

# Відкриті дані в Україні:

Посібник для початківців

Олександр Краковецький

2017



# Зміст

<b>1</b>	<b>Вступ</b>	<b>1</b>
<b>2</b>	<b>Про стан розвитку відкритих даних в Україні</b>	<b>3</b>
2.1	Запущено єдиний державний портал відкритих даних data.gov.ua .	3
2.2	Запущено систему державних закупівель ProZorro . . . . .	3
2.3	Запущено офіційний портал публічних фінансів України Є-Data . .	4
2.4	Запущено пошуково-аналітичну систему 007 . . . . .	5
2.5	Нормативно-правове забезпечення . . . . .	5
2.6	Формати даних для оприлюднення у формі відкритих даних . . . .	6
<b>3</b>	<b>Що таке відкриті дані?</b>	<b>7</b>
3.1	Визначення відкритих даних . . . . .	7
3.2	Класифікація відкритих даних . . . . .	8
3.3	Формати даних . . . . .	9
3.3.1	Формати представлення даних через API . . . . .	10
3.3.2	Формати даних для роботи з геопросторовими даними . . . .	11
<b>4</b>	<b>Публікація та оновлення відкритих даних</b>	<b>13</b>
4.1	Вибір формату даних та конвертація даних з одного формату в інший	13
4.2	Основні помилки при оприлюдненні відкритих даних . . . . .	14
4.2.1	Чому Excel формат не є рекомендованим форматом? . . . . .	14
4.2.2	Оприлюднення занадто агрегованих даних . . . . .	15
4.3	Призначення відповідальних за розкриття даних . . . . .	16
4.4	Аудит наборів відкритих даних і пріоритетність їх публікації . . . .	16
4.5	Підготовка даних до публікації . . . . .	17
4.5.1	Кодування файлів . . . . .	17
4.5.2	Один файл чи декілька? . . . . .	17
4.5.3	Деперсоніфікація даних . . . . .	19
4.5.4	Архівація наборів даних . . . . .	19
4.5.5	Публікація даних, що мають періодичність . . . . .	19
4.5.6	Публікація схожих даних з різними структурами . . . . .	20
4.5.7	Публікація даних у вигляді API та робота з великими обсягами	20

4.5.8	Канали розповсюдження відкритих даних . . . . .	21
<b>5</b>	<b>Відкриті дані в Україні</b>	<b>23</b>
5.1	Законодавство України щодо відкритих даних та готовність країни до впровадження . . . . .	23
5.2	Україна у світових рейтингах . . . . .	23
5.2.1	Open Data Barometer . . . . .	23
5.2.2	Open Data Index . . . . .	24
5.2.3	The Web Index . . . . .	25
5.2.4	Open Budget Survey . . . . .	25
5.2.5	E-Government for the People . . . . .	26
5.3	Проекти на базі відкритих даних . . . . .	27
5.3.1	Нові назви вулиць, площ Дніпропетровська . . . . .	27
5.3.2	ДонорUA . . . . .	27
5.3.3	УкрТрансГаз . . . . .	27
5.3.4	НБУ . . . . .	28
5.3.5	3G.Multitest.ua . . . . .	28
5.3.6	Агенти змін . . . . .	29
5.3.7	Лун.UA . . . . .	29
5.3.8	PER.ORG.UA . . . . .	30
5.3.9	Посіпаки . . . . .	30
5.3.10	CityScale . . . . .	31
5.3.11	ARBUZ . . . . .	31
5.3.12	KyivSmartCity . . . . .	32
5.3.13	Видані направлення авіакомпаніям України . . . . .	32
5.3.14	Tesla Club Ukraine . . . . .	33
5.3.15	Дороги України . . . . .	33
5.3.16	Інші проекти . . . . .	34
5.3.17	Інкубатор на базі відкритих даних 1991 . . . . .	34
<b>6</b>	<b>Ресурси</b>	<b>35</b>
<b>7</b>	<b>Глосарій</b>	<b>41</b>
<b>8</b>	<b>Посилання та література</b>	<b>43</b>

# 1. Вступ

**Мета курсу** — ознайомити представників державних структур та громадського сектору (активістів, членів громадських організацій, журналістів тощо) з основними поняттями, принципами та підходами до роботи з даними. Курс допоможе підготувати та оприлюднити дані у форматах, що відповідають принципам відкритості та зручності обробки програмними засобами. Ви дізнаєтесь про переваги найбільш популярних форматів відкритих даних, міжнародну практику роботи з даними, а також особливості роботи з відкритими даними в Україні.

**Зверніть увагу:**

Ці методичні матеріали також будуть корисними для представників ІТ-сектору та бізнесу, які планують аналізувати публічні дані або працювати із суспільно-корисними ініціативами.



## **2. Про стан розвитку відкритих даних в Україні**

### **2.1 Запущено єдиний державний портал відкритих даних data.gov.ua**

Перша версія порталу була розроблена волонтерами із організації Socialboost за підтримки міжнародних організацій і компанії Microsoft. На сьогоднішній день портал переданий на баланс Державного агентства з питань електронного урядування. На порталі доступно майже 6000 наборів даних від 700+ розпорядників, хоча якісних наборів даних у відсотковому співвідношенні не дуже багато.

### **2.2 Запущено систему державних закупівель ProZorro**

ProZorro – електронна система публічних закупівель яка прийшла на зміну паперовим держтендерам.

У Законі України «Про публічні закупівлі» передбачається:

- Запровадження обов’язковості проведення процедур через електронну систему. (Перший етап — обов’язковість проведення процедур через електронну систему поширюється на головних розпорядників коштів та монополістів (з 1 квітня 2016 року), на другому (з 1 серпня) – на всіх замовників;
- Запровадження електронного аукціону, який передбачає автоматичну оцінку тендерних пропозицій;
- Визначення нових понять «авторизований електронний майданчик», «електронна система закупівель», «централізована закупівельна організація», «система хмарних обчислень»;
- Замість 5 процедур залишити 3 (відкриті торги, конкурентний діалог, переговорна процедура);

- Зміна термінології, зокрема, замість терміну «державна закупівля» вводиться термін «публічна закупівля»; замість термінів «конкурс», «документація конкурсних торгів», «пропозиція конкурсних торгів», «комітет з конкурсних торгів», вводяться поняття «тендер», «тендерна документація», «тендерна пропозиція», «тендерний комітет».

ProZorro отримала міжнародну премію у сфері публічних закупівель Public Sector Procurement Award за створення і впровадження електронної системи з унікальною архітектурою. Розробка цієї системи на базі програмного забезпечення з відкритим кодом була здійснена у партнерстві влади, бізнесу та громадськості та адмініструвалася антикорупційною організацією Transparency International Україна.

З 1 серпня 2016 року ProZorro є обов'язковою системою для всіх державних замовників при закупівлі від 200 тис. грн для товарів і від 1,5 млн грн для робіт.

## 2.3 Запущено офіційний портал публічних фінансів України Є-Data

Є-Data — це офіційний державний інформаційний портал у мережі Інтернет, на якому оприлюднюється інформація про використання публічних коштів та реалізується ідея «Прозорого бюджету». Задача - забезпечити повну прозорість державних фінансів та задовольнити право громадськості на доступ до інформації.

Основні законодавчі та нормативні засади для створення проекту:

- Закон України «Про відкритість використання публічних коштів»
- План заходів з виконання Програми діяльності Кабінету Міністрів України та Стратегії сталого розвитку "Україна - 2020" у 2015 році п. 95: «Впровадження інтегрованої інформаційно-аналітичної системи "Прозорий бюджет" з метою забезпечення доступності інформації про державні фінанси для суспільних потреб із забезпеченням відкритої звітності за всіма коштами, використаними отримувачами бюджетних коштів..»
- Коаліційна угода Парламенту VIII скликання п. 2.5.10.: «Запровадження системи «Прозорий бюджет» з метою забезпечення доступності інформації про державні фінанси для суспільних потреб.

З 15 вересня 2015 року на порталі оприлюднюються всі трансакції Державної казначейської служби, з листопада на ньому доступна інформація про використання коштів державного і місцевих бюджетів, а у січні інформацію почали розкривати суб'єкти господарювання державної і комунальної власності, у статутному



капіталі яких державна або комунальна частка акцій (часток, паїв) перевищує 50 відсотків.

До 2018 року планується забезпечити повний функціонал Інтегрованої інформаційно-аналітичної системи «Прозорий бюджет», яка своєю чергою торкнеться змін бюджетних процесів Міністерства фінансів, автоматизації систем ДФС та Казначейства, автоматизації систем обліку та звітності на місцевих рівнях.

## 2.4 Запущено пошуково-аналітичну систему 007

Пошуково-аналітична система 007 – це веб-сервіс на основі відкритих даних про використання публічних коштів. Проект передбачає сервіс пошуку та візуалізації даних з відкритих джерел про використання державою бюджетних коштів. Його ідея – дати громадськості інструмент контролю влади та можливість стежити за бюджетними витратами, збирати докази зловживань і швидко переводити боротьбу з корупцією в правове поле. Основний акцент зроблено на простоті використання та представлення специфічної інформації з масивів великих даних. Сайт був відкритий 8 квітня 2016 року.

## 2.5 Нормативно-правове забезпечення

9 квітня 2015 року Верховна Рада України ухвалила Закон України № 319 «Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних». Зазначеним Законом внесені зміни до Закону України «Про доступ до публічної інформації» з метою визначення базових норм та засад розвитку відкритих даних в Україні, а саме:

1. 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;
2. Розпорядники інформації зобов'язані надавати публічну інформацію у формі відкритих даних на запит, оприлюднювати і регулярно оновлювати її на єдиному державному веб-порталі відкритих даних та на своїх веб-сайтах;
3. Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації.

21 жовтня 2015 року затверджено Постановою КМУ № 835 «Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних», якою визначено вимоги до формату і структури наборів даних, а також затверджено перелік пріоритетних наборів даних, які підлягають оприлюдненню (більше 300 наборів). Постановою чітко визначений перелік форматів для оприлюднення відкритих даних в залежності від їх виду:

## 2.6 Формати даних для оприлюднення у формі відкритих даних

**Текстові дані** TXT, RTF, ODT, DOC(X), PDF (с текстовим змістом, скановане зображення), (X)HTML

**Структуровані дані** RDF, XML, JSON, CSV, XLS(X), ODS, YAML

**Графічні дані** GIF, TIFF, JPG (JPEG), PNG

**Відеодані** MPEG, MKV, AVI, FLV, MKS, MK3D

**Аудіодані** MP3, WAV, MKA

**Дані, розроблені з використанням програми Macromedia Flash** SWF, FLV

**Архіви даних** ZIP, 7z, Gzip, Bzip2

Також, постановою передбачено, що державні реєстри, які постійно оновлюються, мають бути відкриті за допомогою API.

З метою забезпечення ефективного функціонування Єдиного державного веб-порталу відкритих даних та підвищення відкритості і прозорості діяльності органів виконавчої влади та місцевого самоврядування, Державним агентством з питань електронного урядування в лютому 2016 року розроблено проект постанови Кабінету Міністрів України «Деякі питання оприлюднення публічної інформації у формі відкритих даних», які знаходяться зараз на етапі обговорення.

## 3. Що таке відкриті дані?

### 3.1 Визначення відкритих даних

**Відкриті дані (англ. Open Data)** – це концепція, яка відображає 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.

## 3.2 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/>), 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.

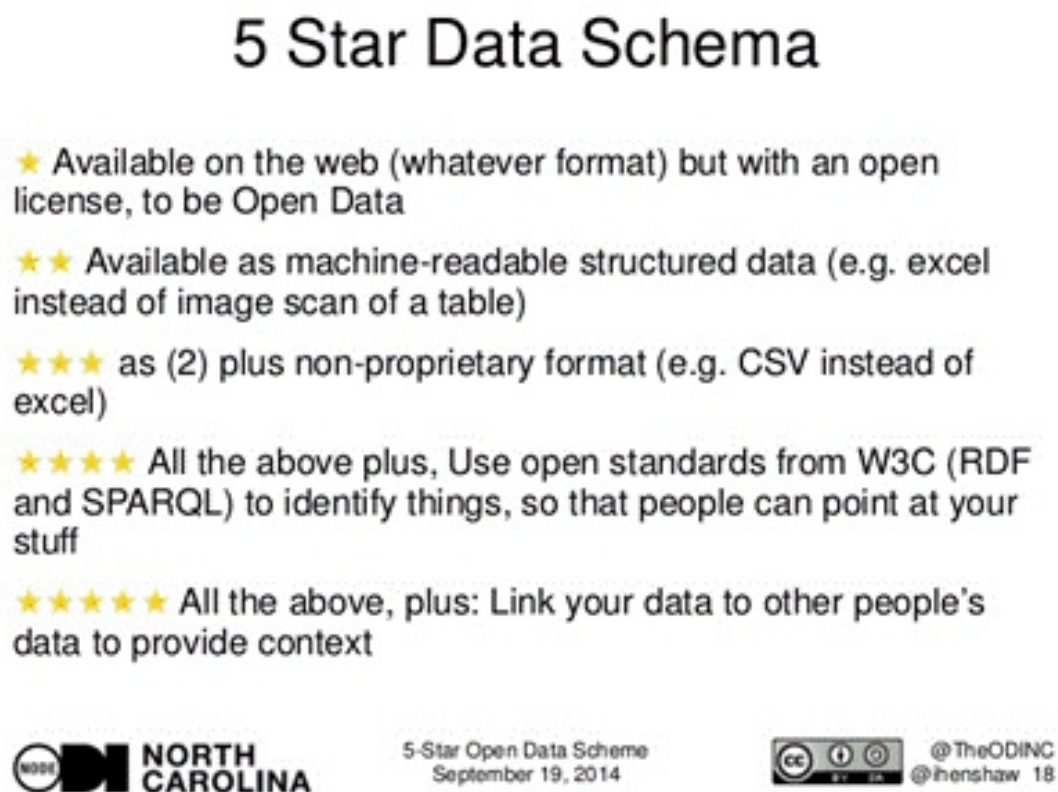


Рис. 3.1: 5 зірок Open Data

- **Одну зірку (\*)** отримує будь-яка інформація, вільно доступна через Інтернет в будь-якому форматі. Під цю класифікацію підпадає файл у форматі PDF або інша (сканована) копія документу, на який веде пряме посилання на офіційному сайті державного органу.
- **Дві зірки (\*\*)** отримує структурована інформація, яку можна обробляти автоматично, наприклад, у форматах для веб-браузерів чи офісних програм

(відкриті формати – TXT, HTML, RSS; пропрієтарні формати, Excel – XLS, Word – DOC, RTF).

- **Три зірки (\*\*\*)** може отримати інформація, представлена у відомих, добре описаних відкритих структурованих форматах (наприклад, CSV, JSON, XML, YAML).
- **Чотири зірки (\*\*\*\*)** надаються у випадку, якщо можна отримати первинні необроблені набори відкритих даних у вигляді файлів або фільтровані дані у запиті до API за вказаними параметрами.
- **П'ять зірок (\*\*\*\*\*)** надається інформації, коли набори відкритих даних пов'язані між собою і представляють собою семантичну мережу.

### 3.3 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.

#### 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.

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

**XLS/XLSX (Document Office Open XML)** Файл XLSX - електронна таблиця, створена в Microsoft Excel - додатку для роботи з таблицями.

#### Графічні формати

**JPEG (Joint Photographic Experts Group)** JPEG (англ. *Joint Photographic Experts Group*) - один з популярних графічних форматів, застосовуваний для зберігання фотозображень.

**PNG (Portable Network Graphics)** PNG (англ. *Portable network graphics*) - растровий формат зберігання графічної інформації, що використовує стиснення без втрат.

**TIFF (Tagged Image File Format)** TIFF (англ. *Tagged Image File Format*) - формат зберігання растрових графічних зображень.

### Текстові формати

**TXT (Textfile)** TXT - це формат, що містить текстові дані, які, як правило, організовані у вигляді рядків.

**Markdown** Markdown - полегшена мова розмітки, створена з метою написання максимально читабельного і зручного для редагування тексту.

### Текстово-графічні формати

**HTML (HyperText Markup Language)** HTML (від англ. *HyperText Markup Language*) - стандартна мова розмітки документів у Всесвітній павутині.

**DOCX (Document Office Open XML)** DOCX (*Document Office Open XML*) - формат файлу для зберігання електронних документів пакетів офісних додатків.

**OpenDocument format (LibreOffice/OpenOffice)** Формат OpenDocument - це формат для текстових, табличних документів та презентацій.

**PDF (Portable Document Format)** Формат переносного документа (PDF) - це формат файлу, який використовується для надійного уявлення і обміну документами.

### 3.3.1 Формати представлення даних через API

**JSON (JavaScript Object Notation)** JSON (від англ. *JavaScript Object Notation*) – текстовий відкритий формат, оснований на Javascript представлені.

```
{
  "firstName": "Иван",
  "lastName": "Иванов",
  "address": {
    "streetAddress": "Московское ш., 101, кв.101",
    "city": "Ленинград",
```

```
    "postalCode": 101101
  },
  "phoneNumbers": [
    "812 123-1234",
    "916 123-4567"
  ]
}
```

**XML (eXtensible Markup Language)** XML (від англ. *eXtensible Markup Language*)

– мова розмітки, що розширюється.

```
<person firstName="Иван" lastName="Иванов">
  <address streetAddress="Московское ш., 101, кв.101"
    city="Ленинград" postalCode="101101"/>
  <phoneNumbers>
    <phoneNumber>812 123-1234</phoneNumber>
    <phoneNumber>916 123-4567</phoneNumber>
  </phoneNumbers>
</person>
```

### 3.3.2 Формати даних для роботи з геопросторовими даними

**GeoJSON та TopoJSON** GeoJSON - відкритий формат, призначений для зберігання графічних структур даних.

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {"type": "Point", "coordinates": [102.0, 0.5]},
      "properties": {"prop0": "value0"}
    },
    {
      "type": "Feature",
      "geometry": {
        "type": "LineString",
        "coordinates": [
          [102.0, 0.0], [103.0, 1.0], [104.0, 0.0], [105.0, 1.0]
        ]
      }
    }
  ]
}
```

```
    },
    "properties": {
      "prop0": "value0",
      "prop1": 0.0
    }
  },
  {
    "type": "Feature",
    "geometry": {
      "type": "Polygon",
      "coordinates": [
        [ [100.0, 0.0], [101.0, 0.0], [101.0, 1.0],
          [100.0, 1.0], [100.0, 0.0] ]
      ]
    },
    "properties": {
      "prop0": "value0",
      "prop1": {"this": "that"}
    }
  }
]
}
```



## 4. Публікація та оновлення відкритих даних

### 4.1 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. **Неструктуровані текстові дані** (тексти законів, розпоряджень, довідкових даних):

- Найкращий формат — Markdown або TXT.
- Markdown підтримує додаткову розмітку (заголовки, абзаци, посилання, зображення тощо).
- Вказуйте різновид Markdown.
- HTML — гірший варіант, doc/docx, pdf/tiff — найгірші (окрім випадків, коли потрібна копія оригінального документу із підписом та печаткою).

2. **Табличні дані:**

- Використовуйте формат CSV, у гіршому випадку — XLS/XLSX.

3. **Скани документів:**

- Найкращий формат — TIFF, потім — PDF.

4. **Зображення:**

- Обов'язкова роздільна здатність для OCR — не менше 300 dpi.

5. **Великі файли або групи файлів:**

- Архівуйте у форматі ZIP/7Z.
- Для файлів понад 4 ГБ — RAR.

#### 6. Дані, що змінюються постійно і формуються «на льоту»:

- Не обов'язково зберігати у вигляді файлів.
- Достатньо надати доступ через Інтернет-адреси або API.
- Рекомендовані формати для API: JSON/XML/RDF.

#### 7. Бінарні формати:

- Використовуйте лише за відсутності альтернативи.

Для конвертації даних існує багато інструментів: онлайн-сервіси, програмні засоби, бібліотеки. Зазвичай немає проблем з конвертацією між відкритими форматами. Проблеми можуть виникати з Word/PDF.

##### Деякі програмні засоби для конвертації:

- Microsoft Word, Excel, Access
- Графічні формати та PDF: Adobe Photoshop, Acrobat Reader, Foxit PDF Reader, Paint.NET, Corel Draw
- Текстові дані: Notepad++
- Програмні засоби: SPSS, Microsoft Azure ML, Matlab
- Онлайн інструменти:
  - JSON to CSV
  - CSV to JSON
  - JSON to C#
- Очищення HTML документів від Word-стилів: [word2cleanhtml.com](http://word2cleanhtml.com)

## 4.2 Основні помилки при оприлюдненні відкритих даних

### 4.2.1 Чому Excel формат не є рекомендованим форматом?

В Постанові дозволяється використовувати XLS/XLSX (Microsoft Excel). Проте Excel дозволяє додавати розширене форматування, об'єднані комірки, вбудовані об'єкти, макроси та формули, що ускладнює машинну обробку.

#### Приклади неправильних Excel-документів:

#### Приклади правильних Excel-файлів:

Дані у правильному форматі легко експортувати у CSV.

Expense	Project	Code	Amount ex GST	Discount	Total Discount	Total	GST	Total Inc GST
Petty Cash	Alpha	AD785	\$ 52.83	0%	\$ -	\$ 52.83	\$ 7.92	\$ 60.75
Insurance	Alpha	TR555	\$ 45.00	0%	\$ -	\$ 45.00	\$ 6.75	\$ 51.75
Meals/ Snacks	Alpha	AD203	\$ 132.75	0%	\$ -	\$ 132.75	\$ 19.91	\$ 152.66
Car rental	Beta	AD985	\$ 150.00	5%	\$ 7.50	\$ 142.50	\$ 21.38	\$ 163.88
Accommodation	Beta	TR256	\$ 560.00	10%	\$ 56.00	\$ 504.00	\$ 75.60	\$ 579.60
Airfares	Beta	TR256	\$ 258.00	0%	\$ -	\$ 258.00	\$ 38.70	\$ 296.70
<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>

Рис. 4.1: Неправильний Excel 1

State	Utilities Company	data1	data2	data3
Alabama	company1	3.458	1.734	0.584
Alabama	company2	20.675	20.675	20.675
Alaska	company3	13.008	13.008	13.008
Alaska	company4	12.1968	12.1968	12.1968
Arizona	company5	16.698		12.35
Arizona	company6	29.56		8.06
Arizona	company7	15.44		4.47
Arkansas	company8	1.576		1.125
Arkansas	company9	15.45	10.43	7.026
California	company10	19.48		8.48
California	company11	24.3		13.28
California	company12	16.06	8.14	4.655
California	company13	9.35		1.4
California	company14	23.67		6.24
California	company15	36		16
California	company16	34.91	18.99	5.78
California	company17	31	15	9
		16		14
		18.3	17.3	14.37

Рис. 4.2: Неправильний Excel 2

### 4.2.2 Оприлюднення занадто агрегованих даних

Очікується:

- ділянка та її координати
- дата ДТП
- кількість постраждалих та тип ушкоджень
- область (місто)
- стан ділянки

Неправильно:

- лише кількість ДТП по областях

Розпорядники повинні публікувати детальну інформацію на рівні транзакцій, випадків або подій.

	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									

Рис. 4.3: Правильний Excel 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	\$7,421	1/16/2012	Canada		

Рис. 4.4: Правильний Excel 2

### 4.3 Призначення відповідальних за розкриття даних

Кожен розпорядник повинен визначити відповідальних осіб за аудит, обробку та публікацію відкритих даних. Вони повинні знати про доступні набори, формати, розміри, роботу з базами даних та користуватись порталом відкритих даних.

### 4.4 Аудит наборів відкритих даних і пріоритетність їх публікації

Публікація даних здійснюється поетапно, враховуючи:

1. Потребу споживачів (за методикою моніторингу)
2. Ступінь готовності (наявність даних, технічна готовність)
3. Витрати на публікацію та підтримку

#### Фактори для формування реєстру:

- Публікуються дані без попередньої обробки



## 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.

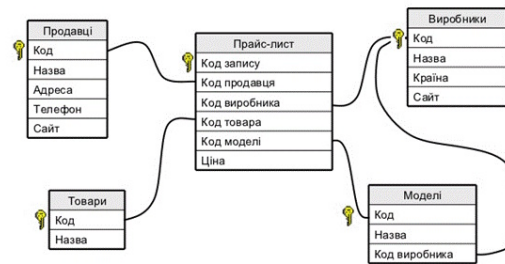


Рис. 4.6: Схема бази

*Models (моделі)*

Id	Name	Vendor Code
1	Surface	1
2	iPhone	2

**Варіант 1: Один файл**

models.csv:

Id;Name;Vendor;Country;Website

1;Surface;Microsoft;USA;http://microsoft.com

2;iPhone;Apple;USA;http://apple.com

**Варіант 2: Два файли**

vendors.csv:

Id;Name;Country;Website

1;Microsoft;USA;http://microsoft.com

2;Apple;USA;http://apple.com

models.csv:

Id;Name;VendorId

1;Surface;1

2;iPhone;2

**Порівняння:**

Критерій	Один файл	Багато файлів
Збитковість даних	Велика	Мала
Фінальний розмір	Більший	Менший
Перевірка на цілісність	Складна	Проста
Готовність до використання через API	Погана	Хороша
Кількість запитів	Один (перевага для малих файлів)	Багато

Другий варіант доцільніший.

### 4.5.3 Деперсоніфікація даних

Згідно Закону “Про захист персональних даних”, персональні дані — будь-які відомості, за якими ідентифікується особа. Такі дані мають бути деперсоніфіковані перед публікацією.

**Приклад:**

*Початкові дані:*

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

*Після деперсоніфікації:*

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

### 4.5.4 Архівація наборів даних

Архівація зменшує розмір файлів і трафік. Текстові дані стискаються до 90%, Word/Excel/PDF — на 10-60%, зображення — майже не стискаються.

**Архівувати потрібно:**

- історичні дані
- файли понад 50 МБ
- застарілі версії наборів
- багатотомні набори

**Рекомендовані формати:** zip/7z/tar. **Рекомендована програма:** 7-zip

### 4.5.5 Публікація даних, що мають періодичність

Набори поділяються на:

- **Частопубліковані** (оновлення частіше 1 раз на тиждень)

- **Рідкопубліковані** (рідше 1 раз на тиждень)

В паспорті набору вказується дата останнього оновлення.

**Частота оновлень:**

- Часто: більше одного разу в день, щодня, щотижня
- Рідко: щомісяця, щокварталу, кожні півроку, щороку, по мірі змін

**Приклади іменування файлів:**

- За датою: `exchange_20160101.csv`
- За валютою: `exchange_usd.csv`

Якщо дані змінюються постійно — організуйте доступ через API.

#### 4.5.6 Публікація схожих даних з різними структурами

Якщо структура схожа — об'єднайте всі поля, відсутні дані залишайте порожніми.

**Приклад:**

A1,A2,A3,A4

1,1,1,1

A1,A2,A3,A5

2,2,2,2

**Фінальний файл:**

A1,A2,A3,A4,A5

1,1,1,1,

2,2,2,,2

Якщо структура відрізняється значно — розглядайте файли окремо.

#### 4.5.7 Публікація даних у вигляді API та робота з великими обсягами

Два основних способи доступу:

- **Data Hub** — для рідко змінюваних даних
- **API** — для оперативних запитів

Рекомендується використовувати OData для API.

В описі набору вказуйте тип даних, посилання на документацію, умови доступу, обмеження, формати, підтримку OData.



#### **4.5.8 Канали розповсюдження відкритих даних**

1. Через сайт `data.gov.ua`
2. Через сайти місцевих органів самоврядування
3. За допомогою API
4. Через ftp-сервер
5. Через BitTorrent



## 5. Відкриті дані в Україні

### 5.1 Законодавство України щодо відкритих даних та готовність країни до впровадження

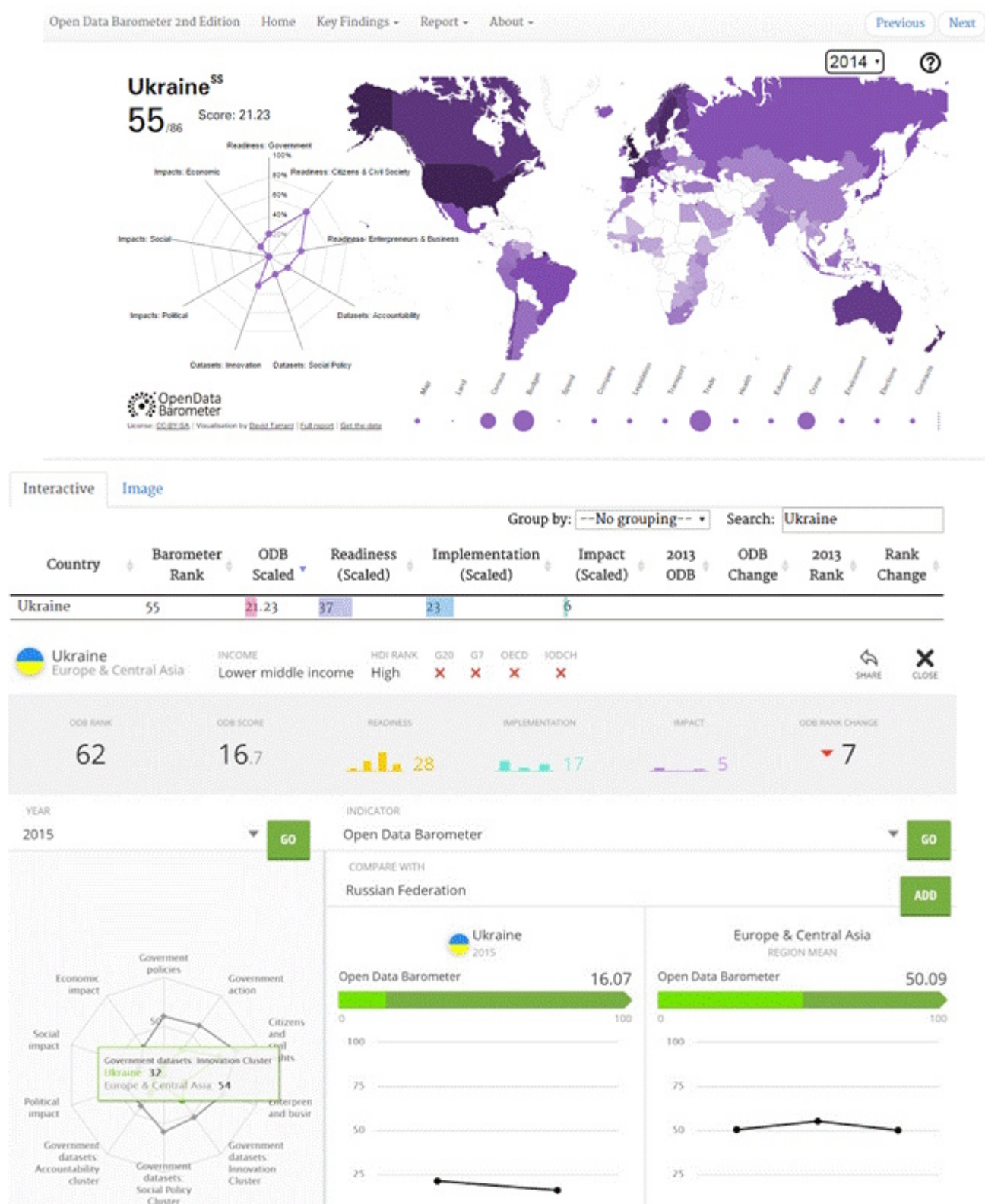
1. Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних № 835 (редакція від 21.10.2015)
2. План заходів (дорожня карта) розвитку відкритих даних на 2016 рік
3. Єдиний державний портал відкритих даних data.gov.ua — Керівництво користувача
4. Єдиний державний портал відкритих даних data.gov.ua — Робота з API
5. Open Data Readiness Assessment (ODRA) Ukraine Final Report

### 5.2 Україна у світових рейтингах

#### 5.2.1 Open Data Barometer

- Сайт: [opendataresearch.org/barometer](http://opendataresearch.org/barometer)

Open Data Barometer — спільний дослідницький проєкт Open Data Institute та World Wide Web Foundation. Аналізує глобальні тенденції розвитку відкритих даних, надає порівняльні дані по країнах і регіонах. Україна займає 62 місце із 86 проаналізованих країн (2015 — 55 місце).

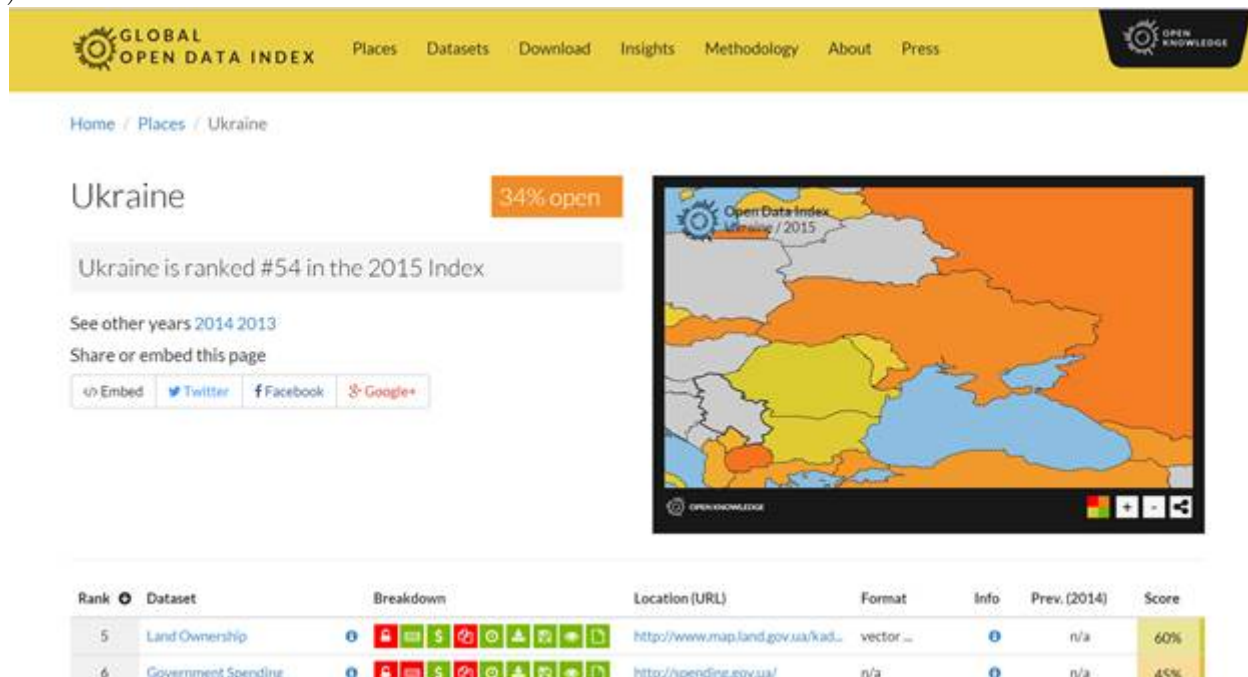


### 5.2.2 Open Data Index

- Сайт: [index.okfn.org](http://index.okfn.org)

Масштабний дослідницький проєкт Open Knowledge Foundation щодо відкритості державних даних у світі. Україна — 54 місце у глобальному рейтингу (2015

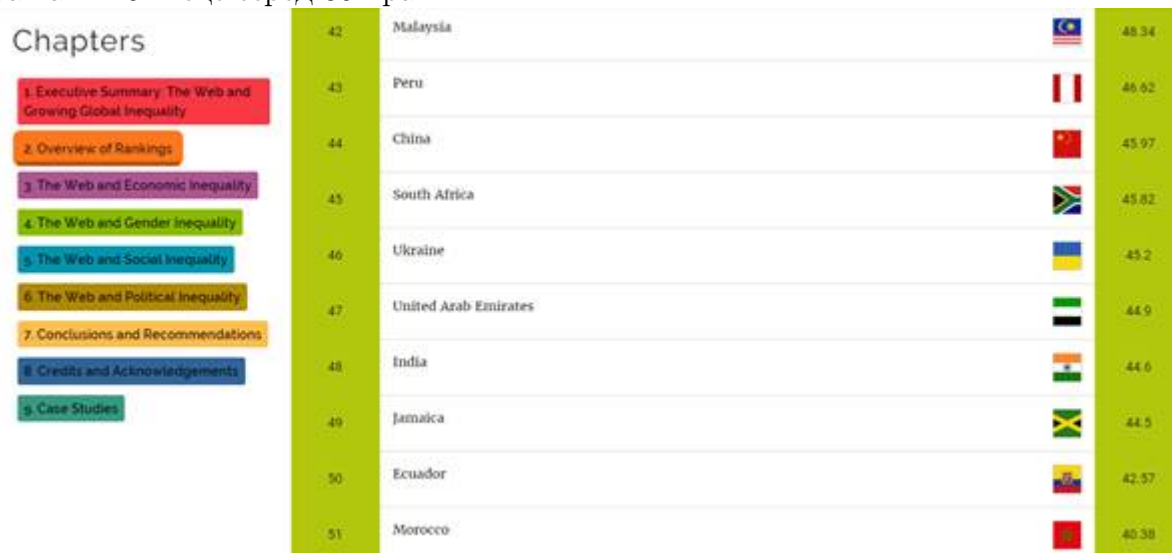
рік).



### 5.2.3 The Web Index

- Сайт: [thewebindex.org](http://thewebindex.org)

Рейтинг розвитку інтернету в країнах світу від World Wide Web Foundation. Україна — 46 місце серед 86 країн.



### 5.2.4 Open Budget Survey

- Сайт: [internationalbudget.org/what-we-do/open-budget-survey/](http://internationalbudget.org/what-we-do/open-budget-survey/)

Рейтинг відкритості державних бюджетів від International Budget Partnership. Оцінки для України:

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



### 5.2.5 E-Government for the People

- Сайт: [unpan.org/egovkb/global\\_reports/08report.html](http://unpan.org/egovkb/global_reports/08report.html)

Рейтинг ООН з розвитку електронного уряду. Коефіцієнт розвитку електронного уряду України — між 0,5 та 0,75.

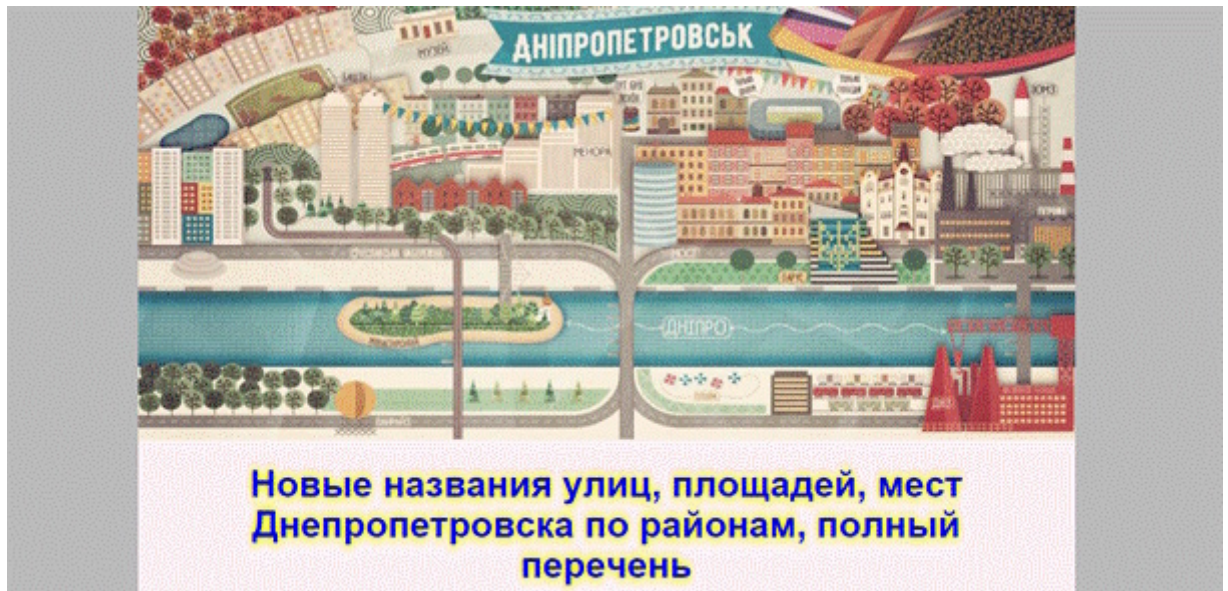
[illegible]



## 5.3 Проєкти на базі відкритих даних

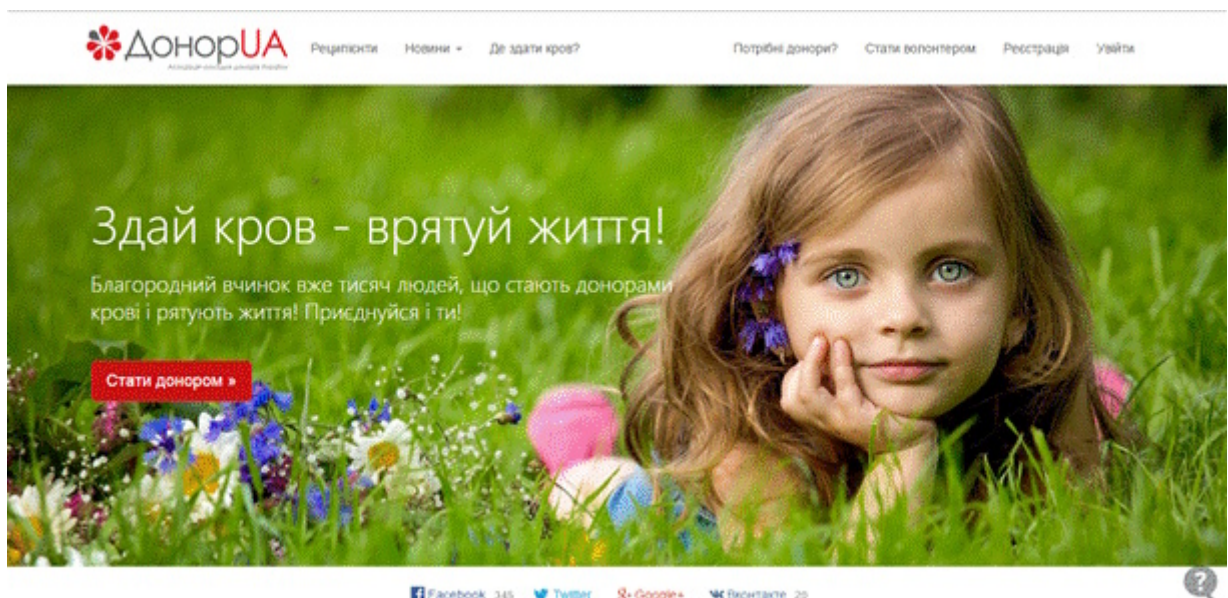
### 5.3.1 Нові назви вулиць, площ Дніпропетровська

- [rename.dp.ua](http://rename.dp.ua) — повний перелік нових назв вулиць і площ Дніпропетровська.



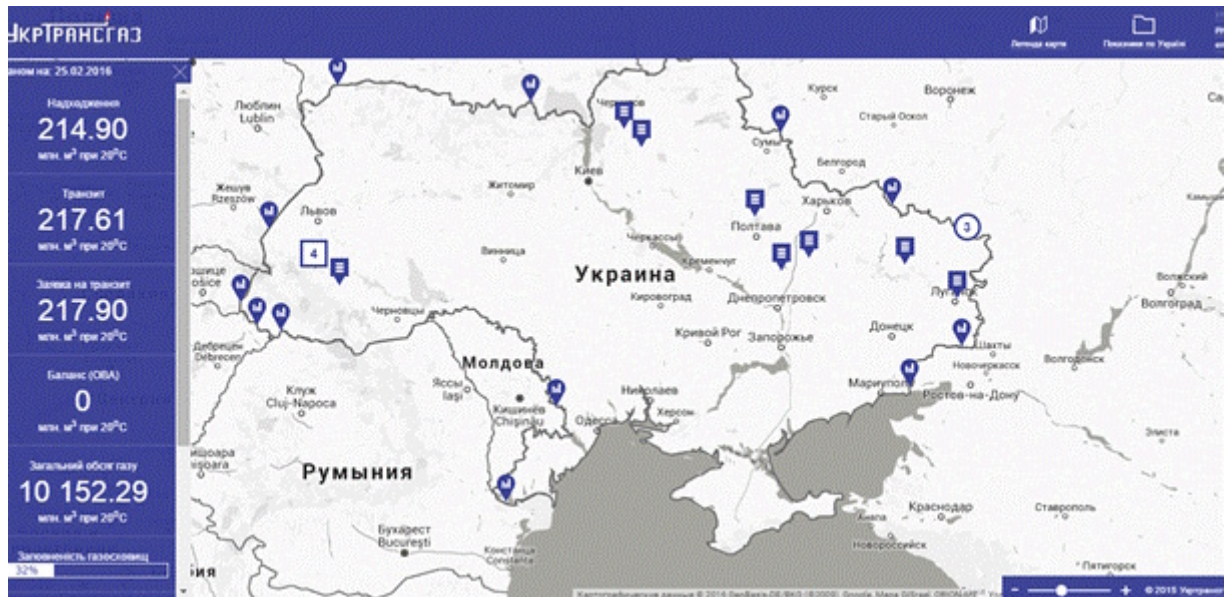
### 5.3.2 ДонорUA

- [donor.ua](http://donor.ua) — координація донорів крові та пропаганда донорства в Україні.



### 5.3.3 УкрТрансГаз

- [utg.ua/live/](http://utg.ua/live/) — карта трубопроводів, сховищ та обсягів газу.



### 5.3.4 НБУ

- nbu.rocks — відкриті дані Національного Банку України про банки.

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

\*А саме: Структури власності, Банківські ліцензії та види діяльності, Дані про структурні підрозділи

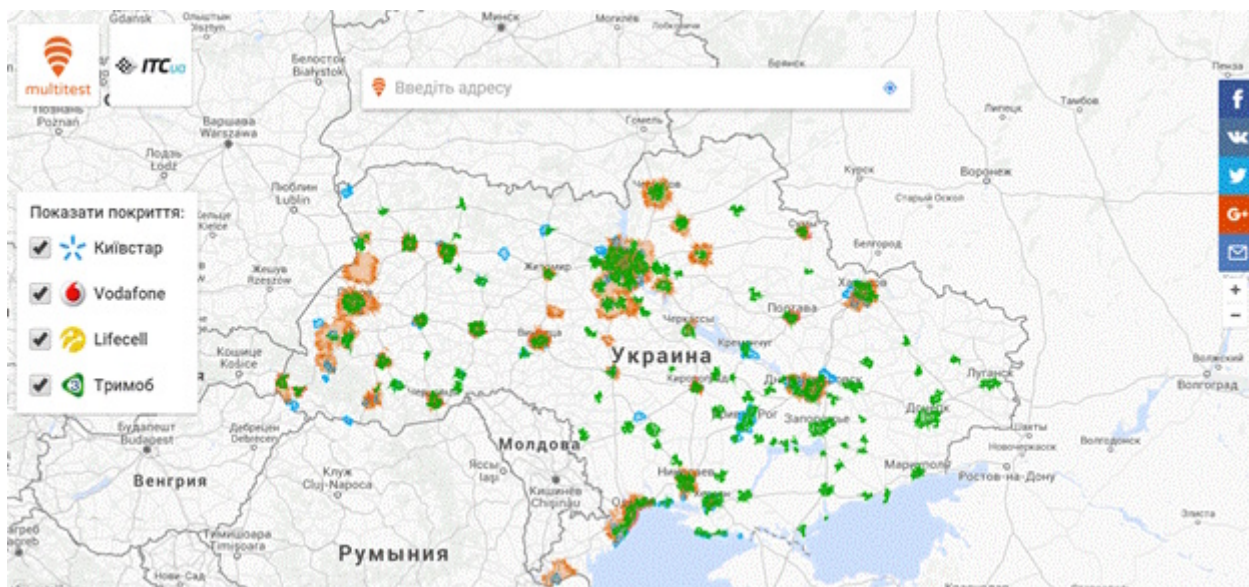
Пошук:

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. Ковпака, 29	263	<a href="#">↓</a>	<a href="#">📄</a>
300056	Публічне акціонерне товариство "Легбанк"	01033, м. Київ, вул. Жилианська, 27		<a href="#">↓</a>	<a href="#">📄</a>
300089	Акціонерний комерційний банк "Трансбанк"	03150, м. Київ, вул. Фізкультури, 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>

### 5.3.5 3G.Multitest.ua

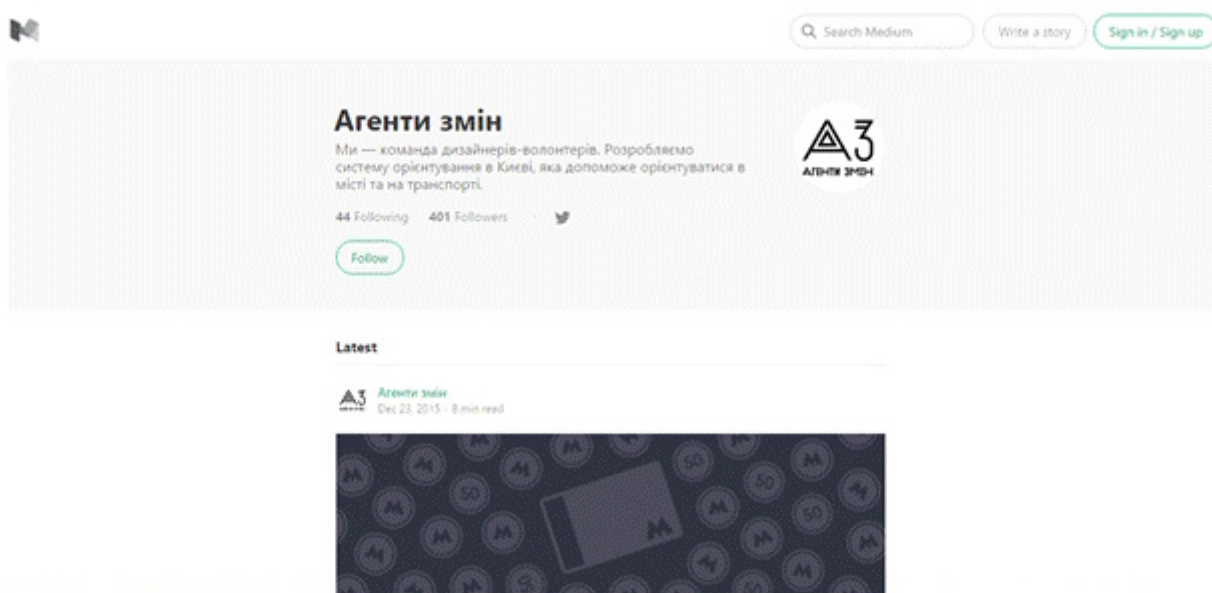
- 3g.multitest.ua — карта покриття 3G GSM операторів України.





### 5.3.6 Агенти змін

- [medium.com/@agencyzmin](https://medium.com/@agencyzmin) — система орієнтування в Києві.



### 5.3.7 Лун.UA

- [lun.ua](https://lun.ua) — оренда та продаж нерухомості у Києві та Україні.

**ЛУН** UA   Продажа   Аренда   Новостройки   Посуточно   + Как разместить объявление?   Войти

Раздел:    Район:    Метро:    Улица:    Адрес:    Комнат:    Цена:    **Найти**

☒ 2+ фото   ☐ новый дом   ☐ возле метро   ☐ вторичный рынок   ☐ без посредников

**Поиск по номеру телефона**  
   **Найти**

**Недвижимость Киева и Украины**  
 Купить квартиру в Киеве  
 Купить квартиру без посредников  
 Купить дом в Киеве  
 Снять квартиру в Киеве  
 Снять квартиру без посредников

**Расчет ипотеки от ПРАВЕ.УА**  
 Стоимость квартиры:  грн.  
 Начальный взнос:   
 Срок:  лет

**Новостройки Киева и Киевской области**

ЖК Левада Борисполь	ЖК Пурпурові вітрила Буча	КГ Новая Александровка с. Безугловка	ЖК Верховина Киев
ЖК Лесной квартал Бровары	ЖК Синергия-2 Ирпень	ЖК Традиция Киев	ЖК Уютный Ирпень

**Новостройки на карте Киева**

### 5.3.8 PEP.ORG.UA

- [per.org.ua/uk/](http://per.org.ua/uk/) — відкритий реєстр національних публічних діячів України.

**PEP.ORG.UA**

**ВІДКРИТИЙ РЕЄСТР НАЦІОНАЛЬНИХ ПУБЛІЧНИХ ДІЯЧІВ УКРАЇНИ**

Поіменно називаємо осіб, які в Україні є публічними діячами.

**Перевірити**

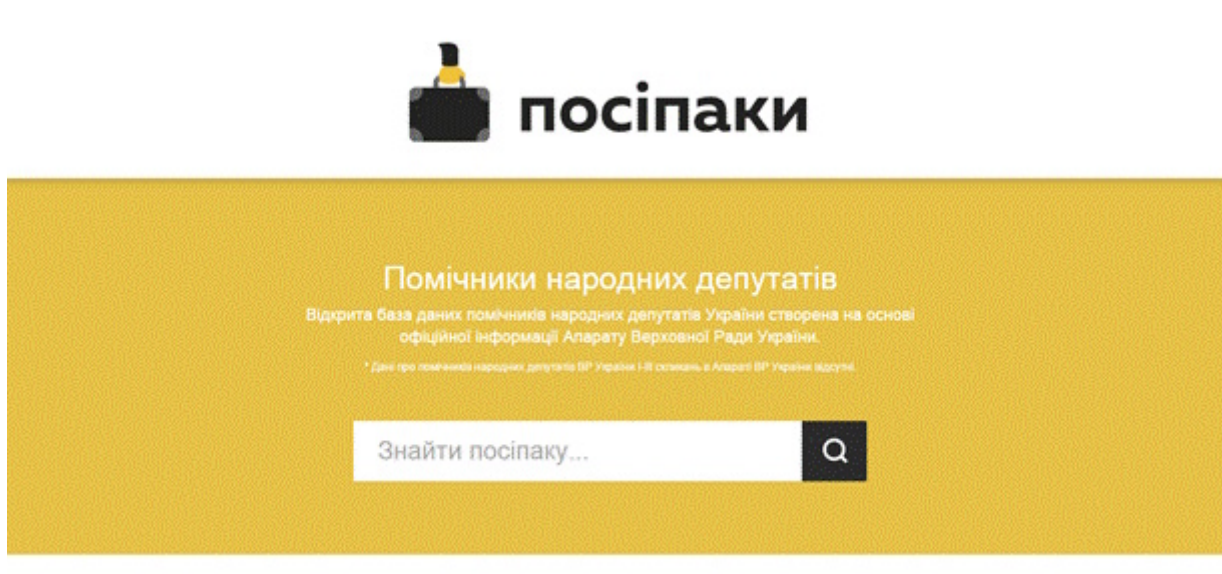
Корумповані високопосадовці легко відмивають брудні гроші через міжнародну банківську систему. Тому у світі запровадили особливі правила контролю за фінансовими операціями публічних осіб: кожна фінансова установа має встановити, хто її клієнт і звідки в нього гроші. Якщо клієнтом є публічний діяч, його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

<b>10495</b> Профіль фізичних осіб	<b>6082</b> Профіль публічних діячів	<b>4413</b> Профіль пов'язаних та близьких осіб
---------------------------------------	---	--

### 5.3.9 Посіпаки

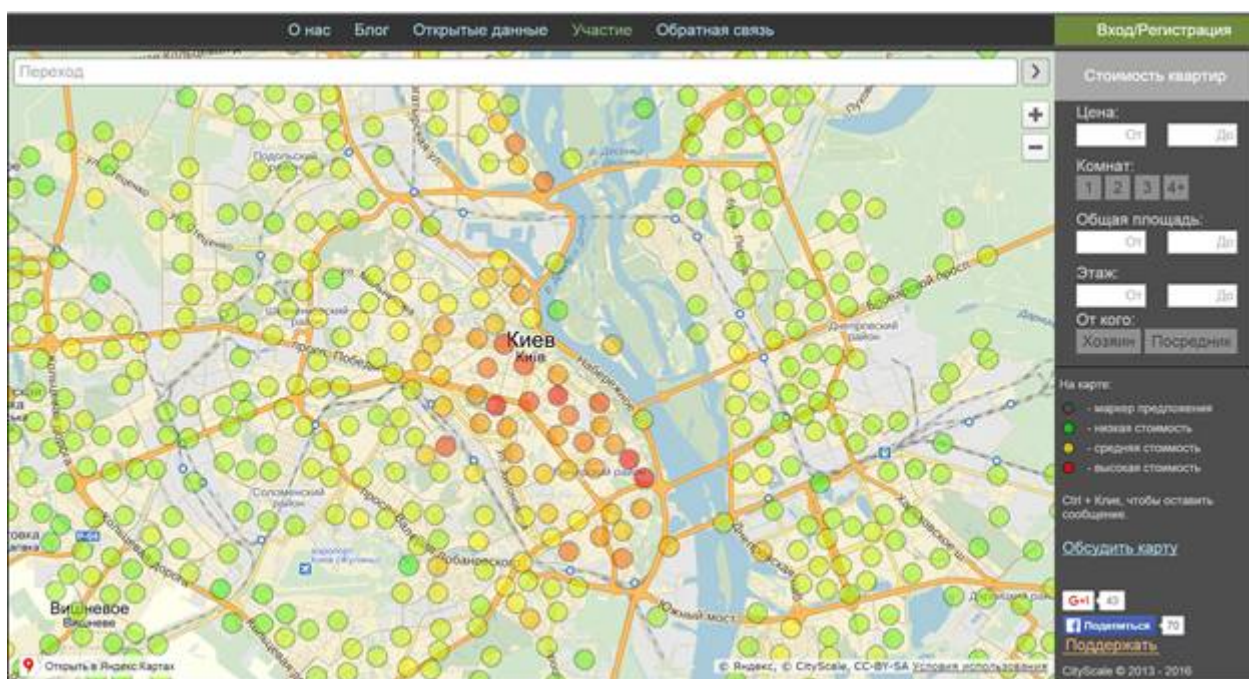
- [posipaky.info](http://posipaky.info) — база даних помічників народних депутатів України.





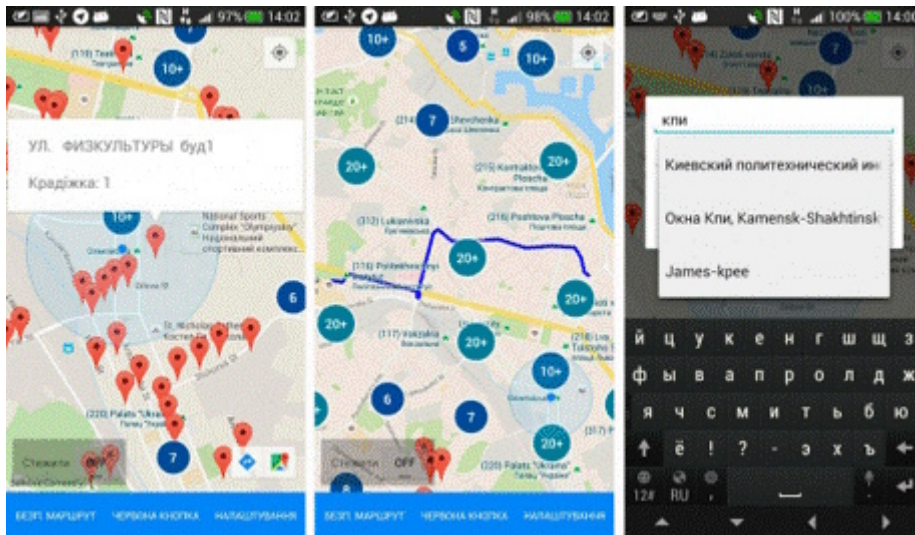
### 5.3.10 CityScale

- cityscale.com.ua — портал з інтерактивними картами умов життя.



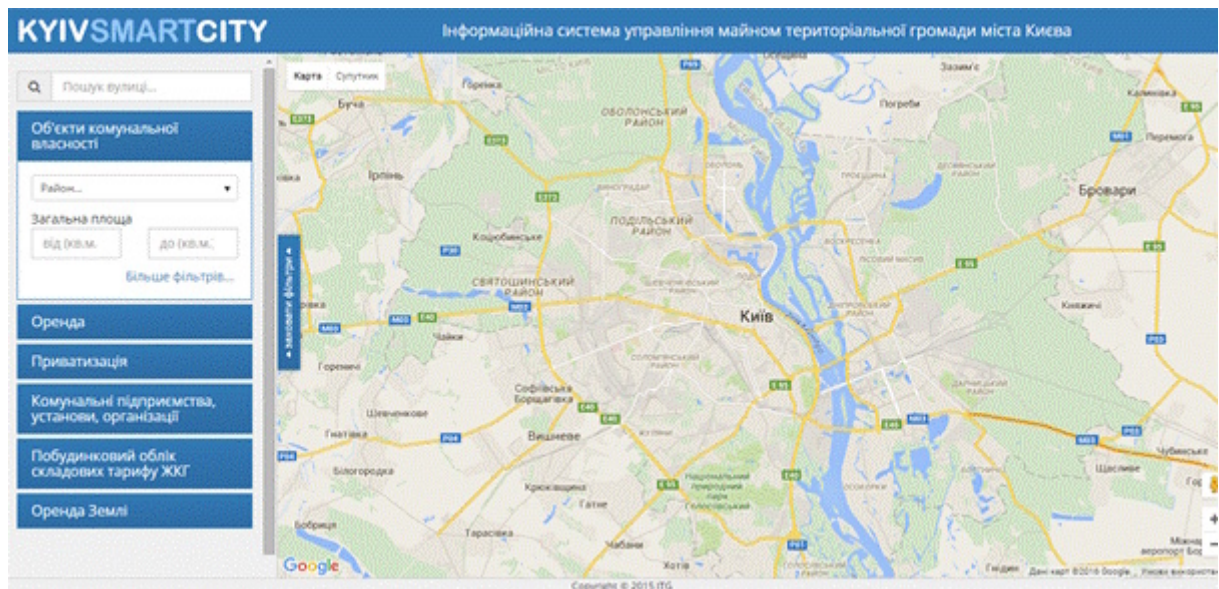
### 5.3.11 ARBUZ

Мобільний додаток для підвищення безпеки у містах на основі офіційних даних поліції.



### 5.3.12 KyivSmartCity

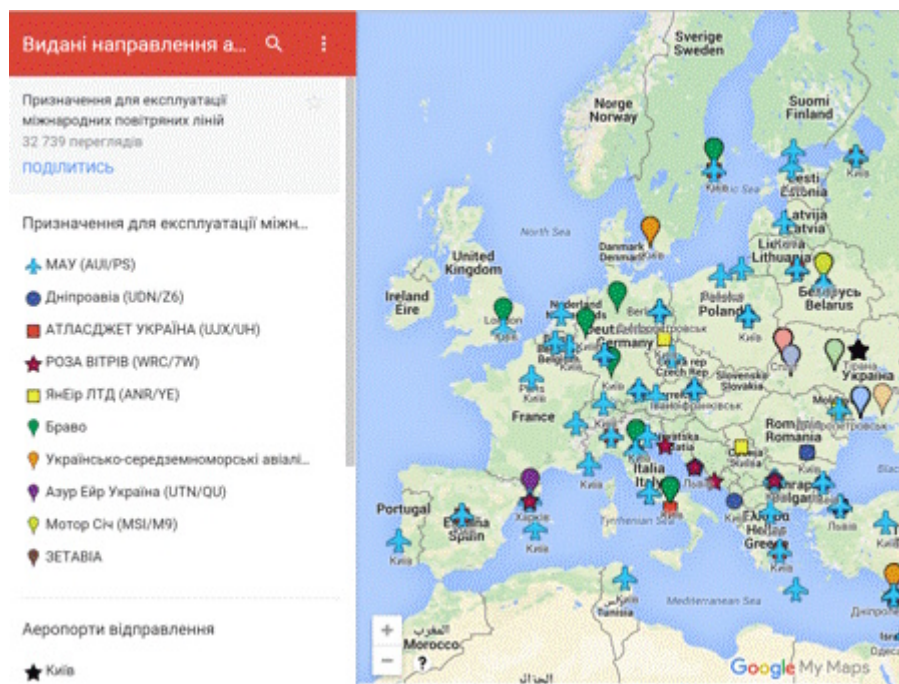
- [re.kievcity.gov.ua/objects](http://re.kievcity.gov.ua/objects) — система управління майном громади Києва.



### 5.3.13 Видані напрямлення авіакомпаніям України

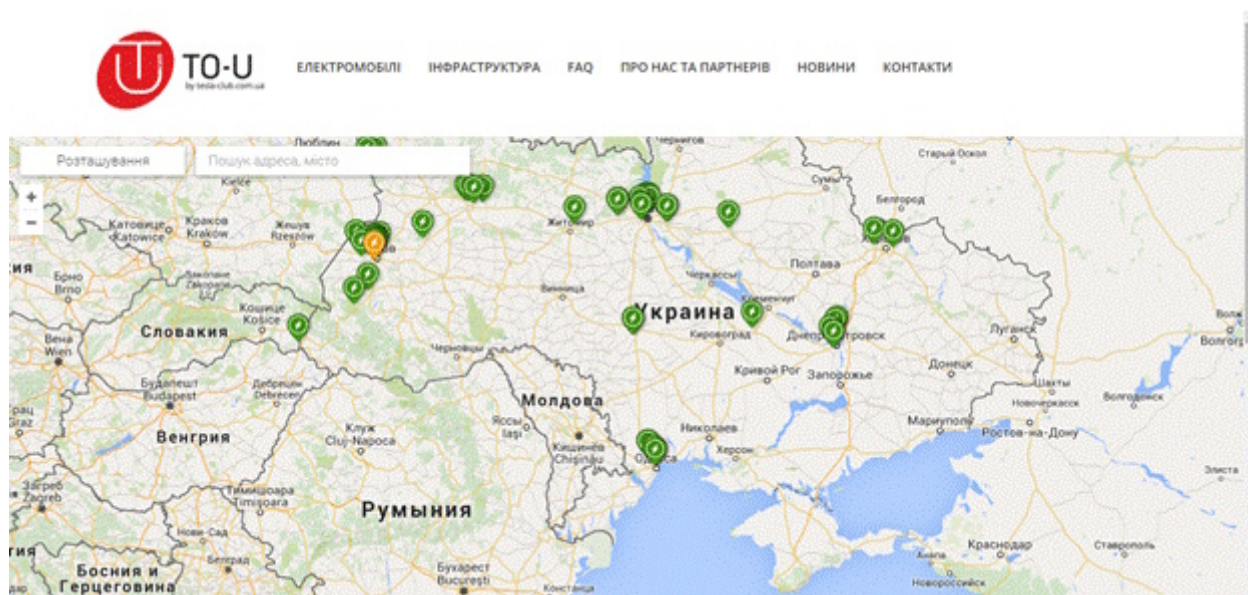
- Google Maps — призначення для експлуатації міжнародних повітряних ліній.





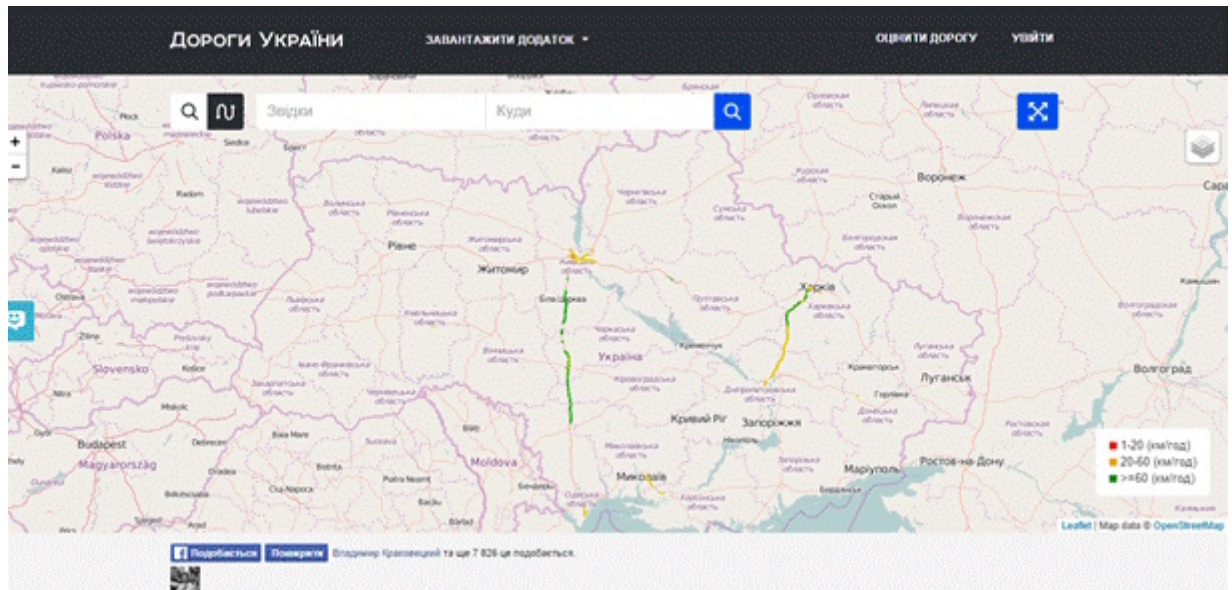
### 5.3.14 Tesla Club Ukraine

- tesla-club.com.ua/gmap — мережа заправок для електромобілів.



### 5.3.15 Дороги України

- uaroads.com — інформація про стан доріг в Україні.



### 5.3.16 Інші проєкти

1. Додатки та сервіси на базі відкритих даних ВРУ
2. Портал відкритих даних ВРУ
3. Як голосують депутати
4. Мобільний додаток «Zakonproekt» для відстеження законопроєктів
5. Кнопкодави
6. Візуалізація даних спільного ініціювання законопроєктів
7. Приймальні депутатів, prujmalni.chesno.org
8. Виборча варта
9. Електронні вибори

### 5.3.17 Інкубатор на базі відкритих даних 1991

1991 — перший в Україні некомерційний інкубатор, що допомагає перетворити відкриті державні дані на стартапи для громадян, бізнесу та держави.

Програми інкубатора:

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

## 6. Ресурси

### 1. **visualizing.org**

<http://www.visualizing.org>

Спільнота і інформаційна платформа про візуалізації даних.

### 2. **Google Chart Tools**

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

Javascript API від Google для простого створення таблиць візуалізації для постійно змінюються даних.

### 3. **GeoCommons**

<http://geocommons.com>

Інструментарій, спільнота і сервіс візуалізації для спільного використання геоданих.

### 4. **Quadrigram**

<http://www.quadrigram.com>

Професійна платформа з можливістю платного побудови кастомізованих візуалізацій.

### 5. **JavaScript InfoVis Toolkit**

<http://thejit.org>

Javascript-інструментарій для створення і підтримки візуалізації різного роду графіків.

### 6. **d3.js - Data-Driven Documents**

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

Маленька і дуже гнучка JavaScript бібліотека з відкритим вихідним кодом для управління документами на основі даних (наприклад, для генерації HTML5 або SVG-діаграми з даних).

### 7. **Google Public Data Explorer**

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

Каталог загального набору даних та інструмент для публікації і візуалізації великих наборів даних.

**8. Maps Marker WP-Plugin**

<http://www.mapsmarker.com>

Wordpress-плагін для відображення карти з анотацією видатних пам'яток в блозі Wordpress.

**9. DataMaps.eu - map your data**

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

Інструмент для створення привабливих карт візуалізації, які можуть бути створені в браузері через веб-сайт без знання програмування.

**10. Ushahidi**

<http://www.ushahidi.com>

Відкрите програмне забезпечення для збору, візуалізації та інтерактивного відображення на основі визначення місця розташування даних в реальному часі (наприклад, від надзвичайних ситуацій, політичних виборів і т.д.).

**11. Eclipse BIRT**

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

Система звітності eclipse для створення візуально привабливих звітів великих обсягів даних.

**12. Chartle.net**

<http://www.chartle.net>

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 вбудується в html-сторінку.

**13. Hohli**

<http://charts.hohli.com/>

Простий і скромний інтерактивний безкоштовний онлайн-інструмент для візуалізації даних за допомогою стандартного набору діаграм. (Немає можливості створювати карти.)

**14. IBM Many Eyes**

<http://www-958.ibm.com/software/data/cognos/manyeyes/>

Популярний онлайн-інструмент для візуалізації даних. Безкоштовний. Є можливість спільної роботи над проектами.

**15. TagCrowd**

<http://www.tagcrowd.com>



Онлайн-додаток для аналізу і візуалізації частотності вживання слів у тексті.  
Безкоштовний.

16. **Wordle**

<http://www.wordle.net>

Онлайн-додаток для аналізу і візуалізації частотності вживання слів у тексті.  
Безкоштовний.

17. **Tableau**

<http://www.tableausoftware.com>

Сімейство 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).

18. **Dundas**

<http://www.dundas.com>

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.

19. **Leximancer**

<https://www.leximancer.com>

Професійна програма для аналізу тексту та візуалізації результатів цього аналізу. Комерційна, платна, кроссплатформенная.

20. **SIMILE Widgets**

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

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 самостійного дослідження представленої бази даних.

21. **GeoCommons**

<http://geocommons.com>

Безкоштовний веб-інструмент зі створення карт на основі даних.

**22. Gephi**

<http://gephi.org>

Програмне забезпечення для візуалізації графів. Використовується як один з інструментів аналізу соцмереж. Безкоштовний, відкритий код, кроссплатформлений.

**23. Graphviz**

<http://www.graphviz.org>

Програма для візуалізації графів. Відкритий код, кроссплатформенна, безкоштовна.

**24. NewRadial**

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

Комплекс інструментів для візуального представлення нечислових даних (в тому числі зображень).

**25. Prefuse**

<http://www.prefuse.org>

Великий набір різних інструментів для створення складних інтерактивних візуалізацій. Вимагають вміння програмувати. Безкоштовний, код відкритий.

**26. ТЕКСТИ.ORG.UA**

<http://texty.org.ua/pg/blog/infoviz>

Блог: Інфовіз: інфографіка, інтерактивна візуалізація даних.

**27. Charted**

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

Швидка візуалізація CSV файлів.

**28. Venngage**

<https://venngage.com/>

Комерційний сервіс для створення інфографіки.

**29. Dipity**

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

Створення гарних таймлайнів.

**30. Easily**

<https://piktochart.com/>

Генератор інфографіки.

### 31. **Automatic Infographic Generator**

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

Генератор інфографіки.



## 7. Глосарій

**Абстрактна модель** — модель, що відображає загальні характеристики явища. Вона представляє інформацію про якісні характеристики об'єкта чи явища.

**Відкрита ліцензія** — документ, що визначає права й обмеження щодо об'єкта, регламентує вільне поширення контенту та/або програмного забезпечення. Свобода означає можливість використання, аналогічну свободі слова.

**Відкриті дані (Open Data)** — систематизована інформація, доступна через Інтернет у форматі, що дозволяє автоматизовану обробку, вільний і безоплатний доступ, а також подальше використання.

**Відкриті державні дані (Open Government Data)** — публічна інформація у вигляді відкритих даних про діяльність органів влади або створена в результаті їх діяльності.

**Власник інформації** — особа, яка створила інформацію або отримала право дозволяти чи обмежувати доступ до неї на підставі закону чи договору.

**Ідентифікатор набору відкритих даних (URN, Uniform Resource Name)** — постійний строковий параметр у визначеному форматі для ідентифікації та формування назв файлів набору, що включає інформацію про орган-власник, систему, тип даних, дати публікації й оновлення тощо.

**Інтерфейс прикладного програмування (API, Application Programming Interface)** — набір класів, функцій, структур і констант, доступних як сервіси через Інтернет для використання у прикладних програмах.

**Машинозчитувані дані** — дані у форматах, придатних для автоматичного використання.

**Метадані (метаінформація)** — структуровані дані, що визначають характеристики відкритих даних для їх ідентифікації та обробки.

**Набір відкритих даних (набір даних)** — сукупність однорідних записів, що містять структуровану інформацію — поля даних і метаінформацію.



## 8. Посилання та література

1. Відео. Як поєднати технології і волонтерський рух? | Дмитро Чаплинський | TEDxKPI
2. Оцифровування статистики, або Перша їжа для Bigdata
3. Данные нам: какие полезные сервисы Украина может получить от огромного массива своей информации
4. Безопасность как услуга: уроки парижских терактов для украинских операторов
5. How Ukraine is embracing open data after the 2014 revolution
6. Відкриті дані в Україні: що і як маємо відкрити? УКМЦ-04-03-2016
7. О развитии сферы открытых данных в Украине