web, characterised by separate static websites rather than continually updated weblogs and social networking tools. *Source: http://en.wiktionary.org/wiki/Web_1.0*

Web 2.0

A colloquial description of the part of the World Wide Web that implements social networking, blogs, user comments and ratings, as well as related human-centred activities. *Source: W3C — http://www.w3.org/TR/ld-glossary/#web-2.0*

Web 3.0

A colloquial description of the part of the World Wide Web that implements machine-readable data and the ability to perform distributed queries and analysis on that data. It is considered synonymous with the terms ‘semantic web’ and ‘the web of data’. *Source: W3C — http://www.w3.org/TR/ld-glossary/#web-3.0*

Web API

An API that is designed to work over the Internet. *Source: ODH*

Web Service

A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards *Source: US OD*

WSDL

An XML-based language (Web Services Description Language) used to describe the services a business offers and to provide a way for individuals and other businesses to access those services electronically. *Source: US OD*

X

XLS(X)

A proprietary spreadsheet format, the native format of the popular Microsoft Excel spreadsheet package. Older versions use .xls files, while more recent ones use the XML-based .xlsx variant. *Source: ODH*

XML

Extensible Markup Language, a simple and powerful standard for representing structured data. It is a markup language that defines a set of rules for encoding documents in a format which is both human readable and machine readable. It is a flexible language for creating common information formats and sharing both the format and content of data over the Internet and elsewhere. XML is a formatting language recommended by the World Wide Web Consortium (W3C). *Sources: https://en.wikipedia.org/wiki/XML ; ODH; US OD*

Key sources

- Directive on open data and the reuse of public sector information, EU/2019/1024 https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024
- EU vocabularies https://op.europa.eu/en/web/eu-vocabularies/home
- Open Knowledge Foundation, Open Data Handbook, https://opendatahandbook.org/glossary/en/
- Publications Office of the European Union, The Basic for the EU Data Providers, 2016, https://op.europa.eu/en/publication-detail/-/publication/e40b5572-7190-4d9c-8d92-9f5357a49b00/language-en/format-PDF/source-search
- The White House, Project Open Data CIO https://project-open-data.cio.gov/glossary/
- W3C, Linked Data Glossary, 2013, https://www.w3.org/TR/ld-glossary/#x5-star-linked-open-data

Annex III - Useful Open Data Websites

Global sources

Data Portals https://dataportals.org/

Dublin Core Metadata Initiative http://dublincore.org/

Global Open Data Index https://index.okfn.org/

Global Initiative for Fiscal Transparency http://www.fiscaltransparency.net/

Open Data Barometer https://opendatabarometer.org/

Open Data Charter https://opendatacharter.net/

Open Data Institute https://theodi.org/, https://odi.ellak.gr/

Open Definition https://opendefinition.org/

Open Fiscal Data http://www.fiscaltransparency.net/ofdp/

Open Knowledge Foundation https://okfn.org/

Open Data Impact Map https://opendataimpactmap.org/

Open Government Partnership, Open Data https://www.opengovpartnership.org/open-data/

School of Data https://schoolofdata.org/

The OECD https://www.oecd.org/gov/digital-government/open-government-data.htm

The Sunlight Foundation, Open Data Guidelines, https://sunlightfoundation.com/opendataguidelines/

The White House, The Open Data Project, https://project-open-data.cio.gov/

The World Bank, Open data readiness http://opendatatoolkit.worldbank.org/en/odra.html

W3C World Wide Web Consortium, Open Data Group https://www.w3.org/community/opendatadir/

The European Union

Current legislation on Open data https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, OJL 172, 26.6.2019 (Open Data Directive) https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

European Data Portal, The EU member states and other European countries https://www.europeandataportal.eu/en/

The EU Open Data Portal, https://data.europa.eu/euodp/en/home

The EU open data portal, Developers corner
https://data.europa.eu/euodp/en/developerscorner

Open data maturity in Europe https://www.europeandataportal.eu/hr/highlights/open-data-maturity-europe

Inspire https://inspire.ec.europa.eu/

Public Sector Information Group https://ec.europa.eu/digital-single-market/en/news/public-sector-information-group-main-page

LAPSI project https://ec.europa.eu/digital-single-market/en/news/legal-aspects-public-sector-information-lapsi-thematic-network-outputs

Open data education project https://odedu-project.eu/project-deliverables/

Share PSI https://www.w3.org/2013/share-psi/

Partner countries' websites

Croatia, Open Data Portal https://data.gov.hr/

The Netherlands, Open Data Portal https://data.overheid.nl/

Greece, Open Data Portal http://data.gov.gr/

Greece, Geo Data http://geodata.gov.gr/content/about-en/

Annex IV - EU legislation on open data and related legislation

EU legislation relevant for the application of the OD Directive (as it appears in the OD Directive)

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, OJL 172, 26.6.2019 (Open Data Directive) https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information (OJ L 345, 31.12.2003, p. 90).

Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (OJ L 175, 27.6.2013, p. 1).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, p. 1).

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (OJ L 77, 27.3.1996, p. 20).

Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC (OJ L 41, 14.2.2003, p. 26).

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (OJ L 108, 25.4.2007, p. 1).

Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport (OJ L 207, 6.8.2010, p. 1).

Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC (OJ L 94, 28.3.2014, p. 243).

Regulation (EC) No 1370/2007 of the European Parliament and of the Council of 23 October 2007 on public passenger transport services by rail and by road and repealing Council Regulations (EEC) Nos 1191/69 and 1107/70 (OJ L 315, 3.12.2007, p. 1).

Regulation (EC) No 1008/2008 of the European Parliament and of the Council of 24 September 2008 on common rules for the operation of air services in the Community (OJ L 293, 31.10.2008, p. 3).

Council Regulation (EEC) No 3577/92 of 7 December 1992 applying the principle of freedom to provide services to maritime transport within Member States (maritime cabotage) (OJ L 364, 12.12.1992, p. 7).

Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection (OJ L 345, 23.12.2008, p. 75).

Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union (OJ L 194, 19.7.2016, p. 1).

Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC (OJ L 94, 28.3.2014, p. 65).

Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies (OJ L 327, 2.12.2016, p. 1).

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) (OJ L 201, 31.7.2002, p. 37).

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (OJ L 167, 22.6.2001, p. 10).

Regulation (EU) No 182/2011 of the European Parliament and of the Council of 16 February 2011 laying down the rules and general principles concerning mechanisms for control by Member States of the Commission's exercise of implementing powers (OJ L 55, 28.2.2011, p. 13).

Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data (OJ L 8, 12.1.2001, p. 1).

Other

European Commission, Guidelines on recommended standard licences, datasets and charging for the reuse of documents, OJ C 240/1, 247.2014, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014XC0724(01)&from=EN

Annex V - Open data legislation in Croatia

Legislation:

Zakon o pravu na pristup informacijama / Law on the Right of Access to Information, NN 25/13 , NN 85/15

Uredba o troškovima ponovne uporabe informacija / Regulation on the costs of the reuse of information, NN 87/18 /

Pravilnik o sadržaju i načinu vođenje evidencije isključivih prava na ponovnu uporabu informacija / Decree on the content and the manner of keeping the register of exclusive rights on the reuse of information, NN 20/16

Pravilnik o vrstama i sadržaju dozvola kojima se utvrđuju uvjeti ponovne uporabe informacija / Decree on types and content of licences for reuse of public information, NN 67/17

Kriteriji za određivanje visine naknade stvarnih materijalnih troškova i troškova dostave informacije / Criteiria for determining costs, NN 12/14, 15/14

Unofficial translations of legislation is available here: https://www.pristupinfo.hr/pravni-okvir/?lang=en

Official documents, analyses and reports:

Politika otvorenih podataka Vlade RH / Government of the Republic of Croatia, Open data policy, 19.7.2018, available in Croatian: https://rdd.gov.hr/politika-otvorenih-podataka/281

Report to the European Commission on the Availability of Public Sector Information for Re-use and the Conditions under which it is made available and the Redress Practices – Croatia, available here

Ministarstvo uprave: Preporuke o prilagodbi skupova podataka za javnu objavu i ponovno korištenje / Ministry of administration, Recommendations on the datasets preparation for publication and reuse, 2015, available here: http://data.gov.hr/sites/default/files/library/Preporukezaobjavu.pdf

Ministarstvo uprave: Uputa o nužnim zahtjevima kod nabave programskih sustava čiju izradu naručuju tijela / Ministry of Administration, A guideline on necessary requirements for the program systems procurement for public bodies, 3.5.2018; available here (in Croatian)

Povjerenik za informiranje, Smjernice za ponovnu uporabu informacija / Information Commissioner, Guidelines for reuse of information, 2018, available in Croatian

Povjerenik za informiranje, Analitičke studije o otvorenim podacima u Hrvatskoj / Information Commissioner, Analytical studies on open data in Croatia, available in Croatian

Musa, A., et al. Otvoreni podaci za sve. Priručnik o ponovnoj uporabi informacija za tijela javne vlasti. / Open data for all, A handbook for reuse of information for public bodies, 2018, available in Croatian

Key institutional websites:

- Central state office for the development of digital society - Središnji državni ured za razvoj digitalnog društva https://rdd.gov.hr/
- Information Commissioner - Povjerenik za informiranje www.pristupinfo.hr and
- Ministry of Administration - Ministarstvo uprave https://uprava.gov.hr/
- National Infrastructure of spatial data - Nacionalna infrastruktura prostornih podataka http://www.nipp.hr/
- Information systems of environmental protection - Informacijski sustavi zaštite okoliša i prirode http://www.haop.hr/hr/informacijski-sustavi

Portals

- Croatia, Open Data Portal https://data.gov.hr/
- City of Zagreb, Open Data Portal http://data.zagreb.hr/
- City of Rijeka, Open Data Portal http://data.rijeka.hr/
- City of Virovitica, Open Data Portal http://opendata.virovitica.hr/