

# **Навчальний курс по відкритим даним**

Александр Краковецкий

# Навчальний курс по відкритим даним

Александр Краковецкий

© 2016 Александр Краковецкий

# Зміст

Вступ . . . . .	1
Що таке відкриті дані? . . . . .	2
Визначення відкритих даних . . . . .	2
Класифікація відкритих даних . . . . .	2
Публікація та оновлення відкритих даних . . . . .	25
Вибір формату даних та конвертація даних з одного формату в інший . . . . .	25
Основні помилки при оприлюдненні відкритих даних . . . . .	26
Призначення відповідальних за розкриття даних . . . . .	29
Аудит наборів відкритих даних і пріоритетності їх публікації . . . . .	29
Підготовка даних до публікації . . . . .	30
Відкриті дані в Україні . . . . .	40
Законодавство України на тему відкритих даних та готовність країни до впровадження . . . . .	40
Україна в світових рейтингах . . . . .	40
Проекти на базі відкритих даних . . . . .	46
Додатки . . . . .	58
Додаток 1. Рекомендовані інструменти та ресурси по візуалізації даних . . . . .	58
Додаток 2. Глосарій . . . . .	68
Посилання та література . . . . .	71

# Вступ

Мета курсу - ознайомити представників державних структур та представників громадського сектору (активістів, членів громадських організацій, журналістів тощо) з основними поняттями, принципами та підходами щодо роботи з даними з метою їх підготовки та викладення у форматах, що відповідають принципам відкритості та зручності обробки програмними засобами, показати переваги окремих, найбільш популярних, форматів відкритих даних, з огляду на міжнародну практику роботи з даними, а також особливості роботи з відкритими даними в Україні.

Також ці методичні матеріали будуть цікавими представникам ІТ сектору та бізнесу, які мають намір аналізувати публічні дані для працювати з суспільно-корисними ініціативами.

# Що таке відкриті дані?

## Визначення відкритих даних

Відкриті дані (англ. Open Data) – це концепція, яка відображує ідею, що визначені дані мають бути доступні для легкої обробки програмними засобами (machine readable) та подальшого використання і розповсюдження без жодних обмежень і контролю, в тому числі й для комерційного використання. Відкриті дані – це не просто інформація, а концепція, тобто система поглядів, підходів, процесів, які мають одну ідею та мету – вільного використання і розповсюдження даних про діяльність державних органів та органів місцевого самоврядування через мережу Інтернет.

Згідно загальноприйнятої в світі концепції, обов'язково безоплатними можуть бути тільки ті дані, які знаходяться у власності держави («ліцензійно чисті»), і якщо вони подаються в первинному необробленому вигляді. Додаткова обробка або доступ до API частіше за все лімітуються або коштують грошей. Крім того, хто завгодно може (з обов'язковим посиланням на джерело) використати дані в комерційний спосіб, створити на базі них власну програму чи обробити та надати нову цінність (розкласифікувати, встановити зв'язки тощо). Безумовно, всі витрати на додаткову роботу редакторів, програмістів та інше оплачує замовник (або кінцевий користувач).

Відкриті дані дозволяють повторно і необмежено використовувати інформацію, поєднувати її між собою, зменшити або виключити зайві витрати на дублювання та опрацювання великих масивів даних, реєстрів, довідників, баз даних тощо, створених в різних органах влади.

Закон “Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних” №2171 від 19.02.2015 (прийнятий 09.04.2015) визначає наступний термін:

Публічна інформація у формі відкритих даних – це публічна інформація у форматі, що дозволяє її автоматизоване оброблення електронними засобами, вільний та безоплатний доступ до неї, а також її подальше використання.

## Класифікація відкритих даних

В законопроекті одночасно використовуються поняття «форма» та «формат», що не є тотожними. Розглянемо детально.

Для того, щоб зрозуміти, які можуть бути форми відкритих даних, ми також звернемося до відомої класифікації «5 зірок Open Data» (<http://5stardata.info/><sup>1</sup>), де якість даних та рівень відкритості визначається кількістю зірок від 1 до 5, чим більше – тим краще. Відкритість даних залежить від способів доступу, форматів та кількості додаткових дій, які потрібні для отримання кінцевої інформації, її обробки та збереження у власному сховищі або базі даних.

## 5 Star Data Schema

★ Available on the web (whatever format) but with an open license, to be Open Data

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context



5-Star Open Data Scheme  
September 19, 2014



Одну зірку (\*) отримує будь-яка інформація вільно доступна через Інтернет в будь-якому форматі. Під цю класифікацію підпадає файл в форматі PDF або інша (сканована) копія документу, на який веде пряме посилання на офіційному сайті державного органу. Якщо цей файл можна відкрити на власному екрані, прочитати, роздрукувати та отримати звідти потрібну інформацію, то це відкриті дані з однією зіркою.

Дві зірки (\*\*) отримує структурована інформація, яку можна обробляти автоматично, наприклад, в форматах для веб-браузерів чи офісних програм (відкриті формати – TXT, HTML, RSS; пропріетарні формати, Excel – XLS, Word – DOC, RTF). Якщо дані знаходяться в тілі вихідної веб-сторінки, але не мають чіткої структури, містять зайві елементи оформлення, навігації, якщо дані потребують додаткових дій – спеціального розбору (парсингу), то вони вважаються «з двома зірками».

---

<sup>1</sup><http://5stardata.info/>

**Три зірки (\*\*\*)** може отримати інформація, представлена у відомих, добре описаних відкритих структурованих форматах (наприклад, CSV, JSON, XML, YAML) і якщо автоматизована її обробка не потребує від користувача особливих ліцензій та додаткових плат. До відкритих форматів також відносяться пов'язані дані (HTML+RDFa) з узгодженою розміткою елементів в атрибутах (див. сніппет для пошукових систем) або текстові файли таблиць, поля яких розділені табуляцією, комами, крапками з комою або іншими символами.

**Чотири зірки (4)** надаються у випадку, якщо можна отримати первинні необроблені набори відкритих даних у вигляді файлів (довідники, списки, таблиці у відкритому форматі, зліпок бази даних, архів документів тощо) або фільтровані дані у запиті до API за вказаними параметрами. Це дає змогу отримувати тільки потрібну інформацію, актуальну на момент запиту, заощаджує ресурси та час користувача. Безумовно, API має бути описаний так само, як і формати даних, а доступ до нього може бути анонімний без обмежень або з реєстрацією, за вказаним ідентифікатором, лімітами на кількість одночасних запитів тощо.

Останній рівень – **п'ять зірок (5)** – надається інформації, коли набори відкритих даних пов'язані між собою (мають спільні довідники, класифікатори, ідентифікатори, посилання між документами та іншими елементами тощо) і представляють собою семантичну мережу, що постійно оновлюється й змінюється відповідно до сучасних запитів.

## Формати даних

В залежності від специфіки даних, їх розміру та тематики (геологія, тендери, реєстри, судові документи тощо), одні проекти відкритих даних створювались на базі наборів PDF чи DOC файлів, таблиць XLS, що перетворювались на прості текстові таблиці CSV, а інші брали за основу формат розмітки XML, проектували власні схеми XSD і використовували складні структури.

Як свідчить остання статистика використання форматів відкритих даних, найбільш поширений в світі формат (як по кількості, так і по об'єму даних) – PDF. Для українських органів влади, де найбільш розповсюджені операційні системи Microsoft Windows, переважають формати DOC та XLS. Разом з новими версіями офісних програм в Інтернет почали з'являтися документи DOCX та XLSX, рідко ODF (*Open Document Format*). Як надбання DOS'івського періоду, поки ще зустрічаються документи в старому форматі Lexicon (TXT) або ранніх версій Word (RTF). Після поширення ініціативи відкриття державних даних та створення порталів, кількість наборів в форматі XML та інших відкритих форматах почала суттєво збільшуватись.

Необроблені дані, сформовані державними структурами за багато років, можуть бути досить неоднорідними, а деякі набори навіть дублюються в різних форматах для зручності користування.

Серед доступних в цей час форматів відкритих даних, які можна автоматично обробляти електронними засобами, є: CSV (текстові дані, розділені комами або іншими розділовими

символами), JSON (формат, орієнтований на обробку складних структурованих даних за допомогою javascript) та XML (універсальний текстовий формат розмітки). Є багато форматів, що основані на XML, зокрема, HTML, KML та інші.

## Табличні формати

### CSV (Comma-Separated Values)

CSV (від англ. *Comma-Separated Values* – значення, що розділені комами) – текстовий відкритий формат, призначений для представлення таблиць (масивів, наборів) даних, де кожний рядок – це запис таблиці, а значення окремих полів у рядку розділені спеціальними символами (delimiter), зазвичай комами. Щоб завантажити записи таблиці за найменуванням полів (а не за порядковими номерами), додатково потрібно мати опис її структури – назви та формат полів.

Більшість програм широко трактують цей формат і допускають використання інших розділових символів, наприклад, табуляції (TSV) чи коми з крапкою.

Приклад даних в форматі CSV:

- 1 1997,Ford,E350,"ac, abs, moon",3000.00
- 2 1999,Chevy,"Venture ""Extended Edition""",",4900.00
- 3 1996,Jeep,Grand Cherokee,"MUST SELL! air, moon roof, loaded",4799.00

Таблиця в форматі CSV має наступні параметри: кодування (Windows-1251, KOI-8, UTF-8 тощо) і формат кінця рядків. Різниця у символах розділення рядків залежить від операційної системи (для Unix – один символ CR (CHR 0x0D), для Mac – LF (CHR 0x0A), а для Windows – пара символів CR LF).

Інститут відкритих даних Open Data Institute випустив оновлену версію 0.4 додатка Comma Chameleon <https://github.com/theodi/comma-chameleon/releases/tag/0.4.0><sup>2</sup> - з валідації CSV файлів. Ця версія найбільш стабільна з усіх попередніх і існує в версіях для Mac, Linux, Windows і просто як відкритий код в репозиторії <https://github.com/theodi/comma-chameleon><sup>3</sup>.

Корисний інструмент для всіх хто готує дані для публікації і думає про автоматизацію і спрощення очищення даних.

Корисні сервіси та інструменти для роботи з CSV файлами:

1. Валідація CSV файлів: Comma Chameleon <https://github.com/theodi/comma-chameleon><sup>4</sup>.
2. CSVLint <http://csvlint.io/><sup>5</sup> - онлайн сервіс з валідації CSV файлів і з відкритим кодом <https://github.com/theodi/csvlint><sup>6</sup>.

<sup>2</sup><https://github.com/theodi/comma-chameleon/releases/tag/0.4.0>

<sup>3</sup><https://github.com/theodi/comma-chameleon>

<sup>4</sup><https://github.com/theodi/comma-chameleon>

<sup>5</sup><http://csvlint.io/>

<sup>6</sup><https://github.com/theodi/csvlint>



3. CSVkit <https://github.com/wireservice/csvkit><sup>7</sup> - бібліотека для Python за численними маніпуляціям з CSV файлами і безліччю інструментів для командного рядка.
4. Textql <https://github.com/dinedal/textql><sup>8</sup> - інструмент із запуску SQL запитів на CSV / TSV файлах.
5. PapaParse <https://github.com/mholt/PapaParse><sup>9</sup> - парсер дуже великих CSV файлів.
6. <https://mledoze.github.io/countries/><sup>10</sup> - країни світу в JSON, CSV, XML і YAML.
7. Tablib <https://github.com/kennethreitz/tablib><sup>11</sup> - бібліотека для роботи з будь-якими табличними даними включаючи CSV.

## XLS/XLSX (Document Office Open XML)

Файл XLSX - електронна таблиця, створена в Microsoft Excel - додатку для роботи з таблицями. Дані в документі зберігаються в комірках, кожна з яких має певну адресу (колонки нумеруються англійськими літерами, рядки - цифрами, наприклад: A1 - ліва верхня клітинка).

Кожна клітинка може містити як фіксовані дані, так і формули, часто пов'язані з даними в інших осередках. Формат XLSX дозволяє користувачеві змінювати форматування тексту: його шрифт, колір, вирівнювання (в комірці) і інші параметри. Крім того, документ може містити зображення, а також діаграми, побудовані на основі даних в певних комірках.

Програми, які дозволяють працювати з XLSX файлами: File Viewer Lite, Microsoft Excel, Microsoft Excel Viewer, Microsoft Works, Corel WordPerfect Office X7, Apache OpenOffice, Kingsoft Spreadsheets, Nuance OmniPage Ultimate, Gnumeric, LibreOffice.

Спочатку формат XLSX створювався як заміна раніше бінарного формату документів (.xls), який раніше використовували додатки Microsoft Office. Але в 2006 році формат Office Open XML був оголошений вільним і відкритим форматом.

## Графічні формати

### JPEG (Joint Photographic Experts Group)

JPEG (англ. *Joint Photographic Experts Group*, за назвою організації-розробника) - один з популярних графічних форматів, застосовуваний для зберігання фотозображень і подібних до них зображень. Файли даних JPEG, зазвичай мають розширення .jpg, .jfif, .jpe або .jpeg. Однак з них .jpg є найпопулярнішим на всіх платформах.

Алгоритм JPEG дозволяє стискати зображення як з втратами, так і без втрат (режим стиснення lossless JPEG). Підтримуються зображення з лінійним розміром не більше 65535 × 65535 пікселів.

---

<sup>7</sup><https://github.com/wireservice/csvkit>

<sup>8</sup><https://github.com/dinedal/textql>

<sup>9</sup><https://github.com/mholt/PapaParse>

<sup>10</sup><https://mledoze.github.io/countries/>

<sup>11</sup><https://github.com/kennethreitz/tablib>

Алгоритм JPEG найбільшою мірою придатний для стиснення фотографій і картин, що містять реалістичні сцени з плавними переходами яскравості і кольору. Найбільшого поширення JPEG отримав в цифровій фотографії і для зберігання і передачі зображень з використанням мережі Інтернет.

Формат JPEG в режимі стиснення з втратами малоприсадаблений для стиснення креслень, текстової та знакової графіки, де різкий контраст між сусідніми пікселями приводить до появи помітних артефактів. Такі зображення доцільно зберігати в форматах без втрат, таких як JPEG-LS, TIFF, GIF, PNG або використовувати режим стиснення Lossless JPEG.

JPEG не підходить для стиснення зображень багатоетапної обробки, так як спотворення в зображення будуть вноситися кожен раз при збереженні проміжних результатів обробки. JPEG не повинен використовуватися і в тих випадках, коли недопустимі навіть мінімальні втрати, наприклад, при стисненні астрономічних або медичних зображень.

Детальніше про формат: <https://ru.wikipedia.org/wiki/JPEG><sup>12</sup>

## PNG (portable network graphics)

PNG (англ. *Portable network graphics*) - растровий формат зберігання графічної інформації, що використовує стиснення без втрат за алгоритмом Deflate. Формат PNG розроблений для заміни форматів GIF та TIFF. Формат PNG позиціонується передусім для використання в Інтернеті і редагування графіки. Формат PNG зберігає графічну інформацію в стислому вигляді, причому це стиснення графічної інформації проводиться без втрат.

PNG підтримує три основних типи растрових зображень:

- Півтонування (з глибиною кольору 16 біт)
- Кольорові індексовані зображення (палітра 8 біт для кольору глибиною 24 біт)
- Повнокольорове зображення (з глибиною кольору 48 біт)

Формат має наступні основні переваги:

- практично необмежену кількість кольорів в зображенні;
- опціональна підтримка альфа-каналу;
- можливість гамма-корекції;
- двовимірний безстроковий розгортка;
- можливість розширення формату для користувача блоками (на цьому заснований, зокрема, APNG).
- будь-яке збережене зображення PNG може бути прочитано в будь-якому іншому додатку, що підтримує PNG.

---

<sup>12</sup><https://ru.wikipedia.org/wiki/JPEG>

PNG є хорошим форматом для редагування зображень, навіть для зберігання проміжних стадій редагування, так як відновлення і повторне збереження зображення здійснюється без втрат в якості.

Детальніше про формат: <https://ru.wikipedia.org/wiki/PNG><sup>13</sup>

## TIFF (Tagged Image File Format)

TIFF (англ. *Tagged Image File Format*) - формат зберігання растрових графічних зображень. TIFF став популярним форматом для зберігання зображень з великою глибиною кольору. Він використовується при скануванні, відправленні факсів, розпізнаванні тексту, в поліграфії, широко підтримується графічними додатками. TIFF був обраний в якості основного графічного формату операційної системи NeXTSTEP і з неї підтримка цього формату перейшла в Mac OS X.

Спочатку формат підтримував стиснення без втрат, згодом формат був доповнений для підтримки стиснення з втратами в форматі JPEG. Файли формату TIFF, як правило, мають розширення .tiff або .tif.

Структура формату гнучка і дозволяє зберігати зображення в режимі кольорів з палітрою, а також в різних колірних просторах:

- Бінарному (двокольоровому, іноді неправильно званому чорно-білим)
- напівтоновому
- з індексованої палітрою
- RGB
- CMYK
- YCbCr
- CIE Lab

Підтримуються режими 8, 16, 32 і 64 біт на канал при цілочисельному, а також 32 і 64 біт на канал при поданні значення пікселя числами з плаваючою комою.

TIFF є тяговим форматом і в ньому є такі види міток:

- основні мітки;
- розширені мітки;
- спеціальні мітки.

Детальніше про формат: <https://ru.wikipedia.org/wiki/TIFF><sup>14</sup>

---

<sup>13</sup><https://ru.wikipedia.org/wiki/PNG>

<sup>14</sup><https://ru.wikipedia.org/wiki/TIFF>

## Текстові формати

### TXT (Textfile)

TXT - це формат, що містить текстові дані, які, як правило, організовані в вигляді рядків; служить основою для багатьох більш спеціалізованих форматів, таких як CHM, PHP, XML, CSV. Може бути переглянутий будь-яким текстовим редактором.

Текстові файли розбиваються на кілька рядків. На сучасних платформах розбивка на рядки кодується керуючим символом переведення рядка, а іноді послідовністю двох керуючих символів. Взагалі, текстові файли можуть містити друковані символи, такі як букви, цифри і розділові знаки і деяку кількість керуючих символів, таких як знаки табуляції і переведення рядка.

Файли TXT містять дуже мало елементів форматування, однак дозволяють співвіднести прийняті набори елементів форматування з системним терміналом або простим текстовим редактором. Файли TXT універсальні за своєю природою, тому що відкрити такі файли в стані будь-який текстовий редактор. При цьому файли TXT можуть використовувати Unicode, яка дозволяє полегшити користувачам використання файлів TXT, які пишуть на різних мовах. Файли TXT, що використовують текст тільки в кодуванні ASCII, можна переносити на різні комп'ютери і пристрої на ОС Unix, Mac і Windows.

### Markdown

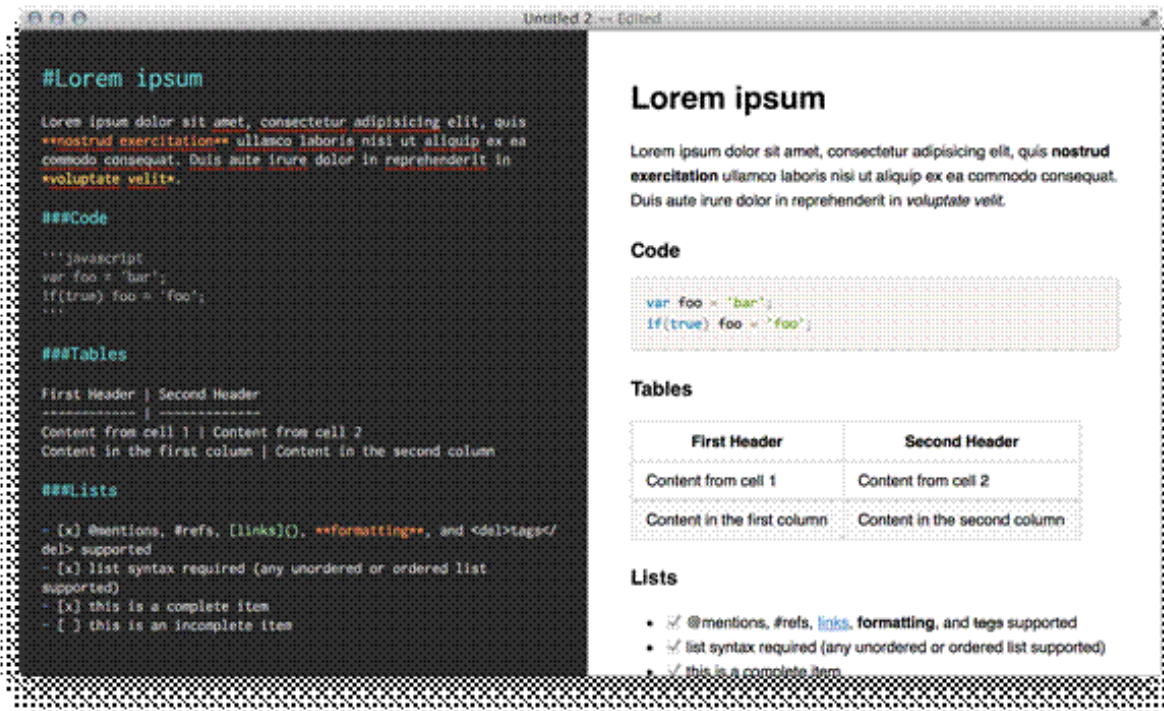
Markdown - полегшена мова розмітки, створена з метою написання максимально читабельного і зручного для редагування тексту, але придатного для перетворення в мови для публікацій (HTML, Rich Text і ін.).

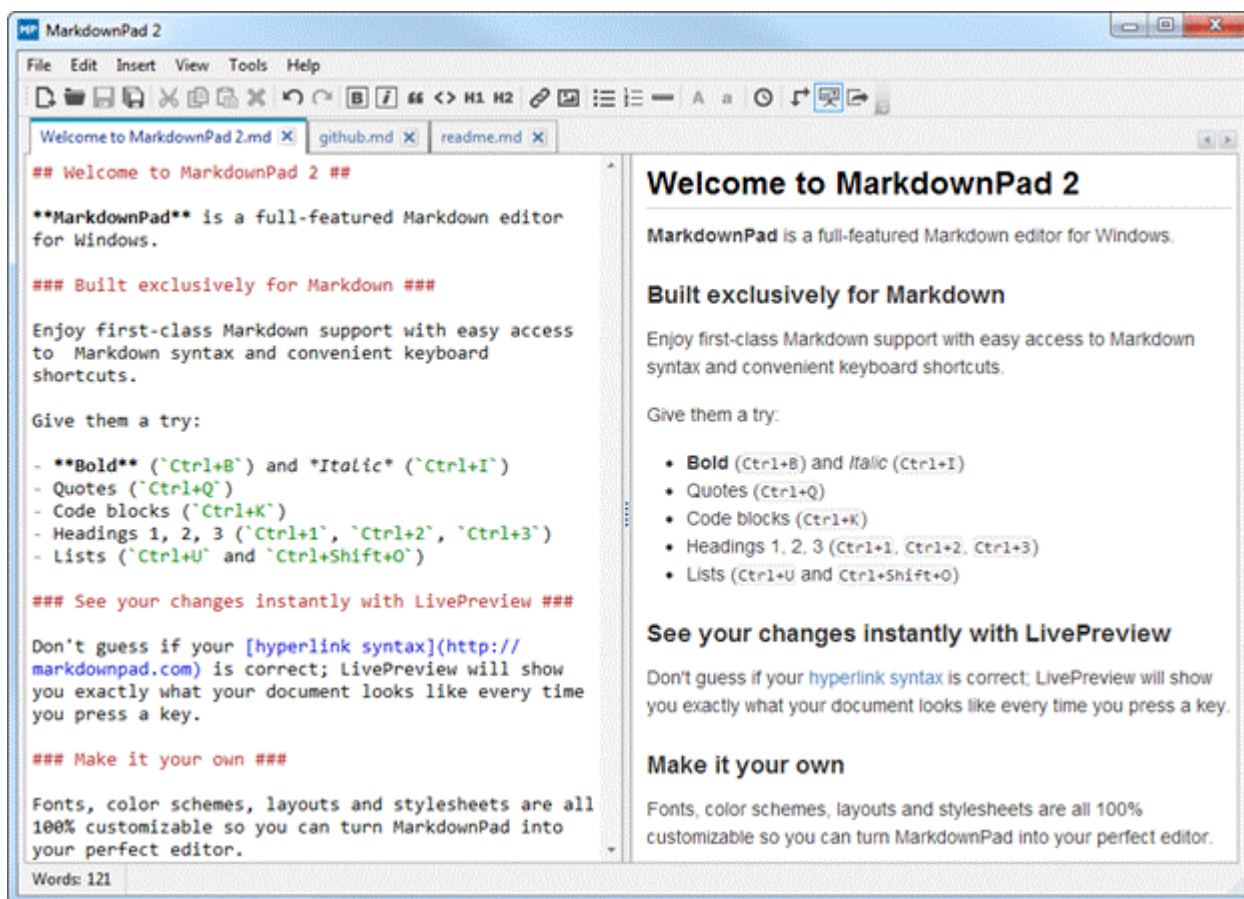
Таким чином, "Markdown" це дві речі:

- простий синтаксис форматування тексту;
- програмний інструмент, написаний на Perl, який перетворює звичайний форматування тексту в HTML.

Першочерговою метою розробки для форматування синтаксису Markdown є зробити його максимально читабельним, наскільки це можливо. Ідея полягає в тому, що Markdown-форматований документ повинен бути опублікований як звичайний текст, використовуючи теги та інструкції форматування.

Приклади форматування тексту мовою розмітки Markdown:





Детальніше про формат та його синтаксис:

1. Markdown <http://daringfireball.net/projects/markdown/><sup>15</sup>
2. Wikipedia <https://uk.wikipedia.org/wiki/Markdown><sup>16</sup>

## Текстово-графічні формати

### HTML (HyperText Markup Language)

HTML (від англ. *HyperText Markup Language* - «мова гіпертекстової розмітки») - стандартна мова розмітки документів у Всесвітній павутині. Більшість веб-сторінок містять опис розмітки на мові HTML (або XHTML). Мова HTML інтерпретується браузерами; отриманий в результаті інтерпретації форматований текст відображається на екрані монітора комп'ютера або мобільного пристрою.

У всесвітній павутині HTML-сторінки, як правило, передаються браузерам від сервера по протоколам HTTP або HTTPS, у вигляді простого тексту або з використанням шифрування.

<sup>15</sup><http://daringfireball.net/projects/markdown/>

<sup>16</sup><https://uk.wikipedia.org/wiki/Markdown>

Текстові документи, що містять розмітку на мові HTML (такі документи зазвичай мають розширення .html або .htm), обробляються спеціальними додатками, які відображають документ в його форматованому вигляді. Такі додатки, звані «браузерами», зазвичай надають користувачеві зручний інтерфейс для запиту веб-сторінок, їх перегляду (і виведення на інші зовнішні пристрої) і, при необхідності, відправки введених користувачем даних на сервер.

HTML - тегова мова розмітки документів. Будь-який документ на мові HTML являє собою набір елементів, причому початок і кінець кожного елемента позначається спеціальними позначками - тегами. Елементи можуть бути порожніми, тобто не містять ніякого тексту та інших даних (наприклад, тег переносу рядка <br>). В цьому випадку зазвичай не вказується закриваючий тег. Крім того, елементи можуть мати атрибути, що визначають будь-які їх властивості (наприклад, розмір шрифту для елемента font). Атрибути вказуються в відкриваючому тегові.

Детальніше про формат: <https://ru.wikipedia.org/wiki/HTML><sup>17</sup>

## DOCX (Document Office Open XML)

DOCX (*Document Office Open XML*) - формат файлу для зберігання електронних документів пакетів офісних додатків - зокрема, Microsoft Office. Формат є zip-архів, що містить текст у вигляді XML, графіку і інших даних, які можуть бути переведені в послідовність бітів (серіалізовані) із застосуванням захищених патентами довічних форматів, специфікації яких були опубліковані Microsoft для користувачів на умовах Microsoft Open Специфікація Promise (англ.).

Спочатку формат створювався як заміна раніше бінарного формату документів, який використовували додатки Microsoft Office аж до версії Office 2003 включно. У 2006 році формат Office Open XML був оголошений вільним і відкритим форматом Ecma International. Він є форматом за умовчанням для додатків Microsoft Office 2007 і пізніших.

Файл з розширенням .DOCX - документ, створений Microsoft Word, програмою обробки тексту; містить текст документа, зображення, форматування, стилі, намальовані об'єкти та інші параметри документа.

Файли DOCX створюються за допомогою відкритого формату XML, в якому зберігаються документи, як збори окремих файлів і папок в стислому пакеті. DOCX-файли містять XML-файли і три папки, docProps, Word, і \_rels, які містять властивості документа, зміст і відносини між файлами.

Детальніше про формат: [https://ru.wikipedia.org/wiki/Office\\_Open\\_XML](https://ru.wikipedia.org/wiki/Office_Open_XML)<sup>18</sup>

## OpenDocument format (LibreOffice/OpenOffice)

Формат OpenDocument - це формат для текстових, табличних документів та презентацій, що був створений з метою підвищити інтероперабельність між різними офісними застосунками

---

<sup>17</sup><https://ru.wikipedia.org/wiki/HTML>

<sup>18</sup>[https://ru.wikipedia.org/wiki/Office\\_Open\\_XML](https://ru.wikipedia.org/wiki/Office_Open_XML)



(програмами, офісними пакетами).

Історична довідка: процес стандартизації форматів офісних документів почався на початку 2000-х років та був обумовлений становищем на ринку офісних пакетів програм. Microsoft на той момент займала домінантне становище. Європейська комісія за скаргою Sun Microsystems, що тоді розробляла альтернативний офісний пакет OpenOffice.org, вирішила, що Microsoft зловживає своїм становищем та веде нечесну конкуренцію. Microsoft змушена була відкрити специфікацію своїх форматів. Паралельно з цим, відбувалась стандартизація офісних форматів. Результатом зусиль зі стандартизації стало створення двох конкуруючих форматів - Office Open XML (Microsoft) та OpenDocument (Sun Microsystems).

Файли формату OpenDocument мають розширення ODT (OpenDocument Text), ODS (OpenDocument Spreadsheet). Його підтримують наступні пакети програм:

- Apache OpenOffice
- LibreOffice
- MS Office (2013 - тільки читання, 2016 - читання та редагування)
- Google Docs (тільки імпорт/експорт)

## PDF (Portable Document Format)

Формат переносного документа (PDF) - це формат файлу, який використовується для надійного уявлення і обміну документами, незалежно від програмного і апаратного забезпечення або операційної системи. В даний час формат PDF, винайдений компанією Adobe, є відкритим стандартом, підтримуваним Міжнародною організацією зі стандартизації (ISO). Файли PDF можуть містити посилання і кнопки, поля форм, аудіо- та відеоелементи, а також бізнеслогіку. Вони також підтримують можливість електронного підпису і можуть переглядатися за допомогою безкоштовного програмного забезпечення Acrobat Reader DC.

Можливості формату:

- мають такий же вигляд, як вихідні документи, незалежно від використовуваної платформи або пристрою;
- відповідають стандартам ISO 32000 з обміну електронними документами, включаючи спеціальні стандарти для PDF;
- можуть бути підписані електронно будь-яким користувачем за допомогою безкоштовного програмного забезпечення Adobe Acrobat Reader DC або мобільного застосування Acrobat DC;
- можуть захищатися паролем для запобігання копіювання та редагування;
- можуть виправлятися для видалення конфіденційної інформації без можливості відновлення;
- зберігають всю інформацію вихідного файлу, включаючи текст, зображення, аудіо, 3D-карти і багато іншого, об'єднану в одному файлі;



- забезпечують зручний пошук, включаючи пошук по сканованому тексту, який був перетворений за допомогою технології оптичного розпізнавання Символів (OCR);
- працюють з допоміжними технологіями, які роблять файли PDF доступнішими для людей з обмеженими можливостями, такими як слабкий зір або сліпота.

Детальніше про формат: [https://ru.wikipedia.org/wiki/Portable\\_Document\\_Format](https://ru.wikipedia.org/wiki/Portable_Document_Format)<sup>19</sup>  
та <https://acrobat.adobe.com/ru/ru/products/about-adobe-pdf.html><sup>20</sup>

## Формати представлення даних через API

### JSON (JavaScript Object Notation)

JSON (від англ. *JavaScript Object Notation*) – текстовий відкритий формат, оснований на Javascript представлені та призначений для обміну даними в Інтернет між сервером та клієнтом або сервером і сервером. Хоча він позиціонується, як незалежний від системи і мови програмування, частіше за все використовується за допомогою програм на Javascript, але як і інші текстові формати, легко читається людиною.

Найбільшу популярність JSON набув після створення інтерактивних веб-сторінок, дані до яких через API передавались під час взаємодії користувача з елементами інтерфейсу (т.з. технологія AJAX). За рахунок своєї лаконічності, на відміну від XML, простоті й швидкості використання саме в програмах на Javascript, широкими можливостями в серіалізації даних – рекурсивного перетворення в текстовий вигляд складних об'єктів, формат активно використовується для формування «на льоту» та передачі структур даних в Інтернет в різних інформаційних системах і сервісах.

Синтаксис JSON (див. <https://ru.wikipedia.org/wiki/JSON><sup>21</sup>) представляє собою текст у вигляді програмного коду Javascript, який описує кілька варіантів структур:

- 1) масив в квадратних дужках (значення через кому),
- 2) іменованний масив (чи об'єкт) у фігурних дужках (пари ключ-значення, розділені двома крапками, через кому),
- 3) просте значення (число, рядок, true, false, null),
- 4) функцію.

Для перетворення формату JSON в об'єкт в пам'яті (десеріалізації) в Javascript використовується функція `eval()`, яка виконує завантажений в текстовий рядок програмний код.

Якщо для обробки використовувати тільки мову Javascript, виникають як мінімум дві проблеми. По-перше, щоб обробити дані, потрібно мати приблизно в два рази більше вільної пам'яті, ніж файл (спочатку завантажуються повний текст JSON, а після його виконання, дані об'єкту завантажуються повторно у вигляді відповідних структур). По-друге, можливість

<sup>19</sup>[https://ru.wikipedia.org/wiki/Portable\\_Document\\_Format](https://ru.wikipedia.org/wiki/Portable_Document_Format)

<sup>20</sup><https://acrobat.adobe.com/ru/ru/products/about-adobe-pdf.html>

<sup>21</sup><https://ru.wikipedia.org/wiki/JSON>

передачі програмного коду, який автоматично стартує, дозволяє зловмисникам отримати несанкціонований доступ до системи. Цю вразливість також можна уникнути, якщо спочатку перевірити і очистити текст JSON від заборонених об'єктів, а дані перетворити на коректні значення.

Через нераціональне використання пам'яті і ресурсів в JSON краще працювати зі об'єктами будь-якою складністю, але не дуже великих розмірів, в швидкому режимі «запит-відповідь».

Приклад формату JSON:

```
1 {  
2   "firstName": "Иван",  
3   "lastName": "Иванов",  
4   "address": {  
5     "streetAddress": "Московское ш., 101, кв.101",  
6     "city": "Ленинград",  
7     "postalCode": 101101  
8   },  
9   "phoneNumbers": [  
10    "812 123-1234",  
11    "916 123-4567"]  
12 }
```

## XML (eXtensible Markup Language)

XML (від англ. *eXtensible Markup Language* – мова розмітки, що розширюється) – мабуть найстаріший текстовий відкритий формат, створений в 1994 році та рекомендований Консорціумом Всесвітньої павутини (W3C), як основний для обміну інформацією в Інтернет. Гіпертекстова розмітка (HTML) – це один з різновидів XML. Разом з таблицями каскадних стилів CSS, які формують зовнішній вигляд документів, вони є тими основними форматами, що обумовлюють розвиток технологій.

Насправді XML (див. <https://ru.wikipedia.org/wiki/XML><sup>22</sup>) – універсальний зручний для людини формат для збереження, передачі структурованих даних і їх автоматичної обробки у формі машиночитних документів. З нього починали розвиток інші відкриті формати, на ньому вдосконалювались підходи обміну даними. Сама ж мова розмітки XML є похідною від ще старішого та більш складного формату SGML (Standard Generalized Markup Language – стандартна узагальнена мова розмітки), стандартизованого за ISO 8879:1986 Information processing—Text and office systems—Standard Generalized Markup Language (SGML).

Перевагами XML є простота та гнучкість розмітки, яка не вимагає формальних, фіксованих назв тегів чи параметрів, і будь-який розробник може доповнювати та змінювати формат, створювати власну схему XSD (XML Schema Definition). Фактично, це мова, яка описує

---

<sup>22</sup><https://ru.wikipedia.org/wiki/XML>

сама себе і будь-які за розміром і складністю структури даних. Безумовно, в цей формат можна конвертувати інші формати (наприклад, XLS). Крім того, формат не залежить від операційної системи чи бази даних. Для простого перегляду чи редагування достатньо текстового редактора. Але є й недоліки – дані в форматі XML займають значно більше місце, ніж це потрібно, за рахунок повторення тегів та відступів, а парсинг значно складніший, ніж CSV чи JSON.

За довгий час існування XML на його базі було розроблено багато форматів і стандартів зі схожим синтаксисом (див. приклади стандартних схем на <http://schema.org><sup>23</sup>). Зазвичай цю групу форматів називають загальною назвою – XML, тому що вони мають єдині механізми опису схем XSD, перевірки правильності даних (валідації), перегляд DOM (Document Object Model), доступу до елементів XPath та трансформації для автоматичного конвертування у інші схеми чи формати (наприклад, альтернативні JSON та YAML) за допомогою мови перетворення XSLT (*eXtensible Stylesheet Language Transformations*).

Використання формату XML (а саме LegalXML) у якості відкритого стандарту нормативно-правового документа – це сучасний спосіб забезпечити обмін інформацією (документами, картками, довідниками тощо) між інформаційними системами або в межах однієї системи при підготовці (опрацюванні) документів.

В одному файлі XML в текстовому вигляді, крім основних даних та тексту електронного нормативного документа, можна розміщати метадані (характеристики, реквізити, опис, класифікацію тощо), вкладені файли (картинки, стилі тексту, таблиці Excel, документи Word і т.і.), необхідні структури чи довідники. Це дозволяє зручно не тільки зберігати, передавати, обробляти документ, отримувати PDF версію для друку, формувати зміст чи робити посилання на конкретну главу, статтю, пункт, підпункт тощо, а й автоматизовано вносити зміни, підготовлені у вигляді, що дозволяє їх програмну обробку.

Приклад даних у форматі XML:

```
1 <person firstName="Иван" lastName="Иванов">
2   <address streetAddress="Московское ш., 101, кв.101"
3       city="Ленинград" postalCode="101101"/>
4   <phoneNumbers>
5       <phoneNumber>812 123-1234</phoneNumber>
6       <phoneNumber>916 123-4567</phoneNumber>
7   </phoneNumbers>
8 </person>
```

## RDF (Resource Description Framework)

RDF - це розроблена консорціумом Всесвітньої павутини модель для представлення даних, особливо - метаданих. RDF представляє твердження про ресурсах у вигляді, придатному для машинної обробки. RDF є частиною концепції семантичної павутини.

---

<sup>23</sup><http://schema.org>

Ресурсом в RDF може бути будь-яка сутність - як інформаційна (наприклад, веб-сайт або зображення), так і неінформаційна (наприклад, людина, місто або якесь абстрактне поняття). Твердження, що висловлюється про ресурс, має вигляд «суб'єкт - предикат - об'єкт» і називається триплетом. Затвердження «небо блакитного кольору» в RDF-термінології можна представити таким чином: суб'єкт - «небо», предикат - «має колір», об'єкт - «блакитний». Для позначення суб'єктів, відносин і об'єктів в RDF використовуються URI.

RDF сам по собі є не форматом файлу, а тільки лише абстрактною моделлю даних, тобто описує пропоновану структуру, способи обробки та інтерпретації даних. Для зберігання і передачі інформації, покладеної в модель RDF, існує цілий ряд форматів запису.

Для обробки RDF-даних пропонується реалізувати мови запитів: SPARQL (стандарт W3C), RQL, RDQL.

Для запису і передачі RDF використовується кілька форматів, в тому числі:

- RDF / XML - запис у вигляді XML-документа;
- RDF / JSON - запис у вигляді JSON-даних;
- RDFa - запис всередині атрибутів довільного HTML- або XHTML-документа;
- N-Triples, Turtle, N3 - компактні форми запису тверджень.

## Формати даних для роботи з геопросторовими даними

### Формати GIS

Файли геопросторових даних можуть містити таку інформацію:

- Просторова або геометрична інформація, яка забезпечує положення і форми конкретних географічних об'єктів.
- Інформація про атрибути, яка забезпечує описову інформацію (числову, текстову, логічну) про кожний об'єкт.
- Інформація про відображення, яке може описувати спосіб оформлення об'єкту.

Деякі цифрові карти не містять всі три типи інформації. Наприклад, растрові карти, як правило, не включають в себе інформацію про атрибути. В свою чергу багато джерел векторних даних не включають інформацію про оформлення.

### Просторова або геометрична інформація

Географічна інформація описує у цифровій формі положення та форму кожного зображуваного об'єкта. На векторній карті, властива об'єкту позиція зазвичай виражається у вигляді наборів координат x, y або x, y, z з використанням системи координат, визначеній для набору геоданих. Більшість векторних географічних інформаційних систем підтримують три основних геометричні об'єкти: \* Точка: одна пара координат, 0-вимірний об'єкт. \*

Лінія: дві або більше точок у певній послідовності, одновимірний об'єкт. \* Полігон: площа, обмежена лінією, двовимірний об'єкт. \* Тіло, об'єм обмежений гранями, тривимірний об'єкт.

Векторні дані характеризуються топологією, яка задає правила просторової взаємодії між векторними об'єктами. В залежності від масштабу даних один і той же об'єкт може зберігатися у вигляді точки, лінії або полігону.

На растровій карті властива об'єкту позиція визначається комірками з відповідними атрибутами. Кожна комірка характеризується розміром та значенням атрибутів, яке може бути відображене у вигляді значень кольорової моделі (спутниковий знімок), або значень атрибуту (GRID), на основі яких растр може бути класифікований за кількісними або якісними характеристиками.

Деякі системи також підтримують більш складні об'єкти, такі як регіони, кола, еліпси, дуги і криві.

### **Інформація про атрибути**

Атрибути описують специфічні характеристики об'єктів, які не є графічними елементами. Для прикладу, атрибутом, що характеризує дорогу може бути її назва або дата, коли її побудували. Зазвичай, атрибути зберігаються окремо від графічної частини карти за допомогою геореляційних структур даних. Атрибути є частиною опису векторних карт і рідко застосовуються разом із растровими зображеннями.

### **Інформація про відображення**

Інформацію, що відображається може описувати цифрова карта або накреслена (план). Загальна інформація на карті відображається за допомогою різних кольорів: озера, моря, вершини, гори, рівнин; ширини і типів ліній: назви доріг, вулиць, нанесення висотної поясності на планах тощо.

Інформація про відображення задає графічні стилі, за допомогою яких відображаються дані на карті. Основними графічними стилями є графічні перемінні, до яких відноситься колір, товщина, форма, орієнтація, символізація, підписи. Разом графічні перемінні утворюють способи умовних позначень на карті географічних об'єктів та формують легенду карти.

**GIS-формати** створені для стандартного кодування географічної інформації в файл. Файли цього формату створюються державними агентствами з картографування, комерційними організаціями або розробниками програмного забезпечення.

Ці формати є розповсюдженими, але існує програмне забезпечення, що не підтримує файли цього формату. В таких випадках шукають способи перетворення GIS-файлу або використовують інший файл. Майже кожен GIS має свій власний внутрішній формат файлу. Ці формати призначені для оптимального використання розробником і часто є запатентованими. Вони не призначені для використання поза системою, що їх розробляла.

### **Векторні формати**

Багато додатків GIS засновані на векторній технології через її широке застосування. Вони є складними, тому що інформацію можливо описати безліччю способів: для зберігання

координат, атрибутів, зв'язків, структур баз даних та відображення інформації. Найбільш поширеним форматом гепросторових даних є формат [Shapefile](https://en.wikipedia.org/wiki/Shapefile)<sup>24</sup> (shp). Цей формат розроблений компанією ESRI та є пропрієтарним форматом даних. Проте сьогодні він став дефакто стандартом обміну та публікації даних. Більшість програмних продуктів працюють з цим форматом, в тому числі ПЗ QGIS, яке є найбільш популярним OpenSource додатком для ГІС. Специфікація формату визначає набір файлів із загальним префіксом імені файлу, які зберігаються в тому ж каталозі. Три обов'язкових файли мають розширення файлів shp, .SHX і .dbf.

.shp - головний файл; містить набори геометричних об'єктів  
.shx - індексний файл; який використовується для зв'язку між файлами .dbf і .shp  
.dbf - атрибутивний файл; містить атрибути об'єктів, описаних в .shp файлі у форматі dBase IV.

На сучасному етапі більшість геоданих зберігаються в СУБД, які підтримують збереження просторових даних. Зараз до них відносяться Oracle, SQLserver, Postgres та ін. ПЗ ArcGis підтримує два типи баз даних: персональна та геобаз даних з відповідними форматами: mdb, gdt.

Значного поширення набули формати ГІС: [MapInfo TAB format](https://en.wikipedia.org/wiki/MapInfo_TAB_format)<sup>25</sup>, [Spatialite](https://en.wikipedia.org/wiki/Spatialite)<sup>26</sup>, [Geography Markup Language \(GML\)](https://en.wikipedia.org/wiki/Geography_Markup_Language)<sup>27</sup>. Також розповсюдженими в Україні є формати даних ПЗ Панорама – SXF, Digital – підтримка формату in4, який довгий час був обмінним файлом кадастрової інформації.

Файли AutoCAD не є ГІС форматом, проте сучасні професійні ГІС підтримують пряму конвертацію з формату DXF та навпаки. Доволі рідкими форматами в Україні є формати [National Transfer Format](https://en.wikipedia.org/wiki/National_Transfer_Format)<sup>28</sup> (NTF), [SOSI](https://en.wikipedia.org/wiki/SOSI)<sup>29</sup>, [Spatial Data File](https://en.wikipedia.org/wiki/Spatial_Data_File)<sup>30</sup>, [TIGER](https://en.wikipedia.org/wiki/TIGER)<sup>31</sup>, [Vector Product Format](https://en.wikipedia.org/wiki/Vector_Product_Format)<sup>32</sup> (VPF).

Для обміну просторовими даними також використовують формат ASCII, який у більшості випадків використовують для збереження растрових даних.

Для прикладу наведено найпоширеніші формати: [AutoCAD DXF](https://en.wikipedia.org/wiki/AutoCAD_DXF)<sup>33</sup>, [Cartesian coordinate system](https://en.wikipedia.org/wiki/Cartesian_coordinate_system)<sup>34</sup> (XYZ), [Digital Line Graph](https://en.wikipedia.org/wiki/Digital_Line_Graph)<sup>35</sup> (DLG), [Esri TIN](https://en.wikipedia.org/wiki/Esri_TIN)<sup>36</sup>, [Geography Markup Language](https://en.wikipedia.org/wiki/Geography_Markup_Language)<sup>37</sup> (GML),

---

<sup>24</sup><https://en.wikipedia.org/wiki/Shapefile>

<sup>25</sup>[https://en.wikipedia.org/wiki/MapInfo\\_TAB\\_format](https://en.wikipedia.org/wiki/MapInfo_TAB_format)

<sup>26</sup><https://en.wikipedia.org/wiki/Spatialite>

<sup>27</sup>[https://en.wikipedia.org/wiki/Geography\\_Markup\\_Language](https://en.wikipedia.org/wiki/Geography_Markup_Language)

<sup>28</sup>[https://en.wikipedia.org/wiki/National\\_Transfer\\_Format](https://en.wikipedia.org/wiki/National_Transfer_Format)

<sup>29</sup><https://en.wikipedia.org/wiki/SOSI>

<sup>30</sup>[https://en.wikipedia.org/wiki/Spatial\\_Data\\_File](https://en.wikipedia.org/wiki/Spatial_Data_File)

<sup>31</sup><https://en.wikipedia.org/wiki/TIGER>

<sup>32</sup>[https://en.wikipedia.org/wiki/Vector\\_Product\\_Format](https://en.wikipedia.org/wiki/Vector_Product_Format)

<sup>33</sup>[https://en.wikipedia.org/wiki/AutoCAD\\_DXF](https://en.wikipedia.org/wiki/AutoCAD_DXF)

<sup>34</sup>[https://en.wikipedia.org/wiki/Cartesian\\_coordinate\\_system](https://en.wikipedia.org/wiki/Cartesian_coordinate_system)

<sup>35</sup>[https://en.wikipedia.org/wiki/Digital\\_line\\_graph](https://en.wikipedia.org/wiki/Digital_line_graph)

<sup>36</sup>[https://en.wikipedia.org/wiki/Esri\\_TIN](https://en.wikipedia.org/wiki/Esri_TIN)

<sup>37</sup>[https://en.wikipedia.org/wiki/Geography\\_Markup\\_Language](https://en.wikipedia.org/wiki/Geography_Markup_Language)

GeoJSON<sup>38</sup>, GeoMedia<sup>39</sup>, ISFC, Keyhole Markup Language<sup>40</sup> (KML), MapInfo TAB format<sup>41</sup>, National Transfer Format<sup>42</sup> (NTF), Spatialite<sup>43</sup>, Shapefile<sup>44</sup>, Simple Features<sup>45</sup>, SOSI<sup>46</sup>, Spatial Data File<sup>47</sup>, TIGER<sup>48</sup>, Vector Product Format<sup>49</sup> (VPF).

## Растрові формати

Растрові файли зазвичай використовуються для зберігання інформації зображення у вигляді сканованих паперових карт чи аерофотозйомки. Вони також використовуються для передачі даних дистанційного зондування Землі. На відміну від інших растрових файлів, які характеризуються розрізненням у розмірах точок на дюйм, розрізнення на зображеннях дистанційного зондування виражається в метрах на піксел растру, що вказує розмір земельної ділянки певної області. Прикладами растрових форматів є: ADRG, RPF, CADRG, CIB<sup>50</sup>, Digital raster graphic<sup>51</sup> (DRG), ECRG, ECW<sup>52</sup>, Esri grid<sup>53</sup>, GeoTIFF<sup>54</sup>, IMG – ERDAS IMAGINE<sup>55</sup>, JPEG2000<sup>56</sup>, MrSID<sup>57</sup>, netCDF<sup>58</sup>-CF.

Детальніше про формати та GIS-формат: [https://en.wikipedia.org/wiki/GIS\\_file\\_formats](https://en.wikipedia.org/wiki/GIS_file_formats)<sup>59</sup>

## KML - геоінформаційний формат Google Maps

KML (від англ. *Keyhole Markup Language*) - мова розмітки на основі XML для представлення тривимірних геопросторових даних в програмі «Google Earth».

Підмножина мови KML 2.0 може використовуватися і для відображення двовимірних карт в сервісі «Карти Google».

KML-файли зазвичай поширюються в ZIP-архіві: KMZ

Зміст файлів KML визначає один або кілька об'єктів для відображення в Google Earth. Цими об'єктами можуть бути:

---

<sup>38</sup><https://en.wikipedia.org/wiki/GeoJSON>

<sup>39</sup><https://en.wikipedia.org/wiki/GeoMedia>

<sup>40</sup>[https://en.wikipedia.org/wiki/Keyhole\\_Markup\\_Language](https://en.wikipedia.org/wiki/Keyhole_Markup_Language)

<sup>41</sup>[https://en.wikipedia.org/wiki/MapInfo\\_TAB\\_format](https://en.wikipedia.org/wiki/MapInfo_TAB_format)

<sup>42</sup>[https://en.wikipedia.org/wiki/National\\_Transfer\\_Format](https://en.wikipedia.org/wiki/National_Transfer_Format)

<sup>43</sup><https://en.wikipedia.org/wiki/Spatialite>

<sup>44</sup><https://en.wikipedia.org/wiki/Shapefile>

<sup>45</sup>[https://en.wikipedia.org/wiki/Simple\\_Features](https://en.wikipedia.org/wiki/Simple_Features)

<sup>46</sup><https://en.wikipedia.org/wiki/SOSI>

<sup>47</sup>[https://en.wikipedia.org/wiki/Spatial\\_Data\\_File](https://en.wikipedia.org/wiki/Spatial_Data_File)

<sup>48</sup><https://en.wikipedia.org/wiki/TIGER>

<sup>49</sup>[https://en.wikipedia.org/wiki/Vector\\_Product\\_Format](https://en.wikipedia.org/wiki/Vector_Product_Format)

<sup>50</sup>[https://en.wikipedia.org/wiki/Controlled\\_Image\\_Base](https://en.wikipedia.org/wiki/Controlled_Image_Base)

<sup>51</sup>[https://en.wikipedia.org/wiki/Digital\\_raster\\_graphic](https://en.wikipedia.org/wiki/Digital_raster_graphic)

<sup>52</sup>[https://en.wikipedia.org/wiki/ECW\\_\(file\\_format\)](https://en.wikipedia.org/wiki/ECW_(file_format))

<sup>53</sup>[https://en.wikipedia.org/wiki/Esri\\_grid](https://en.wikipedia.org/wiki/Esri_grid)

<sup>54</sup><https://en.wikipedia.org/wiki/GeoTIFF>

<sup>55</sup>[https://en.wikipedia.org/wiki/ERDAS\\_IMAGINE](https://en.wikipedia.org/wiki/ERDAS_IMAGINE)

<sup>56</sup><https://en.wikipedia.org/wiki/JPEG2000>

<sup>57</sup><https://en.wikipedia.org/wiki/MrSID>

<sup>58</sup><https://en.wikipedia.org/wiki/NetCDF>

<sup>59</sup>[https://en.wikipedia.org/wiki/GIS\\_file\\_formats](https://en.wikipedia.org/wiki/GIS_file_formats)

- **Позначки на карті.** Задається знаковий символ, який відображається в певному місці на карті, і його параметри: величина, текст і колір, величина написів, яка відображається біля знаку. Є можливість створити позначку без знаку. Наприклад, поставити номери будинків району. Для цього потрібно просто в стилі створити порожнє посилання на значок:

<Icon> <href> </href> </Icon>

- **Багатокутник або набір ліній.** Визначається колір ліній і колір підпису.
- **Зображення.** Визначається положення зображення на поверхні Землі, а також його масштаб. Також можна розмістити зображення на екрані, що не переміщається разом з картою - наприклад, логотип.
- **Тривимірна модель.** Версія мови KML 2.1 (яка відповідає четвертій версії програми Google Earth) дозволяє підключати опис тривимірних об'єктів (наприклад, будівель і споруд). Тривимірну модель можна задати двома способами: завданням висоти плоских фігур (витягуванням) і посиланням на повноцінну модель в форматі Collada.

Крім згаданих характеристик, для кожного об'єкта також задаються основні геоінформаційні властивості (географічна широта і довгота, а також висота над рівнем моря або над рівнем поверхні Землі). Може бути наведено короткий опис об'єкта (який в подальшому відображається в підказці за запитом користувача). Може бути зазначений рекомендований ракурс спостереження зазначеного на карті місця, тобто рекомендована висота, азимут і кут нахилу «віртуальної камери».

Об'єкти всередині KML-файлу можна організувати в ієрархічні структури папок і підпапок, щоб було зручніше спільно ввімкнути або вимкнути відображення логічно взаємопов'язаних груп об'єктів.

KML-файл може містити (в URL-формі) посилання на інші файли в форматі KML або KMZ, розташовані де-небудь в мережі, і задавати умови і регулярність завантаження і відображення даних з цих зовнішніх джерел. Таке мережеве посилання також видно в якості своєрідної підпапки.

Мова KML багато в чому наслідує структуру мови GML - географічного мови розмітки.

Приклад KML-розмітки:



```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <kml xmlns="http://earth.google.com/kml/2.1">
3  <Placemark>
4      <name>Київ</name>
5      <description><![CDATA[<p>Київ, Київська область, Україна.</p>
6          Столиця України. Місто побудоване на березі річки Дніпро.]]>
7      </description>
8      <LookAt id="khLookAt540_copy0">
9          <longitude>38.0576198113139</longitude>
10         <latitude>44.56963150481845</latitude>
11         <altitude>0</altitude>
12         <range>14693.40972993507</range>
13         <tilt>49.10268313434742</tilt>
14         <heading>37.85562764777833</heading>
15     </LookAt>
16     <Style>
17         <IconStyle>
18             <scale>0.9</scale>
19             <Icon>
20                 <href>root://icons/palette-4.png</href>
21                 <x>32</x>
22                 <y>128</y>
23                 <w>32</w>
24                 <h>32</h>
25             </Icon>
26         </IconStyle>
27         <LabelStyle>
28             <scale>0.9</scale>
29         </LabelStyle>
30     </Style>
31     <Point id="khPoint541_copy0">
32         <coordinates>
33             38.06284424434902,44.56842733252498,0
34         </coordinates>
35     </Point>
36 </Placemark>
37 </kml>

```

## GeoJSON та TopoJSON

GeoJSON - відкритий формат, призначений для зберігання графічних структур даних, заснований на форматі JSON.

Формат може зберігати примітивні типи для опису географічних об'єктів, такі як: точки (адреси та місця розташування), лінії (вулиці, шосе, кордони), полігони (країни, штати, ділянки землі). Також можуть зберігатися так звані мультитипи, які представляють собою об'єднання декількох примітивних типів.

Формат GeoJSON відрізняється від інших стандартів ГІС тим, що він був написаний і підтримується не певною організацією зі стандартизації, а за допомогою робочої групи розробників.

Подальшим розвитком GeoJSON є TopoJSON, розширення GeoJSON, яке кодує геопросторову топологію, і, як правило, забезпечує менший розмір файлів. Сьогодні ці формати підтримуються більшістю ГІС, а також безпосередньо застосовуються при публікації інтерактивних картографічних даних в веб, зокрема засобами бібліотек OpenLayers та Leaflet.

```
1 { "type": "FeatureCollection",
2   "features": [
3     { "type": "Feature",
4       "geometry": { "type": "Point", "coordinates": [102.0, 0.5] },
5       "properties": { "prop0": "value0" }
6     },
7     { "type": "Feature",
8       "geometry": {
9         "type": "LineString",
10        "coordinates": [
11          [102.0, 0.0], [103.0, 1.0], [104.0, 0.0], [105.0, 1.0]
12        ]
13      },
14      "properties": {
15        "prop0": "value0",
16        "prop1": 0.0
17      }
18    },
19    { "type": "Feature",
20      "geometry": {
21        "type": "Polygon",
22        "coordinates": [
23          [ [100.0, 0.0], [101.0, 0.0], [101.0, 1.0],
24            [100.0, 1.0], [100.0, 0.0] ]
25        ]
26      },
27      "properties": {
28        "prop0": "value0",
29        "prop1": { "this": "that" }
30      }
31    }
32  ]
33 }
```

```
31         }  
32     ]  
33 }
```

# Публікація та оновлення відкритих даних

## Вибір формату даних та конвертація даних з одного формату в інший

Вибір формату даних залежить від способу збереження та інших факторів, проте в більшості випадків можна користуватись такими загальними правилами:

1. Найкращим форматом для неструктурованих текстових даних (тексти законів, розпоряджень, довідкових даних) є формат Markdown або TXT. На відміну від TXT, Markdown також може містити додаткову розмітку - заголовки, абзаци, посилання, зображення тощо. Варто пам'ятати, що існує декілька різновидів Markdown, тому варто вказувати який саме різновид використовується. Гіршим варіантом є формат HTML, найгіршими - doc/docx, pdf/tiff (окрім випадків, коли потрібна копія оригінального документу із підписом та печаткою).
2. Для табличних даних потрібно використовувати формат CSV, в гіршому випадку - XLS/XLSX.
3. Для сканів документів найкращий формат - TIFF, потім - PDF.
4. До зображень встановлюється обов'язкова вимога до роздільності здатності, яка має дозволяти гарантоване розпізнавання (OCR) - не менше 300 dpi.
5. Великі файли або групи файлів потрібно архівувати в форматі ZIP/7Z. Для файлів, що перевищують розмір 4 ГБ - RAR.
6. Якщо дані змінюються постійно і формуються «на льоту», то їх взагалі можна не записувати у вигляді файлів. Достатньо вказувати доступ до них через адреси в Інтернет чи описаний алгоритм формування цих адрес на основі необхідних реквізитів та параметрів (через API). Рекомендованими форматами для доступу через API є JSON/XML/RDF, тому зберігати файли в цих форматах не потрібно (окрім випадків, коли дані зберігаються в цих форматах з самого початку).
7. Бінарні формати потрібно використовувати лише за відсутності альтернативи.

Часто дані зберігаються в форматах, що не відповідають визначеним вимогам відкритості. В такому разі на допомогу приходять різноманітні інструменти, за допомогою яких можна перевести той чи інший файл в інший формат. Це можуть бути як онлайн інструменти, так і програмні засоби, що встановлюються на робочу машину користувача.

Програмних засобів, бібліотек та сервісів існує дуже багато, тому ви легко зможете знайти підходящий для вас інструмент. Зазвичай немає жодних проблем з конвертацією даних з відкритих форматів в інші відкриті формати. Проблеми, скоріш за все виникнуть з форматами Word/PDF, тому при виникненні такої задачі потрібно аналізувати ці файли окремо.

Перелік деяких програмних засобів, що дозволяють конвертувати дані з одного формату в інший (перелік є не вичерпним):

1. Microsoft Word, Microsoft Excel, Microsoft Access.
2. Графічні формати та PDF: Adobe Photoshop, Adobe Acrobat Reader, Foxit PDF Reader, Paint.NET, Corel Draw.
3. Текстові дані: Notepad++.
4. Програмні засоби: SPSS, Microsoft Azure ML, Matlab.
5. Онлайн інструменти: [JSON to CSV](http://konklone.io/json/)<sup>60</sup>, [CSV to JSON](http://www.convertcsv.com/csv-to-json.htm)<sup>61</sup>, [JSON to C#](http://json2csharp.com/)<sup>62</sup>.
6. Очищення HTML документів від Word стилів: [https://word2cleanhtml.com](https://word2cleanhtml.com/)<sup>63</sup>.

## Основні помилки при оприлюдненні відкритих даних

### Чому Excel формат не є рекомендованим форматом?

В Постанові дозволяється викладати файли у форматі XLS/XLSX, що є форматам Microsoft Excel. Основна причина, по якій не рекомендується викладати Excel в тому, що програмне забезпечення дозволяє додавати розширене форматування, файли можуть містити пропуски, об'єднані комірки, вбудовані об'єкти, макроси та формули, що не дозволяють машинну обробку цих файлів.

Приклади Excel документів, що не відповідають правилам представлення даних у машиночитабельному форматі:

---

<sup>60</sup><http://konklone.io/json/>

<sup>61</sup><http://www.convertcsv.com/csv-to-json.htm>

<sup>62</sup><http://json2csharp.com/>

<sup>63</sup><https://word2cleanhtml.com/>

120203 Travel Expenses for Anne Example - completed exercise.xlsx - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View

Clipboard Font Alignment Number Styles Cells

Calibri 15 A A

Wrap Text

General

\$ % + -

Conditional Formatting Format as Table Cell Styles

Insert Delete Format

A5 April

	A	B	C	D	E	F	G	H	I	J
1	<b>Travel Expenses</b>									
2										
3	GST Calculated at:		15%							
4										
5	<b>April</b>									
6										
7	Expense	Project	Code	Amount ex GST	Discount	Total Discount	Total		GST	Total inc GST
8	Petty Cash	Alpha	AD765	\$ 52.83	0%	\$ -	\$ 52.83	\$ 7.92	\$ 60.75	
9	Insurance	Alpha	TR955	\$ 45.00	0%	\$ -	\$ 45.00	\$ 6.75	\$ 51.75	
10	Meals/Snacks	Alpha	AD203	\$ 132.75	0%	\$ -	\$ 132.75	\$ 19.91	\$ 152.66	
11	Car rental	Beta	AD985	\$ 150.00	5%	\$ 7.50	\$ 142.50	\$ 21.38	\$ 163.88	
12	Accommodation	Beta	TR256	\$ 560.00	10%	\$ 56.00	\$ 504.00	\$ 75.60	\$ 579.60	
13	Airfares	Beta	TR256	\$ 258.00	0%	\$ -	\$ 258.00	\$ 38.70	\$ 296.70	
14										
15	<b>Total</b>			<b>\$ 1,198.58</b>		<b>\$ 63.50</b>	<b>\$ 1,135.08</b>	<b>\$ 170.26</b>	<b>\$ 1,305.34</b>	
16										

R25

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											

companies

	State	Utilities Company	Summer Period		
			data1	data2	data3
12	Alabama	company1	3.458	1.734	0.584
13		company2	20.675	20.675	20.675
14	Alaska	company3	13.008	13.008	13.008
15		company4	12.1968	12.1968	12.1968
16		company5	16.698		12.35
17	Arizona	company6	29.56		8.06
18		company7	15.44		4.47
19	Arkansas	company6	1.576		1.125
20		company7	15.45	10.43	7.026
21		company8	10.48		8.48
22		company9	24.3		13.28
23		company10	16.06	8.14	4.655
24		company11	9.35		1.4
25	California	company12	23.67		6.24
26		company13	26		16
27		company14	34.91	18.99	5.78
28		company15	31	15	9
29		company16	16		14
30		company17	18.3	17.3	14.37



Приклади правильних файлів Excel:

A2    fx    Smith									
	A	B	C	D	E	F	G	H	I
1	Last Name	Sales	Country	Quarter					
2	Smith	\$16,753.00	UK	Qtr 3					
3	Johnson	\$14,808.00	USA	Qtr 4					
4	Williams	\$10,644.00	UK	Qtr 2					
5	Jones	\$1,390.00	USA	Qtr 3					
6	Brown	\$4,865.00	USA	Qtr 4					
7	Williams	\$12,438.00	UK	Qtr 1					
8	Johnson	\$9,339.00	UK	Qtr 2					
9	Smith	\$18,919.00	USA	Qtr 3					
10	Jones	\$9,213.00	USA	Qtr 4					
11	Jones	\$7,433.00	UK	Qtr 1					
12	Brown	\$3,255.00	USA	Qtr 2					
13	Williams	\$14,867.00	USA	Qtr 3					
14	Williams	\$19,302.00	UK	Qtr 4					
15	Smith	\$9,698.00	USA	Qtr 1					
16									
17									

A2 $f_x$ 1								
	A	B	C	D	E	F	G	H
1	Order ID	Product	Category	Amount	Date	Country		
2	1	Carrots	Vegetables	\$4,270	1/6/2012	United States		
3	2	Broccoli	Vegetables	\$8,239	1/7/2012	United Kingdom		
4	3	Banana	Fruit	\$617	1/8/2012	United States		
5	4	Banana	Fruit	\$8,384	1/10/2012	Canada		
6	5	Beans	Vegetables	\$2,626	1/10/2012	Germany		
7	6	Orange	Fruit	\$3,610	1/11/2012	United States		
8	7	Broccoli	Vegetables	\$9,062	1/11/2012	Australia		
9	8	Banana	Fruit	\$6,906	1/16/2012	New Zealand		
10	9	Apple	Fruit	\$2,417	1/16/2012	France		
11	10	Apple	Fruit	\$7,431	1/16/2012	Canada		

Дані в такому форматі легко можна експортувати в CSV формат, що є рекомендованим форматом.

## Оприлюднення занадто агрегованих даних

Розглянемо уявний приклад, коли потрібно оприлюднити інформацію про небезпечні ділянки, де найбільш часто трапляється велика кількість ДТП.

Очікується оприлюднення такої інформації:

- ділянка та її координати;
- дата ДТП;
- кількість постраждалих та тип їх ушкоджень;
- область (місто);
- стан ділянки, де сталось ДТП.

Натомість розпорядник викладає таблицю, де вказано кількість ДТП по областях. Зрозуміло, що дані в такому вигляді є малоінформативними і не несуть в собі великої користі.

Розпорядники не повинні викладати статистичну інформацію. Натомість потрібно викладати детальну інформацію на рівні окремих транзакцій, випадків (наприклад, ДТП) або певної події. Всі бажаючі легко зможуть зробити статистичні таблиці чи агреговану статистику на базі цих даних.

## Призначення відповідальних за розкриття даних

Кожен розпорядник даних повинен визначити відповідальну особу або декілька осіб, що будуть відповідати за аудит, обробку та викладення відкритих даних. Відповідальні особи повинні мати уявлення про те, які набори даних доступні та в якому вигляді, мати розуміння технічних аспектів роботи з даними (формати, розмір, аспекти роботи з реляційні і не реляційними базами даних тощо) та вміти користуватись єдиним порталом відкритих даних. Часто перед публікацією наборів даних потрібно здійснити їх аудит, обробку та/або конвертацію з одного формату в інший.

## Аудит наборів відкритих даних і пріоритетності їх публікації

Відомості, що публікуються у формі відкритих даних містять інформацію про діяльність державних органів і органів місцевого самоврядування, яка повинна розміщуватись в мережі відповідно до закону, а також відомості, що містять інформацію, яка збирається, зберігається, обробляється чи опублікована державними органами або органами місцевого самоврядування в рамках реалізації своїх повноважень.

Опублікування інформації необхідно здійснювати поетапно, враховуючи при цьому такі чинники:



1. потреба потенційними споживачами відповідних наборів відкритих даних, яка оцінюється відповідно до методики моніторингу і оцінки потреби відкритих даних;
2. ступінь готовності, що характеризується наявністю необхідних даних в електронному вигляді, а також готовністю організаційних, технічних, технологічних та інших засобів, необхідних для опублікування наборів відкритих даних;
3. витрати на публікацію (серед них: фінансові, тимчасові, трудові), необхідні для опублікування наборів відкритих даних і підтримки їх в актуальному стані.

Відомості, що володіють високим попитом і ступенем готовності, які потребують при цьому мінімальних витрат, повинні публікуватися в першочерговому порядку державними органами та органами муніципальної влади у формі відкритих даних.

При формуванні, як реєстру наборів відкритих даних, так і самих наборів відкритих даних слід враховувати такі фактори:

- публікації підлягають дані без попередньої первинної обробки;
- для кожного набору даних повинна бути визначена особа, відповідальна за їх утримання: коректність введення, повноту і актуальність;
- для кожного набору даних повинна бути встановлена періодичність оновлення з урахуванням потреб споживачів відкритих даних.

Реєстр відкритих даних формується та затверджується державним органом або органом місцевого самоврядування і після чого публікується в мережі на офіційному сайті органу.

## Підготовка даних до публікації

### Кодування файлів

Те, що ми бачимо на екрані як текст, насправді зберігається в текстовому файлі у вигляді числових значень. Комп'ютер перетворює числові значення на видимі символи. Це здійснюється за допомогою стандартів кодування.

Стандарт кодування є схемою нумерації, яка призначає числовому значенню текстовий символ із набору символів. Набір символів може містити абеткові символи, числа та інші символи. Різні мови зазвичай складаються з різних наборів символів, тому існує багато різноманітних стандартів кодування для відображення наборів символів, які використовуються в різних мовах.

Стандарт кодування, який зберігається разом із текстовим файлом, містить інформацію, потрібну для відображення тексту на екрані комп'ютера. Наприклад, у кодуванні «Кирилиця (Windows)» буква «Й» має числове значення «201». Якщо на комп'ютері, що використовує кириличне кодування (для Windows), відкрити файл, який містить цей символ, комп'ютер зчитує числове значення «201» та відображає букву «Й» на екрані.

Однак, якщо той самий файл відкрити на комп'ютері, що використовує інше кодування, буде відображено символ, який відповідає числовому значенню «201» у стандарті кодування, використовуваному на цьому комп'ютері за промовчанням. Наприклад, якщо комп'ютер використовує західноєвропейський стандарт кодування (для Windows), то символ у вихідному файлі на базі кирилиці буде відображено як «Е», а не як «Й», тому що в західноєвропейському стандарті кодування (для Windows) числове значення «201» відповідає символу «Е».

*Windows-1251* (також вживаються назви *Win1251*, *CP1251*) — кодування символів, що є стандартним 8-бітовим кодуванням для всіх локалізованих українських і російських версій Microsoft Windows. Користується досить великою популярністю. *Windows-1251* вигідно відрізняється від інших кирилических кодувань наявністю практично всіх символів, що використовуються в слов'янській кирилическій типографії для звичайного тексту (відсутній тільки значок наголосу); вона містить всі символи для російської, української, білоруської, сербської і болгарської мов. ([Вікіпедія](https://uk.wikipedia.org/wiki/Windows-1251)<sup>64</sup>)

*UTF-8* (від англ. *Unicode Transformation Format* — формат перетворення Юнікоду) — кодування, що реалізовує представлення Юнікоду, сумісне з 8-бітовим кодуванням тексту. Текст, що складається тільки з символів з кодом меншим, ніж 128, при записі в UTF-8 перетворюється на звичайний текст ASCII. І навпаки, в тексті UTF-8 будь-який байт із значенням менше, ніж 128 зображає символ ASCII з тим же кодом. Решта символів Юнікоду зображається послідовностями завдовжки від 2 до 6 байтів (реально тільки до 4 байтів, оскільки використання кодів більших, ніж 221 не планується), в яких перший байт завжди має вид 11xxxxxx, а решта — 10xxxxxx. Простіше кажучи, у форматі UTF-8 символи латинського алфавіту, розділові знаки і символи управління ASCII записуються кодами US-ASCII, а решта символів кодується за допомогою октетів із старшим бітом 1. В результаті, навіть якщо програма не розпізнає Юнікод, то англійська мова і розмітка все одно відображатимуться правильно. ([Вікіпедія](https://uk.wikipedia.org/wiki/UTF-8)<sup>65</sup>)

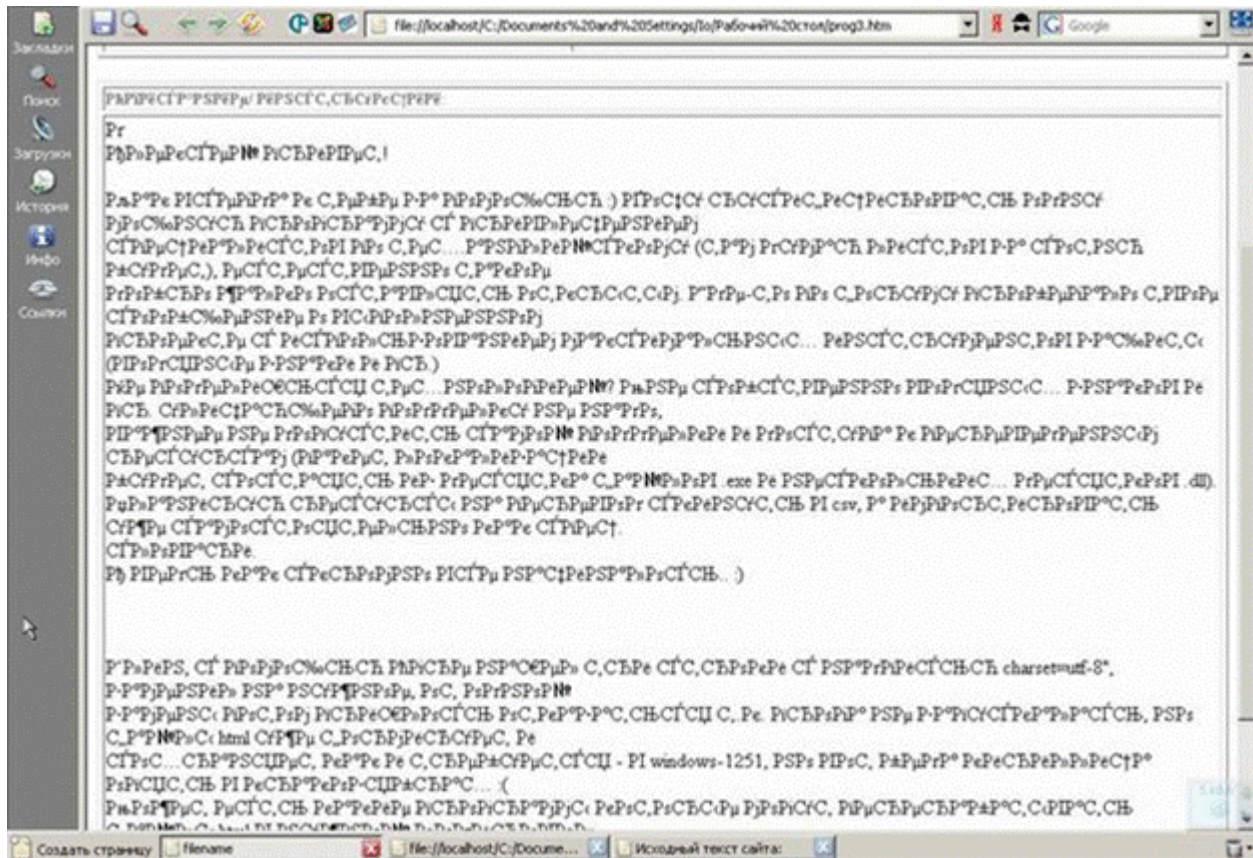
Проблема заключається в тому, що більшість державних органів користуються операційною системою Windows та зберігають всі файли та тексти у кодуванні Windows-1251, що часто є несумісними з іншими програмними засобами та операційними системами. Тому перед викладенням на data.gov.ua потрібно переконатись, що файли збережені з кодуванні UTF-8.

Ось такі символи можуть побачити користувачі при наявності проблем з кодуванням:

---

<sup>64</sup><https://uk.wikipedia.org/wiki/Windows-1251>

<sup>65</sup><https://uk.wikipedia.org/wiki/UTF-8>



## Один файл чи декілька?

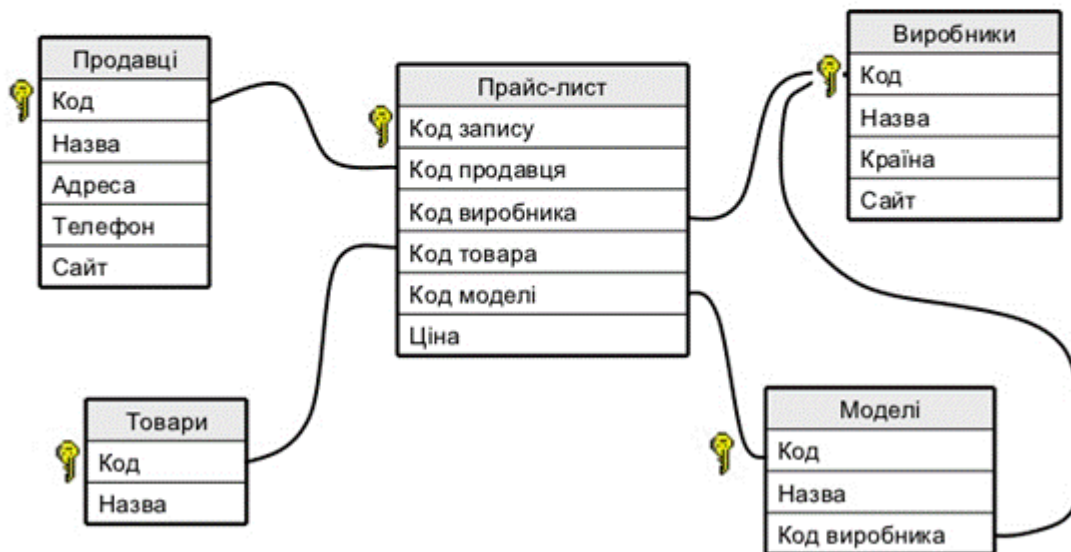
У випадку, якщо потрібно експортувати дані, які в початковому вигляді є реляційною базою (MySQL, Microsoft SQL, Oracle на ін.), наприклад:

## Реляційні БД

17

1970-і р. Е. Кодд, англ. *relation* – відношення.

Реляційна база даних – це набір простих таблиць, між якими встановлені зв'язки (відношення) за допомогою числових кодів.



то краще дані кожної таблиці експортувати в окремий файл.

Розглянемо приклад. Нехай маємо дві таблиці - виробники та моделі:

*Vendors (виробники)*

Id	Name	Country	Website
1	Microsoft	USA	<a href="http://microsoft.com">http://microsoft.com</a>
2	Apple	USA	<a href="http://apple.com">http://apple.com</a>

*Models (моделі)*

Id	Name	Vendor Code
1	Surface	1
2	iPhone	2

Ці дані можна представити у вигляді одного файлу *models.csv*:

```

1 Id;Name;Vendor;Country;Website
2 1;Surface;Microsoft;USA;http://microsoft.com
3 2;iPhone;Apple;USA;http://microsoft.com

```

У вигляді двох файлів:

*vendors.csv:*

```

1 Id;Name;Country;Website
2 1;Microsoft;USA;http://microsoft.com
3 2;Apple;USA;http://apple.com

```

*models.csv:*

```

1 Id;Name;VendorId
2 1;Surface;1
3 2;iPhone;2

```

Порівняння двох підходів

Критерій	Один файл	Багато файлів
Збитковість даних	Велика збитковість	Мала збитковість
Фінальний розмір	Більший	Менший
Перевірка на цілісність	Складна	Проста
Готовність до використання через API	Погана	Хороша, по суті відображає методи API
Кількість запитів	Один (перевага лише для невеликих за розміром файлів)	Багато

Як видно з таблиці, другий варіант використовувати доцільніше.

При публікації даних на єдиному порталі структура та призначення кожного файлу повинні бути описані та опубліковані як окремі файли. Якщо розмір хоча б одного файлу велика, то допускається публікація всіх файлів у вигляді одного архівного файлу (але з детальним описом всіх файлів).

## Деперсоніфікація даних

Згідно закону “Про захист персональних даних”<sup>66</sup> до персональних даних можна віднести будь-які відомості, за якими ідентифікується або може бути ідентифікована фізична особа,

<sup>66</sup><http://zakon3.rada.gov.ua/laws/show/2297-17>

зокрема: прізвище, ім'я, по батькові, адреса, телефони, паспортні дані, національність, освіта, сімейний стан, релігійні та світоглядні переконання, стан здоров'я, матеріальний стан, дата і місце народження, місце проживання та перебування тощо, дані про особисті майнові та немайнові відносини цієї особи з іншими особами, зокрема членами сім'ї, а також відомості про події та явища, що відбувалися або відбуваються у побутовому, інтимному, товариському, професійному, діловому та інших сферах життя особи (за винятком даних стосовно виконання повноважень особою, яка займає посаду, пов'язану із здійсненням функцій держави або органу місцевого самоврядування) тощо. Вказаний перелік не є вичерпним.

Така інформація про фізичну особу та членів її сім'ї є конфіденційною і може оброблятися в тому числі поширюватись тільки за їх згодою, крім випадків, визначених законом, і лише в інтересах національної безпеки, економічного добробуту та прав людини.

Тому, якщо дані містять персональну інформацію, то такі набори повинні пройти процес деперсоналізації даних, що означає видалення таких відомостей.

В більшості випадків прив'язку даних до конкретної людини можна замінити ідентифікатором (але це не може бути номер паспорту чи ІНН) шляхом генерації порядкового номеру або GUID, що можна зробити будь-якими програмними засобами.

Приклад початкових даних:

ППП	Місто	Телефон	Група крові
Іванов І.І.	Київ	011 111 11 11	A(I)+
Петров П.П.	Одеса	011 111 11 12	B(II)-

Кроводачі:

ППП	Центр	Дата
Іванов І.І.	Охматдит	01.01.2016
Петров П.П.	Охматдит	01.02.2016

Дані після деперсоналізації:

ППП	Місто	Група крові
1	Київ	A(I)+
2	Одеса	B(II)-

Кроводачі:

ППП	Центр	Дата
1	Охматдит	01.01.2016
2	Охматдит	01.02.2016

## Архівація наборів даних

Архівація даних необхідна для зменшення розмірів наборів даних, а, відповідно, необхідних ресурсів для зберігання даних на порталі, а також з метою зменшення трафіку для порталу та кінцевих споживачів.

Потрібно пам'ятати, що архівація текстових даних може зменшити розміри файлів до 90%. Word/Excel/PDF дані можуть стискатись в середньому на 10-60%, майже не стискаються зображення (лише на 1-3%).

Архівувати потрібно:

- історичні дані (наприклад, курси валют за попередній рік чи десятиліття);
- усі файли набору, що мають розмір більше 50 МБ.
- застарілі версії наборів даних, якщо є доступ до оновлених версій набору відкритих даних
- якщо публікується багатотомний набір, тоді рекомендується всі томи не копіювати і зберігати в одному архіві.

Архівувати потрібно в форматах zip/7z/tar, що є кросплатформеними та не вимагають платних програм для деархівації. Рекомендується використовувати програму 7-zip <http://7-zip.org.ua/><sup>67</sup> для архівування/деархівування даних для ОС Windows та вбудовані засоби в інших ОС.

## Публікація даних, що мають періодичність

Набори відкритих даних поділяються на частопубліковані дані (висока періодичність актуалізації даних - частіше чим 1 раз в тиждень), рідкопубліковані (рідше, чим раз в тиждень).

Для кожного набору даних в паспорті набору має бути вказана дата останнього оновлення.

**Частота оновлень може бути наступною:**

**Частооновлювані:**

- більше одного разу в день;
- кожного дня;
- кожного тижня

**Рідкооновлювальні:**

- кожного місяця;

---

<sup>67</sup><http://7-zip.org.ua/>

- кожного квартала;
- кожних півроку;
- кожного року;
- по мірі зміни даних.

Для наборів даних, які оновлюються періодично (наприклад, курси валют або список денних транзакцій), то перш за все, необхідно визначити критерій, за яким відбуватиметься групування: дата, місяць, скликання (для депутатів) тощо.

Цей критерій має чітко бути описаний в паспорті даних, а назви файлів повинні містити цей критерій.

Наприклад, курси валют за 1 січня 2016 року при виборі критерію - дати:

```
exchange_20160101.csv (exchange_yyyymmdd.csv):  
USD;20,3333  
RUB;0,3456  
EUR;27,9090
```

Приклад публікації курсів валют при виборі критерію - валюти:

```
exchange_usd.csv (exchange_yyyymmdd.csv):  
20160103;20,3333  
20160102;20,4455  
20160101;20,3333
```

Для цього випадку логічніше групувати дані по даті.

Якщо набір даних містить інформацію, зміна яких відбувається досить рідко і не регулярно, то в такому разі дані повинні оновлюватись кожен раз, коли відбувається така зміна. Приклади: перелік дитсадочків, перелік аптек міста, розклад міського транспорту. В паспорті даних потрібно вказати орієнтовну кількість оновлень протягом місця чи року.

Якщо набір даних містить інформацію про показники чи транзакції, що змінюються чи доповнюються з однаковою регулярністю, то викладення цих даних повинно прив'язуватись до цього періоду. Приклади таких наборів: курси валют (оновлюються щоденно), державні закупівлі (оновлюються щоденно). Якщо кількість даних за певний період є великою, то можливо розбиття на менші кратні проміжки (тижні, дні, години), і навпаки, якщо кількість даних за певний проміжок мала, то можна збільшувати цей проміжок. Хорошою практикою є публікація проміжкових даних, наприклад, курси валют за сьогодні, і історичні дані за певний період (по рокам, десятиліттям чи за весь доступний період - в залежності від специфіки даних).



Якщо кількість даних є великою, зміни постійними і з непередбачуваною періодичністю (по суті, якщо дані оновлюються в реальному часі), зокрема, це стосується всіх реєстрів, то потрібно передбачити доступ до даних у вигляді API для того, щоб була змога завжди отримати актуальні дані. Архіви даних можуть також публікуватись з певною періодичністю, але з припискою, що вони є актуальними на певний момент часу.

## Публікація схожих даних з різними структурами

Якщо суть даних схожа (однакова), проте в залежності від різних факторів структура відрізняється, то потрібно керуватись такими рекомендаціями:

Якщо структура даних всіх файлів схожа і відрізняється не значно, то приймається за основу одна структура (яка складатиметься з усіх стовбців всіх файлів), а відсутні дані заповнюються пустими значеннями.

Наприклад, маємо два файли з такими полями:

```
A1,A2,A3,A4
1, 1, 1, 1
```

```
A1,A2,A3,A5
2, 2, 2, 2
```

то за основу приймається структура A1,A2,A3,A4,A5 і фінальний файл матиме вигляд:

```
A1,A2,A3,A4,A5
1, 1, 1, 1,
2, 2, 2, , 2
```

У випадку, якщо структура файлів відрізняється значно, то такі файли потрібно розглядати як незалежні один від одного.

## Публікація даних у вигляді API та робота з даними великого обсягу

Є два основних способи організації доступу до даних - Data Hub (сховище даних) та API (Application Programming Interface - інтерфейс доступу до даних).

Data Hub краще підходить для даних, що змінюються рідко або час від часу, а API – для швидких оперативних запитів і міжсерверної взаємодії. Можлива комбінація двох способів, коли частина відкритих даних (довідники, списки) знаходяться в статичному вигляді на порталі, окремі елементи (картки чи тексти документів, фотографії, приєднані файли тощо)

доступні за відповідною URI-адресою на іншому сервері, як зовнішні ресурси, а доступ до інформації, що постійно змінюється (записи реєстру) - через API.

Для API рекомендується використовувати [Open Data Protocol](http://www.odata.org/)<sup>68</sup> (OData) - відкритий веб-протокол для запиту і оновлення даних. Протокол дозволяє виконувати операції з ресурсами, використовуючи в якості запитів HTTP-команди, і отримувати відповіді в форматах XML або JSON.

При публікації даних у вигляді API потрібно вказати тип даних - API. Крім того, в описі набору потрібно вказати посилання на документацію та інформацію про те, на яких умовах можна отримати дані - після реєстрації, платно/безоплатно, про умови використання, чи є обмеження на кількість та тип запитів, доступні формати, чи реалізований протокол OData.

## Канали розповсюдження відкритих даних

Відкриті дані можна розповсюджувати за допомогою таких каналів:

1. Через сайт [data.gov.ua](http://data.gov.ua).
2. Через сайти місцевих органів самоврядування.
3. За допомогою API.
4. Через ftp-сервер.
5. Через файлообмінний протокол BitTorrent.

---

<sup>68</sup><http://www.odata.org/>

# Відкриті дані в Україні

## Законодавство України на тему відкритих даних та готовність країни до впровадження

1. Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних № 835 — редакція від 21.10.2015<sup>69</sup>
2. ПЛАН ЗАХОДІВ<sup>70</sup> (ДОРОЖНЯ КАРТА) РОЗВИТКУ ВІДКРИТИХ ДАНИХ НА 2016 РІК
3. Єдиний державний портал відкритих даних data.gov.ua - Керівництво користувача
4. Єдиний державний портал відкритих даних data.gov.ua - Робота з АПІ
5. Open Data Readiness Assessment (ODRA) Ukraine final Report <https://docs.google.com/document/d/1g5VUh>

## Україна в світових рейтингах

### Open Data Barometer

Сайт: <http://www.opendataresearch.org/barometer><sup>72</sup>

Барометр відкритих даних. Спільний науково-дослідницький проект Open Data Institute и World Wide Web Foundation.

Open Data Barometer аналізує глобальні тенденції розвитку та поширення відкритих даних, а також надає порівняльні дані по країнах і регіонах за допомогою методології поглибленого комбінування контекстних даних, технічних оцінок та вторинних показників.

Проаналізувавши дані дослідники створили рейтинг країн в залежності від розвитку відкритих даних. Наша країна знаходиться на 62 місці із 86 проаналізованих країн (2015 - 55 місце).

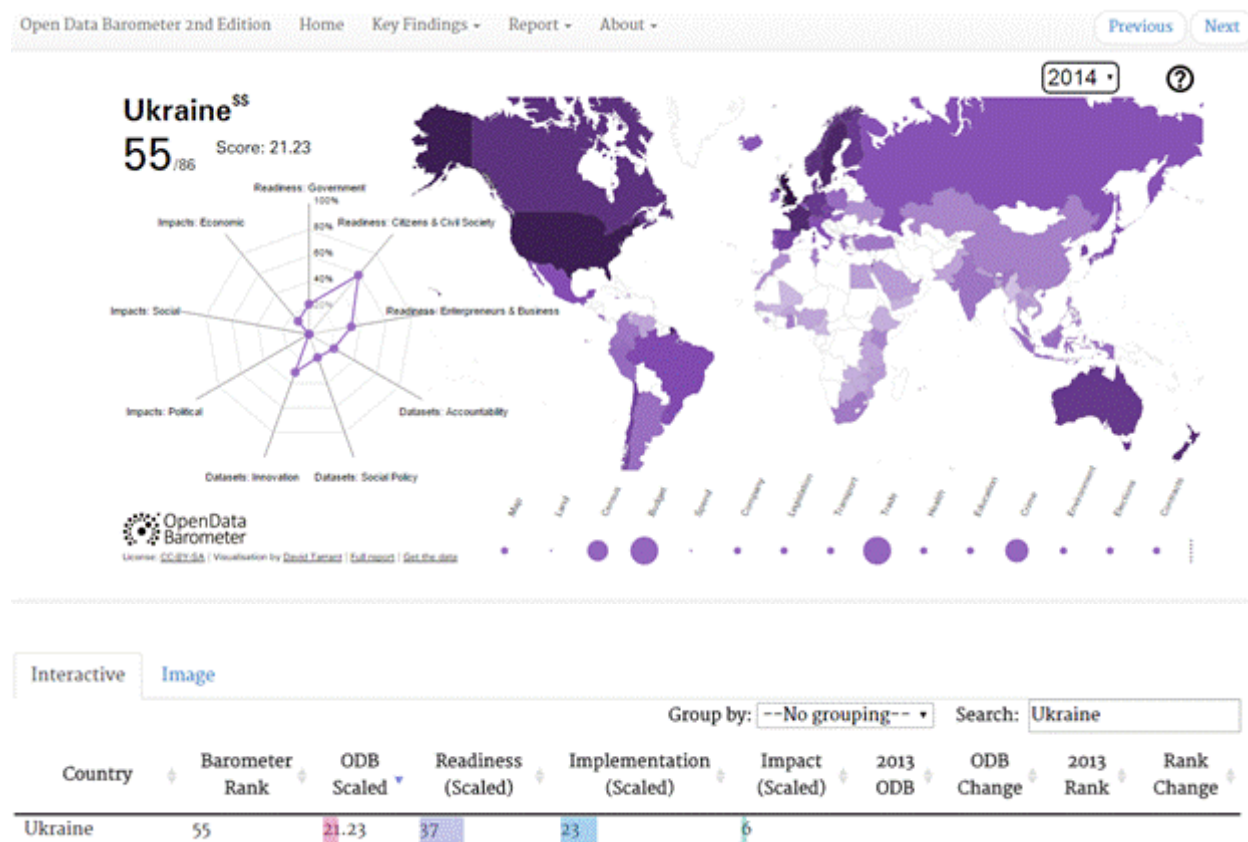
---

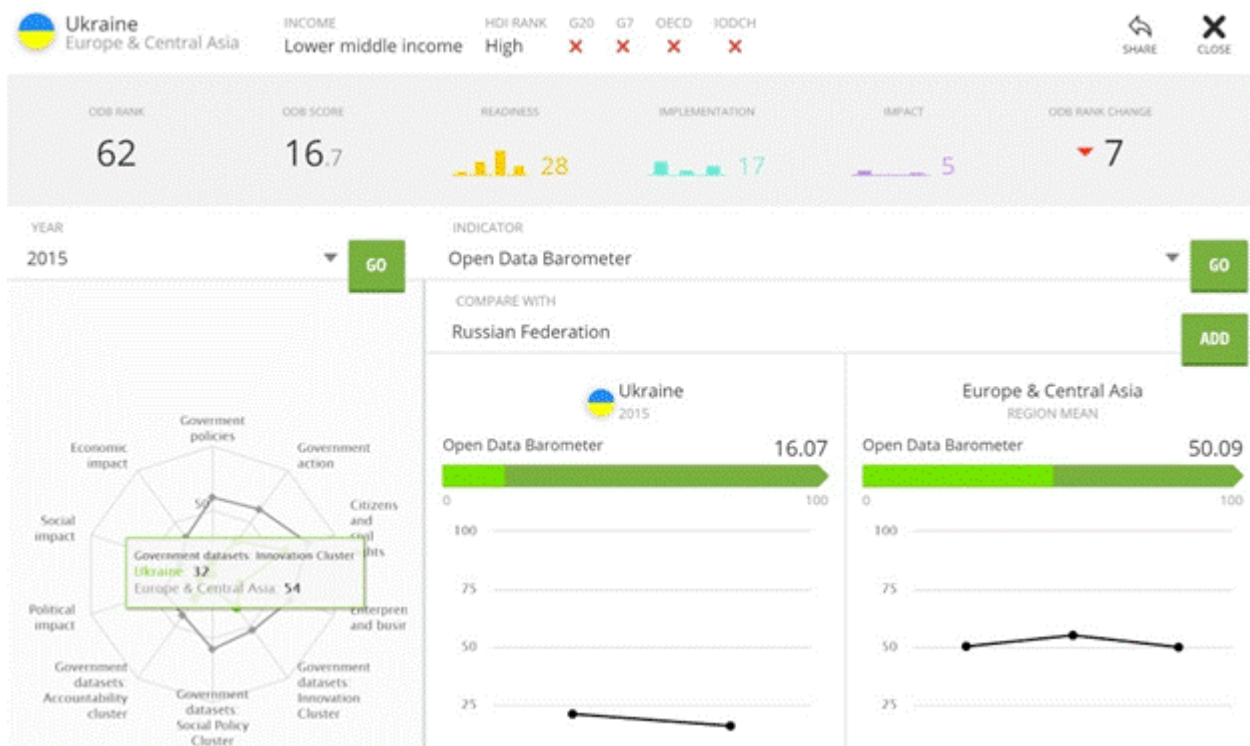
<sup>69</sup><http://zakon2.rada.gov.ua/laws/show/319-19>

<sup>70</sup>[https://drive.google.com/file/d/0B1kGsKt9XV\\_QaFZVaTZiT19aRTA/view](https://drive.google.com/file/d/0B1kGsKt9XV_QaFZVaTZiT19aRTA/view)

<sup>71</sup><https://docs.google.com/document/d/1g5VUhUzjgTVvcp7zvlWbCUsmoJNaQ6w4T3vacYSu34U/edit>

<sup>72</sup><http://www.opendataresearch.org/barometer>





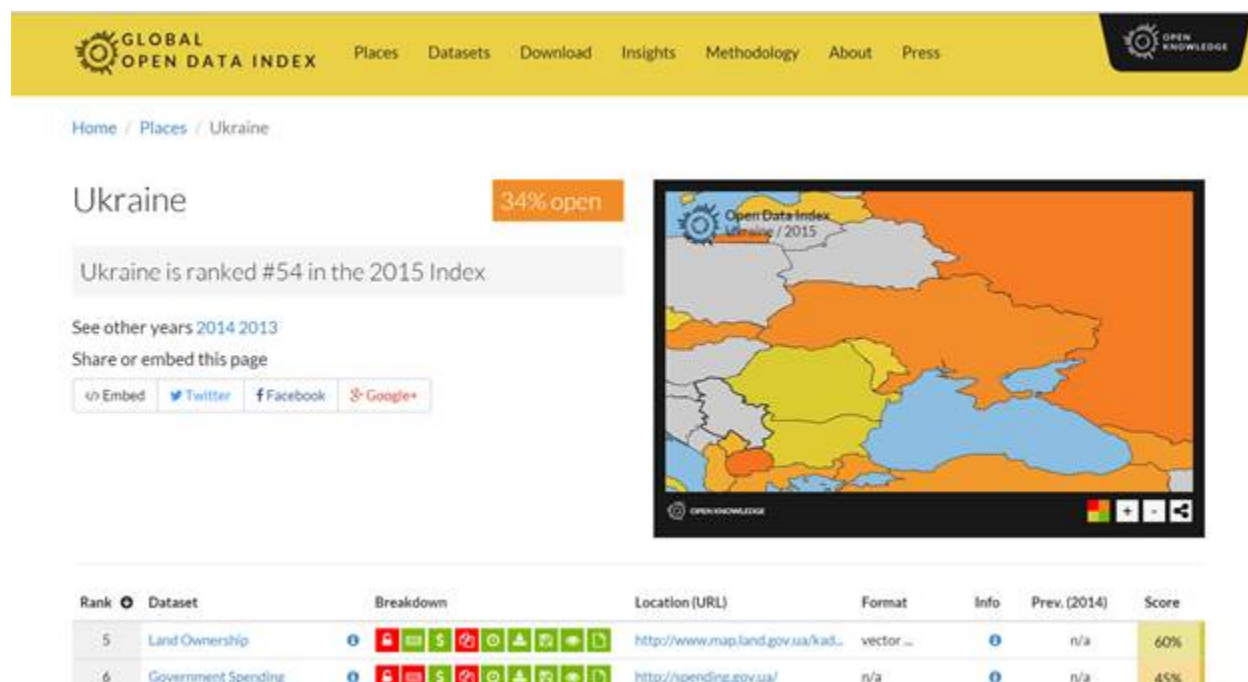
## Open Data Index

Сайт: <https://index.okfn.org/><sup>73</sup>

Масштабний дослідний проект Open Knowledge Foundation по вивченню відкритості державних даних в країнах світу.

Україна знаходиться на 54 місці в глобальному рейтингу країн станом на 2015 рік.

<sup>73</sup><https://index.okfn.org/>



## The Web Index

Сайт: <http://thewebindex.org/><sup>74</sup>

Рейтинг розвитку інтернету в країнах світу, дослідницький проект World Wide Web Foundation. Web Index аналізує кроки, що зроблені для розвитку інтернету. Оновлення кожного року дає можливість оцінити тенденції розвитку певної країни в порівнянні з іншими. Всього було проаналізовано 86 країн, серед яких наша займає 46 місце.

<sup>74</sup><http://thewebindex.org/>

Chapters

1. Executive Summary: The Web and Growing Global Inequality

2. Overview of Rankings

3. The Web and Economic Inequality

4. The Web and Gender Inequality










5. The Web and Social Inequality

6. The Web and Political Inequality

7. Conclusions and Recommendations

8. Credits and Acknowledgements

9. Case Studies

42	Malaysia		48.34
43	Peru		46.62
44	China		45.97
45	South Africa		45.82
46	Ukraine		45.2
47	United Arab Emirates		44.9
48	India		44.6
49	Jamaica		44.5
50	Ecuador		42.57
51	Morocco		40.38

## Open Budget Survey

Сайт: <http://internationalbudget.org/what-we-do/open-budget-survey/><sup>75</sup>

Рейтинг відкритості державних бюджетів, підготовлених International Budget Partnership. Open Budget Survey оцінює показники країни по чотирьох критеріях. Для нашої країни вони складають:

- Прозорість (відкритий бюджет) 46/100
- Участь громадськості 23/100
- Нагляд державним органом 79/100
- Нагляд вищого органу фінансового контролю 83/100.

<sup>75</sup><http://internationalbudget.org/what-we-do/open-budget-survey/>



## E-Government for the People

Сайт: [http://www.unpan.org/egovkb/global\\_reports/08report.html](http://www.unpan.org/egovkb/global_reports/08report.html)<sup>76</sup>

Рейтинг ООН по розвитку електронного уряду.

На Конференції Організації Об'єднаних Націй з питань розвитку технологій, що відбулася в Ріо-де-Жанейро в червні 2012 року було досягнуто згоди, що для розвитку суспільства потрібно мати інститути на всіх рівнях, які є ефективними, прозорими, підзвітними і демократичними. Електронний уряд має величезний потенціал для поліпшення надання державних послуг та активізації участі широкого кола зацікавлених сторін в державній службі. Це дозволить громадськості брати участь у прийнятті рішень, надати доступ до інформації та усунення бар'єрів між держслужбовцями та громадськістю.

<sup>76</sup>[http://www.unpan.org/egovkb/global\\_reports/08report.html](http://www.unpan.org/egovkb/global_reports/08report.html)



Table 1.2. Countries grouped by ECR in alphabetical order

Very high COVID-19 cases (over 100,000)	High COVID-19 cases (between 10,000 and 100,000)	Medium COVID-19 cases (between 1,000 and 10,000)	Low COVID-19 cases (less than 1,000)		
Australia	Albania	Indonesia	Algeria	Norway	Algeria
Austria	Armenia	Malta	Angola	Mali	Bosnia
Belgium	Azerbaijan	Mauritius	Antigua	Morocco	Bosnia Herz
Bulgaria	Bahrain	Mexico	Bangladesh	Nepal	Brazil
Canada	Bangladesh	Nigeria	Barbados	Nicaragua	Central American
Denmark	Belize	Paraguay	Belize	Philippines	Costa Rica
Ecuador	Bermuda	Peru	Bolivia	Poland	Cuba
Egypt	Bhutan	Poland	Bosnia and Herzegovina	Portugal	Czechia
France	Bolivia	Qatar	Brazil	Romania	China
Germany	Brazil	Paraguay	Bulgaria	Romania	Colombia
Greece	Bulgaria	Peru	Cameroon	Romania	Colombia
India	Bulgaria	Poland	Cameroon	Romania	Colombia
Italy	Chile	Portugal	Canada	Romania	Colombia
Japan	China	Qatar	Canada	Romania	Colombia
Lebanon	Czechia	Qatar	Canada	Romania	Colombia
Netherlands	Czechia	Qatar	Canada	Romania	Colombia
New Zealand	Czechia	Qatar	Canada	Romania	Colombia
Norway	Czechia	Qatar	Canada	Romania	Colombia
Republic of Korea	Czechia	Qatar	Canada	Romania	Colombia
Saudi Arabia	Czechia	Qatar	Canada	Romania	Colombia
Spain	Czechia	Qatar	Canada	Romania	Colombia
Sweden	Czechia	Qatar	Canada	Romania	Colombia
Switzerland	Czechia	Qatar	Canada	Romania	Colombia
United Kingdom	Czechia	Qatar	Canada	Romania	Colombia
United States	Czechia	Qatar	Canada	Romania	Colombia
Various	Czechia	Qatar	Canada	Romania	Colombia

Як видно із статистики, коефіцієнт розвитку електронного уряду лежить у діапазоні між 0,5 та 0,75, що є непоганим показником, але ще далеким від ідеалу.

## Проекти на базі відкритих даних

## Нові назви вулиць, площ Дніпропетровська

Посилання: <http://rename.dp.ua/><sup>77</sup>

Нові назви вулиць, площ Дніпропетровська по районах, повний перелік Всі назви українською мовою згідно з оригіналами розпоряджень про перейменування 2015 року.

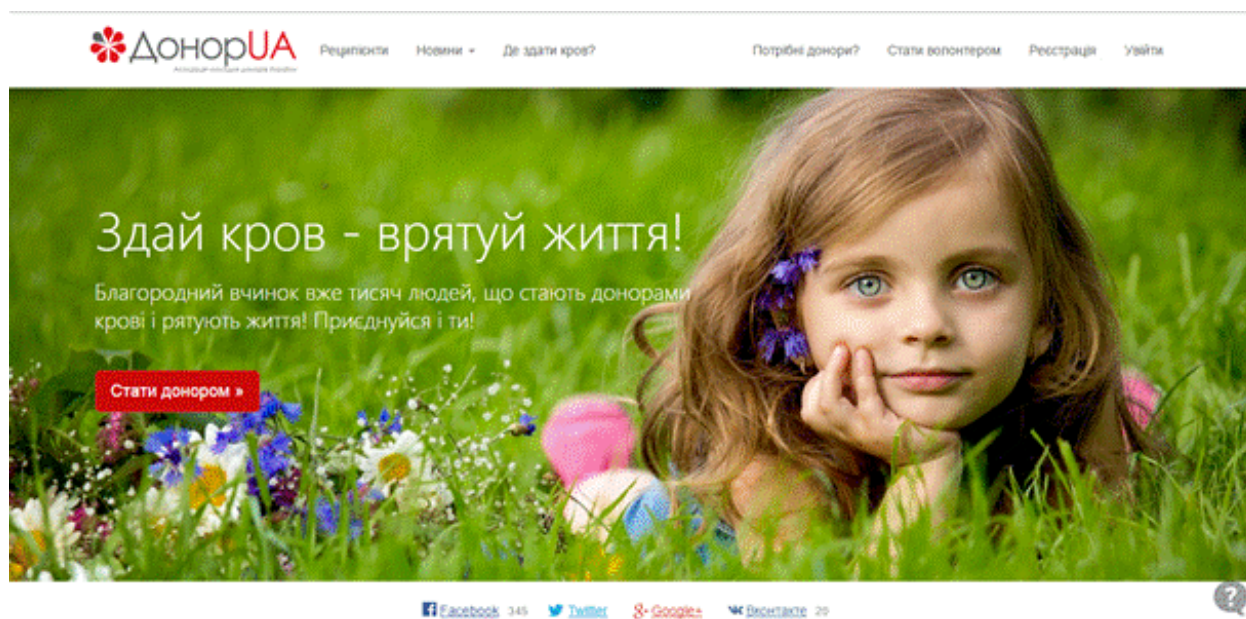
<sup>77</sup><http://rename.dp.ua/>



## ДонорUA

Посилання: <http://donor.ua/><sup>78</sup>

Проект спрямований на координацію донорів крові та пропаганду безоплатного донорства в Україні. Унікальна розробка - автоматизована система дозволяє встановити миттєвий контакт донор - реципієнт, що дає змогу значно скоротити процес пошуку донорів.

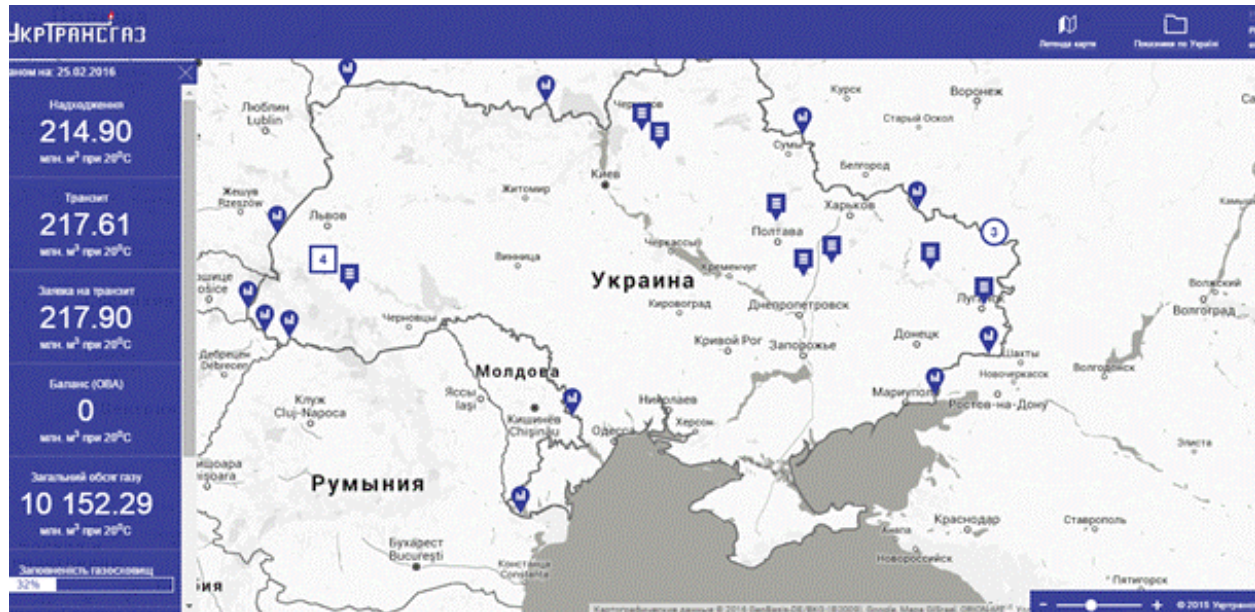


<sup>78</sup><http://donor.ua/>

## УкрТрансГаз

Посилання: <http://utg.ua/live/><sup>79</sup>

На карті показані трубопроводи через які газ потрапляє до України і через які залишає її. Показані сховища а також скільки газу закачано та викачано за добу і багато іншого.



## НБУ

Посилання: <http://nbu.rocks/><sup>80</sup>

Відкриті дані Національного Банку України про українські банки.

<sup>79</sup><http://utg.ua/live/>

<sup>80</sup><http://nbu.rocks/>



Відкриті дані **НБУ\*** про українські банки // (JSON, 10mb)

\*А саме: Структури власності, Банківські ліцензії та види діяльності, Діючі структурні парозвіли

For me on GitHub

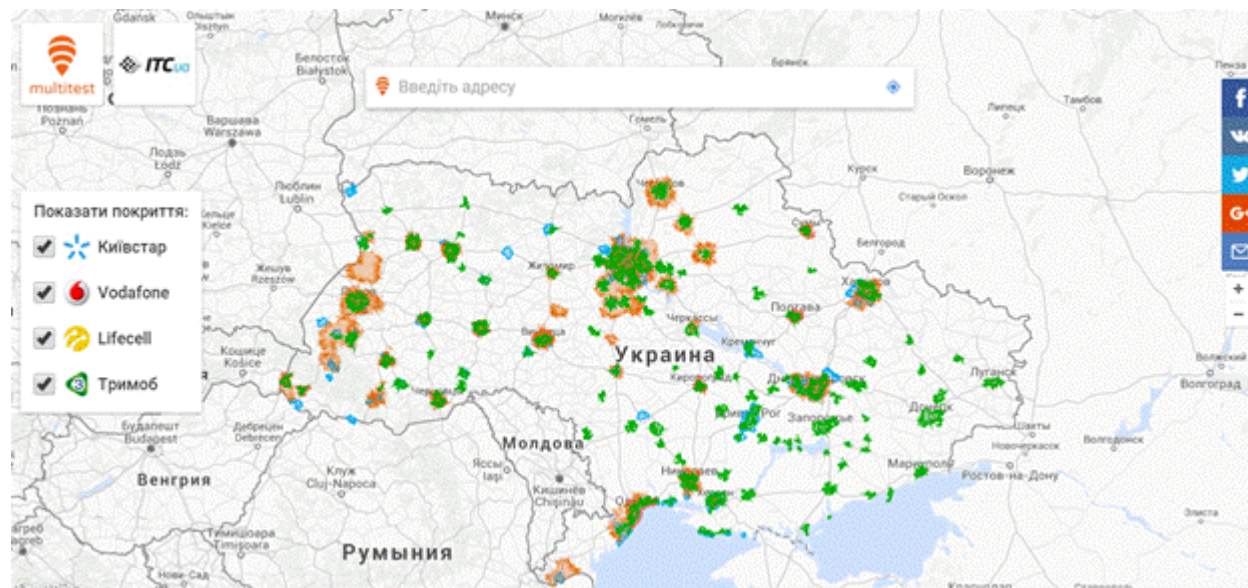
Пошук:

МФО	Назва	Адреса	Філії	Структура	Деталі
300012	Публічне акціонерне товариство "Акціонерний комерційний промислово-інвестиційний банк"	01001, м. Київ, провулок Шевченка, 12	108	<a href="#">↓</a>	<a href="#">📄</a>
300023	ПУБЛІЧНЕ АКЦІОНЕРНЕ ТОВАРИСТВО "УКРСОЦБАНК"	03150, м. Київ, вул. Копака, 29	263	<a href="#">↓</a>	<a href="#">📄</a>
300056	Публічне акціонерне товариство "Легбанк"	01033, м. Київ, вул. Жилинська, 27		<a href="#">↓</a>	<a href="#">📄</a>
300089	Акціонерний комерційний банк "Трансбанк"	03150, м. Київ, вул. Факультури, 9			<a href="#">📄</a>
300090	Акціонерний комерційний банк "Народний банк"	01025, м. Київ, вул. Велика Житомирська, 24			<a href="#">📄</a>
300119	Публічне акціонерне товариство "Банк Альянс"	04073, м. Київ, проспект Московський, буд. 8, корпус 15-А	0	<a href="#">↓</a>	<a href="#">📄</a>
300120	Публічне акціонерне товариство "Банк Петрокоммерц-Україна"	01025, м. Київ, вулиця Велика Житомирська, 20	2	<a href="#">↓</a>	<a href="#">📄</a>
300131	Публічне акціонерне товариство "Банк "Фінанси та Кредит"	04050, м. Київ, вул. Артема, 60		<a href="#">↓</a>	<a href="#">📄</a>

## 3G.Multitest.ua

Посилання: <http://3g.multitest.ua/><sup>81</sup>

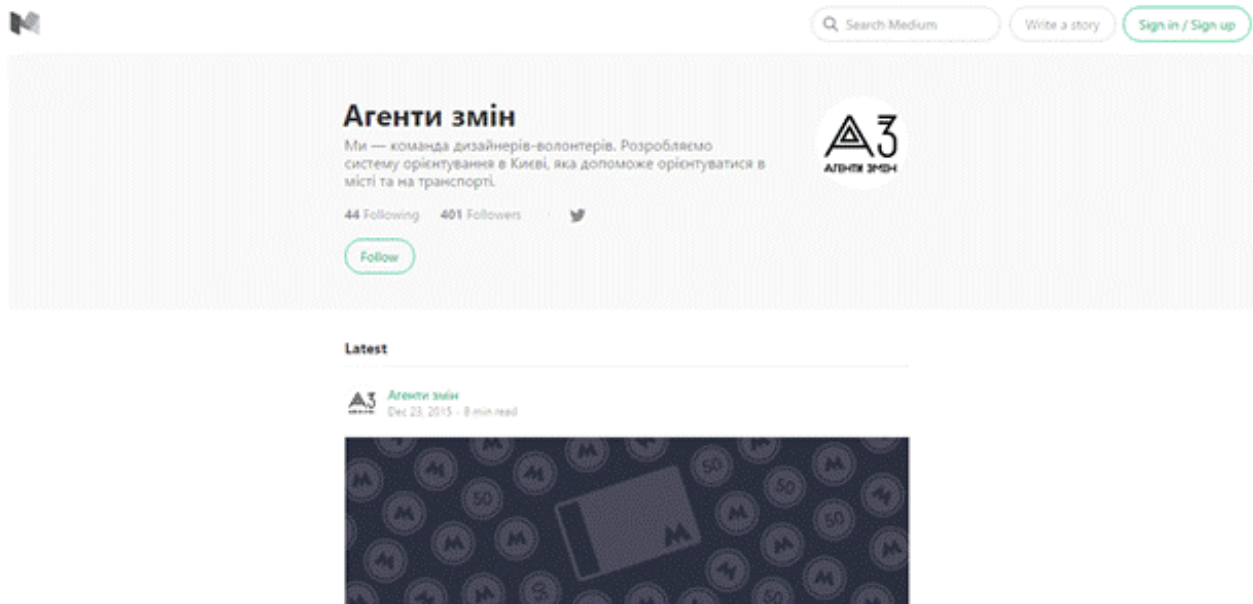
Зведена карта покриття всіх 3G GSM операторів України.

<sup>81</sup><http://3g.multitest.ua/>

## Агенти змін

Посилання: <https://medium.com/@agencyzmin><sup>82</sup>

Команда дизайнерів-волонтерів, розробляє систему орієнтування в Києві, яка допоможе орієнтуватися в місті та на транспорті.



## Лун.UA

Посилання: <http://www.lun.ua/><sup>83</sup>

Оренда та продаж нерухомості у Києві та Україні.

---

<sup>82</sup><https://medium.com/@agencyzmin>

<sup>83</sup><http://www.lun.ua/>

## PEP.ORG.UA

Посилання: <http://pep.org.ua/uk/><sup>84</sup>

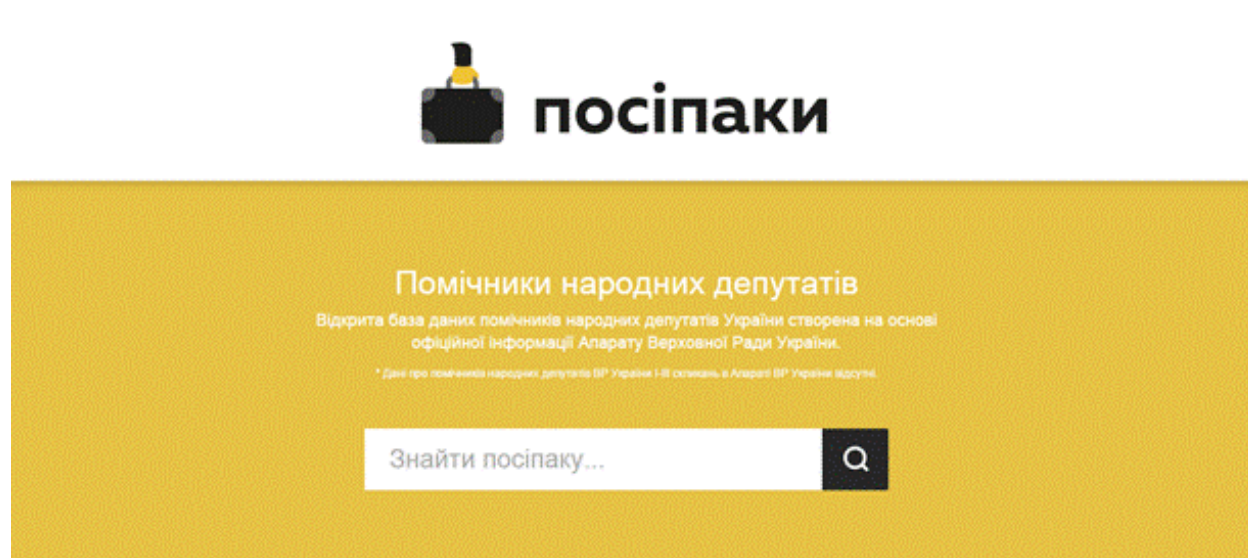
Відкритий реєстр національних публічних діячів України. Корумповані високопосадовці легко відмивають брудні гроші через міжнародну банківську систему. Тому у світі запровадили особливі правила контролю за фінансовими операціями публічних осіб: кожна фінансова установа має встановити, хто її клієнт і звідки в нього гроші. Якщо клієнтом є публічний діяч, його фінансові операції мають бути ретельно перевірені.

<sup>84</sup><http://pep.org.ua/uk/>

## Посіпаки

Посилання: <http://posipaky.info/><sup>85</sup>

Відкрита база даних помічників народних депутатів України створена на основі офіційної інформації Апарату Верховної Ради України



## CityScale

Посилання: <http://www.cityscale.com.ua/><sup>86</sup>

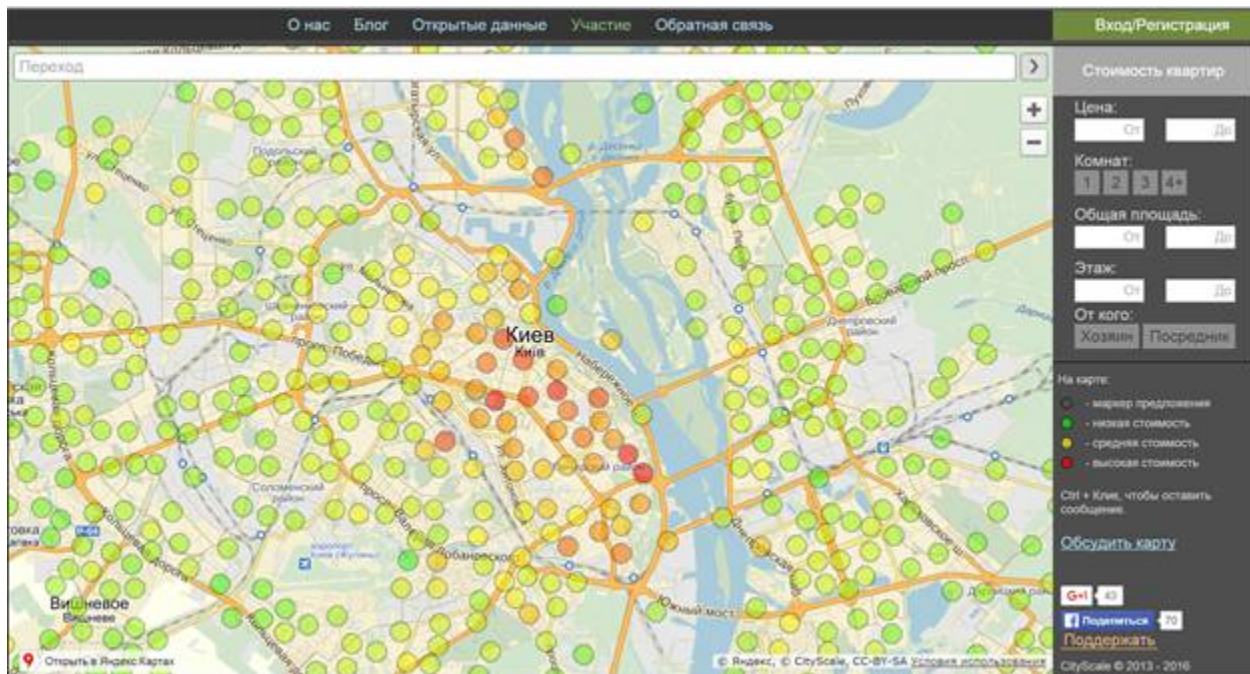
Інформаційний портал, який надає жителям і фахівцям зручний доступ до даних про різних умовах життя, таких як: вартість житла, комунальні тарифи, злочинність, екологія, інфраструктура та інші, у вигляді інтерактивних карт, файлів і ресурсів даних, починаючи з окремого будинки і закінчуючи всією країною.

---

<sup>85</sup><http://posipaky.info/>

<sup>86</sup><http://www.cityscale.com.ua/>





## ARBUZ

Мобільний додаток, задача якого - вирішити проблему безпеки в Києві і інших містах. Сервіс базується на офіційних даних, отриманих від Управління патрульної поліції Києва і включає в себе карту небезпечних районів і вулиць міста Києва, пошук безпечного маршруту в будь-яку точку міста і повідомлення про рівень безпеки в місці перебування.





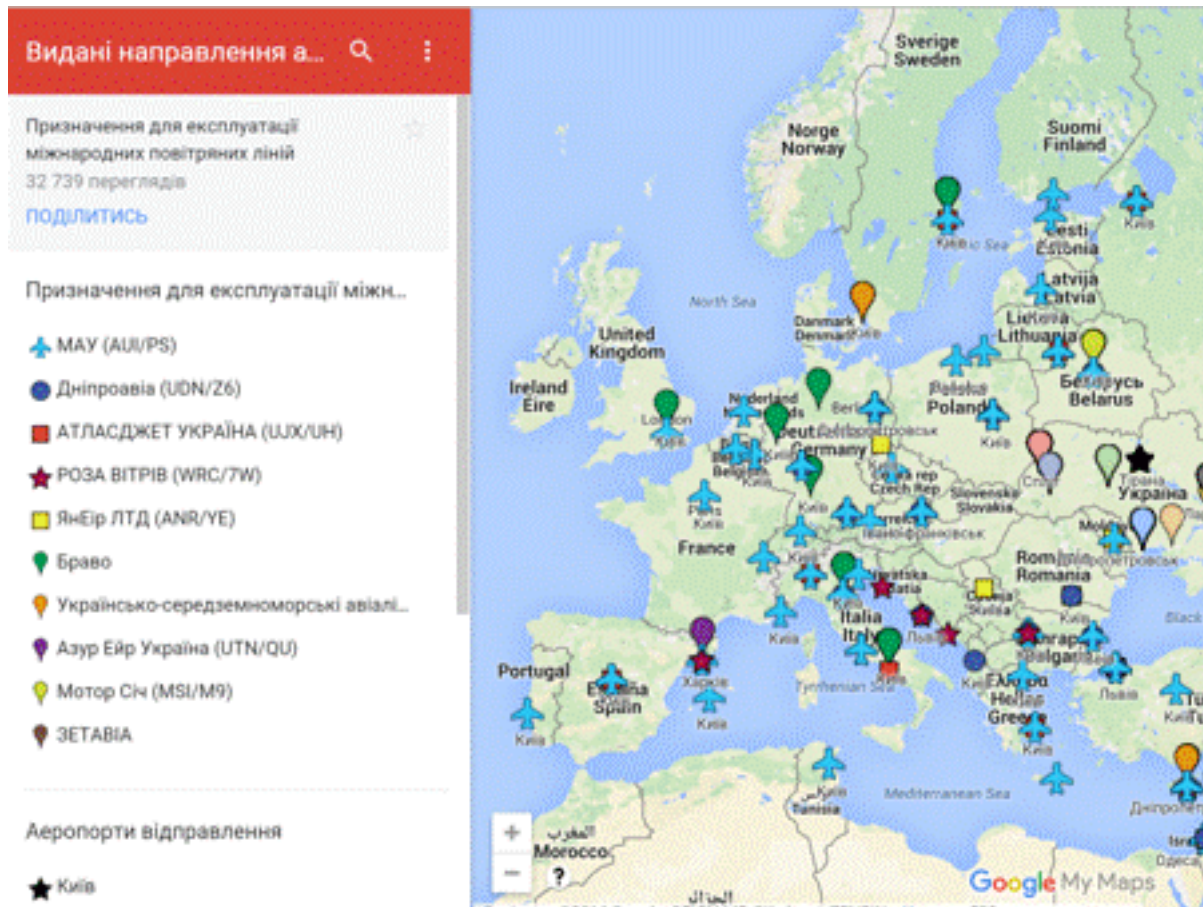
Інформаційна система управління майном територіальної громади міста Києва



## Видані напрямлення авіакомпаніям України

Посилання: [https://www.google.com/maps/d/viewer?mid=zQyzchol-IIQ.kqJvVjvXIY-s&hl=en\\_US](https://www.google.com/maps/d/viewer?mid=zQyzchol-IIQ.kqJvVjvXIY-s&hl=en_US)<sup>88</sup>

Призначення для експлуатації міжнародних повітряних ліній.



## Tesla Club Ukraine

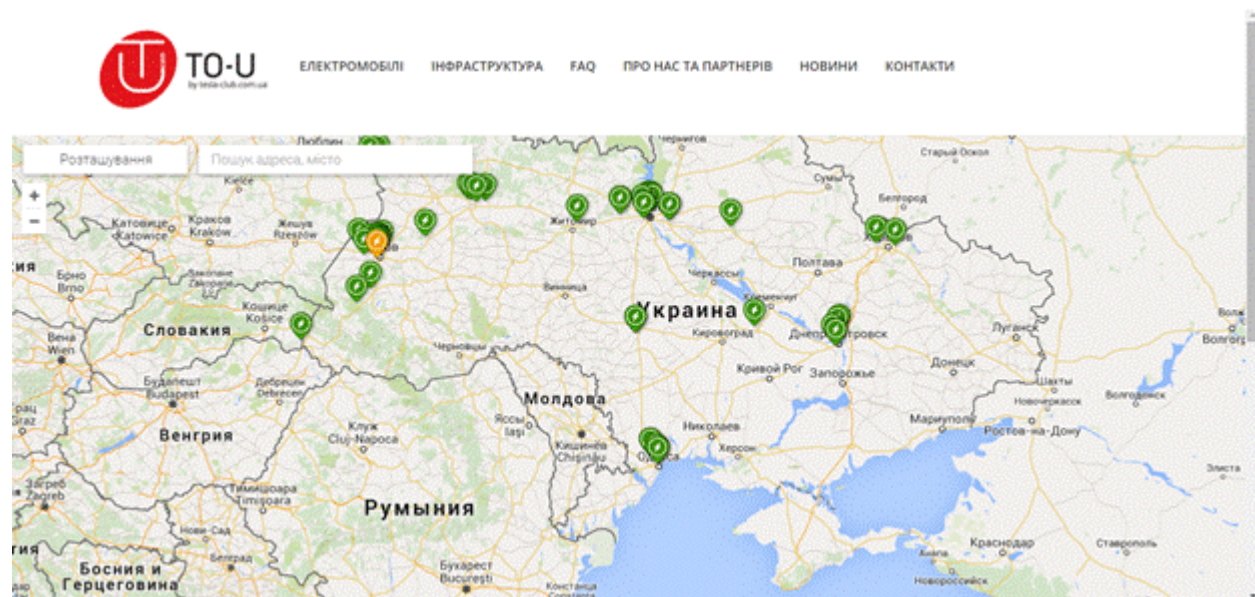
Посилання: <http://tesla-club.com.ua/gmap><sup>89</sup>

Мережа заправок для електромобілів.

<sup>88</sup>[https://www.google.com/maps/d/viewer?mid=zQyzchol-IIQ.kqJvVjvXIY-s&hl=en\\_US](https://www.google.com/maps/d/viewer?mid=zQyzchol-IIQ.kqJvVjvXIY-s&hl=en_US)

<sup>89</sup><http://tesla-club.com.ua/gmap>

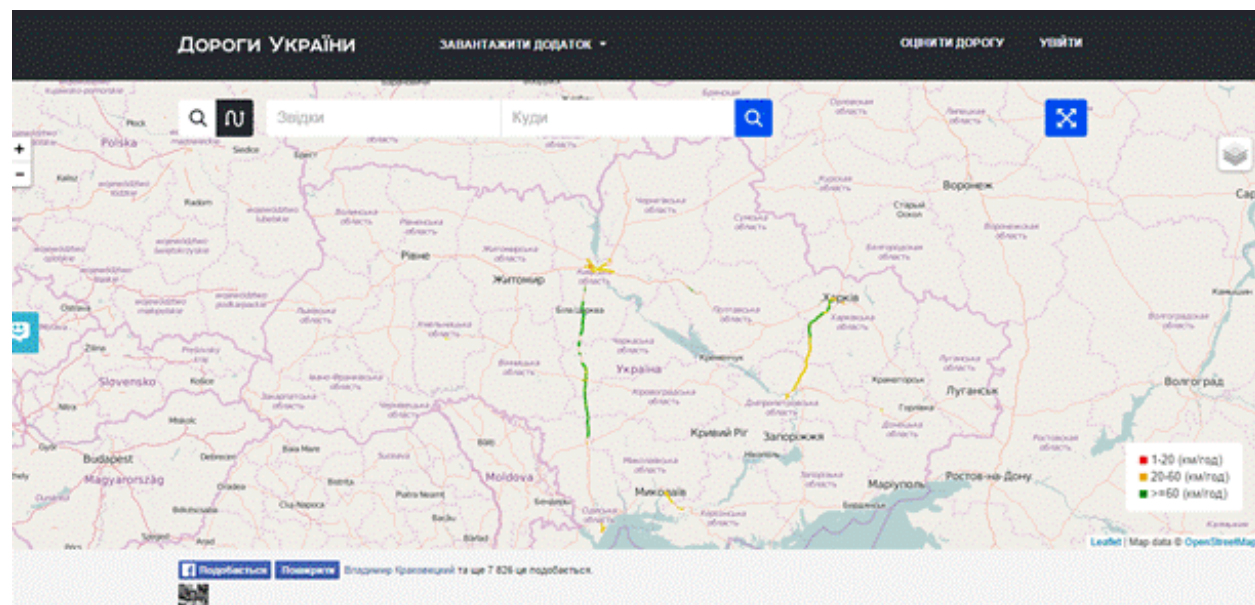




## Дороги України

Посилання: <http://uaroads.com/><sup>90</sup>

Інформація про стан доріг в Україні з можливістю оцінювати їх.



<sup>90</sup><http://uaroads.com/>

## Інші проекти

1. Додатки, сервіси та інші ресурси, що побудовані на базі відкритих даних ВРУ <http://data.rada.gov.ua/op>
2. Портал відкритих даних ВРУ <http://opendata.rada.gov.ua/><sup>92</sup>
3. Як голосують депутати <http://rada4you.org><sup>93</sup>
4. Мобільний додаток для прослідковування за законопроектами, поправками, прийняттям, зручні фільтри - «Zakonoprojekt»
5. Кнопкодави <http://knopkodavy.chesno.org/><sup>94</sup>
6. Візуалізація даних спільного ініціювання проектів законів народними депутатами <http://groups.chesno.org/><sup>95</sup>
7. Приймальні депутатів з контактами і годинами прийому <https://rada.oporaua.org/map/><sup>96</sup>, <https://pryjmalni.chesno.org><sup>97</sup>
8. Виборча варта <http://openvote.in.ua/><sup>98</sup>
9. Електронні вибори <https://e-vybory.org/><sup>99</sup>

## Інкубатор на базі відкритих даних 1991

1991 — перший в Україні некомерційний інкубатор, який допомагає перетворити тонни відкритих державних даних на реальні стартапи, що надають сервіси українським громадянам, підприємствам та державним органам.

Інкубаційні програми розраховані на інтердисциплінарні команди, що хочуть отримати дані від державного/корпоративного сектору та побудувати стартап на їх основі. Програми розділяють на групи:

- Галузеві рішення (для інфраструктури, агро, енергетики)
- Електронні послуги (для населення, в т.ч. на умовах ДПП)
- Аналітичні системи (для міністерств і органів влади)
- Smart-City рішення (для місцевих адміністрацій)

---

<sup>91</sup><http://data.rada.gov.ua/open/main/apps>

<sup>92</sup><http://opendata.rada.gov.ua/>

<sup>93</sup><http://rada4you.org>

<sup>94</sup><http://knopkodavy.chesno.org/>

<sup>95</sup><http://groups.chesno.org/>

<sup>96</sup><https://rada.oporaua.org/map/>

<sup>97</sup><https://pryjmalni.chesno.org>

<sup>98</sup><http://openvote.in.ua/>

<sup>99</sup><https://e-vybory.org/>

# Додатки

## Додаток 1. Рекомендовані інструменти та ресурси по візуалізації даних

Найменування інструменту	Посилання	Опис
visualizing.org	<a href="http://www.visualizing.org">http://www.visualizing.org</a> <sup>100</sup>	Онлайн інформаційна платформа про візуалізації даних.
Google Chart Tools	<a href="http://code.google.com/apis/chart/">http://code.google.com/apis/chart/</a> <sup>101</sup>	JavaScript API від Google для простого створення таблиць візуалізації для постійно змінюються даних.
GeoCommons	<a href="http://geocommons.org">http://geocommons.org</a> <sup>102</sup>	Інструментарій, спільнота і сервіс візуалізації для спільного використання геоданих.

<sup>100</sup><http://www.visualizing.org>

<sup>101</sup><http://code.google.com/apis/chart/>

<sup>102</sup><http://geocommons.com>

Найменування інструменту	Посилання	Опис
Quadrigram	<a href="http://www.quadrigram.com">http://www.quadrigram.com</a> <sup>103</sup>	Професійна платформа з можливістю платного побудови кастомізованих візуалізацій.
JavaScript InfoVis Toolkit	<a href="http://thejit.org">http://thejit.org</a> <sup>104</sup>	Javascript-інструментарій для створення і підтримки візуалізації різного роду графіків.
d3.js - Data-Driven Documents	<a href="http://mbostock.github.com/d3/">http://mbostock.github.com/d3/</a> <sup>105</sup>	Малює дуже гнучка JavaScript бібліотека з відкритим вихідним кодом для управління документами на основі даних (наприклад, для генерації HTML5 або SVG-діаграми з даних).

---

<sup>103</sup><http://www.quadrigram.com>

<sup>104</sup><http://thejit.org>

<sup>105</sup><http://mbostock.github.com/d3/>

Найменування інструменту	Посилання	Опис
Google Public Data Explorer	<a href="http://www.google.com/publicdata/home">http://www.google.com/publicdata/home</a> <sup>106</sup>	Каталог загального набору даних та інструмент для публікації і візуалізації великих наборів даних.
Maps Marker WP-Plugin	<a href="http://www.mapsmarker.com/wordpress/">http://www.mapsmarker.com/wordpress/</a> <sup>107</sup>	Wordpress. плагін для відображення карти з анотацією видатних пам'яток в блозі
DataMaps.eu - map your data	<a href="http://www.datamaps.eu/">http://www.datamaps.eu/</a> <sup>108</sup>	Wordpress. інструмент для створення привабливих карт візуалізації, які можуть бути створені в браузері через веб-сайт без знання програмування.

---

<sup>106</sup><http://www.google.com/publicdata/home>

<sup>107</sup><http://www.mapsmarker.com>

<sup>108</sup><http://www.datamaps.eu/>

Найменування інструменту	Посилання	Опис
Ushahidi	<a href="http://www.ushahidi.com">http://www.ushahidi.com</a> <sup>109</sup>	Відкрито програмне забезпечення для збору, візуалізації та інтер-активного відображення на основі визначення місця розташування даних в реальному часі (наприклад, від надзвичайних ситуацій, політичних виборів і т.д.).
Eclipse BIRT	<a href="http://www.eclipse.org/birt/phoenix/">http://www.eclipse.org/birt/phoenix/</a> <sup>110</sup>	Система звітності eclipse ( <a href="http://www.eclipse.org">http://www.eclipse.org</a> ) для створення візуально привабливих звітів великих обсягів даних.

---

<sup>109</sup><http://www.ushahidi.com>

<sup>110</sup><http://www.eclipse.org/birt/phoenix/>



Найменування інструменту	Посилання	Опис
Chartle.net	<a href="http://www.chartle.net">http://www.chartle.net</a> <sup>111</sup>	Безкоштовне інтерактивний онлайн-додаток по створенню графіків. Інтуїтивно зрозумілий інтерфейс, не вимагає спеціальних навичок, проте і набір можливостей обмежений. Застосовується, коли потрібен швидкий результат: круглі і стовпчасті діаграми, лінійні графіки, динамічні схеми, географічна карта двох видів. Підсумкова візуалізація інтерактивна, і її код легко вбудовується в html-сторінку.

---

<sup>111</sup><http://www.chartle.net>



Найменування інструменту	Посилання	Опис
Wordle	<a href="http://www.wordle.net">http://www.wordle.net</a>	Онлайн-додаток для аналізу і візуалізації частотності вживання слів у тексті. Безкоштовний.

---

<sup>114</sup><http://www.wordle.net>

Найменування інструменту	Посилання	Опис
Tableau	<a href="http://www.tableausoftware.com">http://www.tableausoftware.com</a> <sup>115</sup>	<p>програм для візуалізації даних. Асортимент опцій для кастомізації, а також можливість комбінувати кілька візуалізацій на одній панелі. За підсумками створення візуалізації можна отримати html-код, який можна вбудувати в веб-сторінку. Графіки інтерактивні. Закритий софт, однак, є безкоштовна версія з великою кількістю доступних можливостей (Tableau Public).</p>

<sup>115</sup><http://www.tableausoftware.com>

Dundas	<a href="http://www.dundas.com">http://www.dundas.com</a> <sup>116</sup>	Програмне забезпечення для створення інтерактивних візуалізацій. Може обробляти великі масиви даних. Створює візуалізації, в числі іншого, у вигляді панелей з декількох компонентів, що дозволяє одночасно представити кілька вимірів. Працює онлайн, комерційне, платне. Пропонують 45-денний безкоштовний випробувальний термін.
Leximancer	<a href="https://www.leximancer.com">https://www.leximancer.com</a> <sup>117</sup>	Професійна програма для аналізу тексту та візуалізації результатів цього аналізу. Комерційна, платна, кроссплатформенная.

---

<sup>116</sup><http://www.dundas.com>

<sup>117</sup><https://www.leximancer.com>

SIMILE Widgets	<a href="http://www.simile-widgets.org">http://www.simile-widgets.org</a> <sup>118</sup>	Збірка різноманітних віджетів і їх кодів. Коди відкриті з можливістю адаптувати під свої потреби, але для цього потрібні відповідні навички. Серед іншого, є інструменти, що дозволяють обробляти великі масиви даних і конструювати карти, тайм-лайни, інтерактивні таблиці і багато іншого. Інструмент Exhibit дозволяє створювати цілі інтерактивні веб-сторінки з можливістю пошуку і самотійного дослідження представленої бази даних.
GeoCommons	<a href="http://geocommons.com">http://geocommons.com</a> <sup>119</sup>	Безкоштовний веб-інструмент зі створення карт на основі даних.
Gephi	<a href="http://gephi.org">http://gephi.org</a> <sup>120</sup>	Програмне забезпечення для візуалізації графів. Використовується як один з інструментів аналізу соцмереж. Безкоштовний, відкритий код, кроссплатформлений.
Graphviz	<a href="http://www.graphviz.org">http://www.graphviz.org</a> <sup>121</sup>	Програма для візуалізації графів. Відкритий код, кроссплатформенна, безкоштовна.

---

<sup>118</sup><http://www.simile-widgets.org>

<sup>119</sup><http://geocommons.com>

<sup>120</sup><http://gephi.org>

<sup>121</sup><http://www.graphviz.org>

NewRadial	<a href="http://sourceforge.net/projects/newradial/">http://sourceforge.net/projects/newradial/</a> <sup>122</sup>	Комп'ютерні інструменти для візуального представлення нечислових даних (в тому числі зображень).
Prefuse	<a href="http://www.prefuse.org/">http://www.prefuse.org/</a> <sup>123</sup>	Великий набір різних інструментів для створення складних інтерактивних візуалізацій. Вимагають вміння програмувати. Безкоштовний, код відкритий.
ТЕКСТИ.ORG.UA	<a href="http://texty.org.ua/pg/blog/infviz/">http://texty.org.ua/pg/blog/infviz/</a> <sup>124</sup>	Ботів: Інфовіз: інфографіка, інтерактивна візуалізація даних
Charted	<a href="http://www.charted.co/">http://www.charted.co/</a> <sup>125</sup>	Швидка візуалізація CSV файлів
Venngage	<a href="https://venngage.com/">https://venngage.com/</a> <sup>126</sup>	Комерційний сервіс для створення інфографіки
Dipity	<a href="http://www.dipity.com/">http://www.dipity.com/</a> <sup>127</sup>	Створення гарних таймлайнів
Easily	<a href="https://piktochart.com/">https://piktochart.com/</a> <sup>128</sup>	Генератор інфографіки
Automatic Infographic Generator	<a href="http://petercv.com/aig/">http://petercv.com/aig/</a> <sup>129</sup>	Генератор інфографіки

## Додаток 2. Глосарій

**Абстрактна модель** - модель, що відображає загальні характеристики модельованого явища. Даний тип моделі представляє інформацію про якісні характеристики модельованого об'єкта чи явища.

**Відкрита ліцензія** - документ, зведення прав і обмежень даного об'єкту, який регламентує вільне поширення контенту та/або програмного забезпечення. Під свободою мається на увазі термін аналогічний визначенню свободи слова.

**Відкриті дані (Open Data)** – це систематизована інформація, доступна через Інтернет, у форматі, що дозволяє її автоматизоване оброблення електронними засобами, вільний та

<sup>122</sup><http://sourceforge.net/projects/newradial/>

<sup>123</sup><http://www.prefuse.org>

<sup>124</sup><http://texty.org.ua/pg/blog/infviz>

<sup>125</sup><http://www.charted.co/>

<sup>126</sup><https://venngage.com/>

<sup>127</sup><http://www.dipity.com/>

<sup>128</sup><https://piktochart.com/>

<sup>129</sup><http://petercv.com/aig/>

безоплатний доступ до неї, а також її подальше використання.

**Відкриті державні дані (Open Government Data)** – публічна інформація у вигляді відкритих даних про діяльність державних органів та органів місцевого самоврядування або створена в результаті їх діяльності.

**Власник інформації** - особа, яка самостійно створила інформацію або отримала на підставі закону або договору право дозволяти чи обмежувати доступ до інформації, яка визначається за будь-якими ознаками.

**Ідентифікатор набору відкритих даних (URN, Uniform Resource Name)** – це постійний строковий параметр у визначеному форматі, що використовується для швидкої ідентифікації та формування назв файлів набору, й включає інформацію, якщо є, про орган влади – розпорядника даних, назву системи або бази даних, з якої сформований набір даних, тип або групу даних, дату та час першої публікації, внутрішній номер або назву набору даних, дату та час оновлення.

**Інтерфейс прикладного програмування (API, Application Programming Interface)** – набір готових класів, функцій, структур і констант, доступних у вигляді сервісів через Інтернет, що можна використовувати у прикладних програмах.

**Машинозчитувані дані** - дані, представлені в форматах, придатних для автоматичного і автоматизованого використання.

**Метадані (метаінформація)** – це структуровані дані, що визначає характеристики відкритих дані для цілей їх ідентифікації та обробки.

**Набір відкритих даних (набір даних)** – це сукупність однорідних записів, що включають структуровану інформацію – поля даних та метаінформацію про них.

**Онтологія** - засіб зв'язування атрибутів наборів відкритих даних між собою, являє собою термінологічний словник синонімів/співвідношень.

**Першочерговий набір відкритих даних** - набір відкритих даних, визначений в якості першочергового для публікації і вдосконалення відповідно до методики моніторингу і оцінки затребуваності відкритих даних.

**Реєстр наборів відкритих даних** – це систематизований перелік наборів відкритих даних, що дозволяє в автоматичному режимі ідентифікувати та обробити наявні набори відкритих даних, отримати їх ключові параметри та адреси для доступу в Інтернет.

**Паспорт відкритих даних** – це сукупність відомостей про набір відкритих даних для його ідентифікації та обробки.

**Першочерговий набір відкритих даних** – це такий набір даних, який потребує найбільша кількість користувачів, та який має найбільшу значимість або популярність

**Пов'язані дані (Linked Data)** – це спосіб формування даних, що дозволяє інтерпретувати їх у сукупності з наборами відкритих даних.



**Портал відкритих даних (Data Hub)** – це комплексна інформаційна система, яка реалізує функцію необмеженого доступу до реєстру відкритих даних.

**Реєстр наборів відкритих даних** - систематизований перелік наборів відкритих даних, що дозволяє в автоматичному режимі здійснювати навігацію по наборам відкритих даних і отримувати ключові параметри цих наборів, включаючи гіперпосилання, необхідні для доступу до наборів відкритих даних.

**Розпорядник інформації** – це особа, яка самостійно створила інформацію або отримала на підставі закону чи угоди право використовувати інформацію, дозволяти або забороняти доступ до неї.

**Структура відкритих даних (Schema)** – це опис елементів та їх параметрів, з яких складаються відкриті дані.

**JSON** - (JavaScript Object Notation) - текстовий формат обміну даними, заснований на JavaScript і часто використовується саме з цією мовою. Як і багато інших текстові формати, JSON легко читається людьми.

**RDF** (Resource Description Framework) - модель, розроблена консорціумом W3C для представлення даних, особливо - метаданих. RDF подає відомості про земельні ресурси в придатному для машинної обробки вигляді.

**XML** - (eXtensible Markup Language - розширювана мова розмітки) - мова розмітки, рекомендований Консорціумом Всесвітньої павутини (W3C). Специфікація XML описує XML-документи і частково описує поведінку XML-процесорів (програм, які читають XML-документи і забезпечують доступ до їх вмісту).

# Посилання та література

1. Відео. Як поєднати технології і волонтерський рух? | Дмитро Чаплинський | TEDxKPI [https://www.youtube.com/watch?v=nR9CXA6\\_7-I](https://www.youtube.com/watch?v=nR9CXA6_7-I)<sup>130</sup>
2. Оцифрування статистики, або Перша їжа для Bigdata <http://www.epravda.com.ua/publications/2015/08/20/554624><sup>131</sup>
3. Данные нам: какие полезные сервисы Украина может получить от огромного массива своей информации <http://platfor.ma/magazine/text-sq/projects/open-data-incubator/><sup>132</sup>
4. Безопасность как услуга: уроки парижских терактов для украинских операторов <http://itc.ua/articles/bezopasnost-kak-usluga-uroki-parizhskih-teraktov-dlya-ukrainskih-operatorov><sup>133</sup>
5. How Ukraine is embracing open data after the 2014 revolution <http://theodi.org/blog/ukraine-embracing-open-data-2014-revolution><sup>134</sup>
6. Відкриті дані в Україні: що і як маємо відкрити? УКМЦ-04-03-2016 <https://www.youtube.com/watch?v=gyHEOpX1yes><sup>135</sup>
7. О развитии сферы открытых данных в Украине <https://habrahabr.ru/post/306414/><sup>136</sup>

---

<sup>130</sup>[https://www.youtube.com/watch?v=nR9CXA6\\_7-I](https://www.youtube.com/watch?v=nR9CXA6_7-I)

<sup>131</sup><http://www.epravda.com.ua/publications/2015/08/20/554624>

<sup>132</sup><http://platfor.ma/magazine/text-sq/projects/open-data-incubator/>

<sup>133</sup><http://itc.ua/articles/bezopasnost-kak-usluga-uroki-parizhskih-teraktov-dlya-ukrainskih-operatorov>

<sup>134</sup><http://theodi.org/blog/ukraine-embracing-open-data-2014-revolution>

<sup>135</sup><https://www.youtube.com/watch?v=gyHEOpX1yes>

<sup>136</sup><https://habrahabr.ru/post/306414/>