

Лекція: Концепції і платформи відкритих даних Open Data

Вступ

Концепція відкритих даних (Open Data) передбачає забезпечення вільного доступу до даних у придатному для використання форматі, що сприяє прозорості, ефективності та інноваціям. У цій лекції ми розглянемо основи Open Data, їх значення, формати, приклади проєктів і специфіку відкритих даних в Україні.

1. Open data як концепція. Відкриті дані державних реєстрів

Open data – об'ємне поняття, що охоплює одночасно і концепцію, і явище, в якому вона втілюється.

Open data як концепція – це ідея про те, що інформація, яка не є предметом авторського права, не захищена патентами та іншими механізмами контролю, повинна бути доступна всім. Вважається, що обмеження доступу до даних йдуть проти суспільного блага. Відкриті дані не передбачають обмежень та/або оплати. Більш того, місія відкритих даних полягає у підзвітності влади перед громадянами.

Відкриті дані фактично – це будь-яка інформація, що може виявитися корисною. Дані державних реєстрів, відомості про житлово-комунальне господарство або, грубо кажучи, місця розташування громадських туалетів – усе це open data. При цьому формат даних повинен бути зручним для читання комп'ютерними алгоритмами та застосування. Придатні формати open data-файлів: JSON, XML, RDF. Електронні таблиці Microsoft Excel, текстові документи txt та PDF не рекомендуються для open data через відсутність в них чіткої структури даних.

Разом з open data часто використовується термін «big data» («великі дані», база даних) для позначення різних типів баз даних, у тому числі суспільно значущих комп'ютерних даних, які можуть зберігатися у загальному доступі та без перешкод використовуватися широкою аудиторією за допомогою опендатабот.

Відкриті реєстри (бази даних) для українців – це в першу чергу джерело інформації юридичного характеру. Їх грамотне використання допомагає захистити себе від шахрайства, виявити фіктивні підприємства та таким чином уникнути неправомірних дій на свою адресу. У будь-якому випадку, відкриті дані реєстрів дають доступ до певних знань, які ви можете вільно (безкоштовно, не підзвітно) застосувати в свою користь.

З відкритих реєстрів можна дізнатися, наприклад:

- генеральні плани та плани територій
- реєстри вузів та громадських організацій
- інформацію про державні закупівлі
- реєстраційні дані підприємців
- судові провадження
- об'єкти державної власності та багато інших

Ще в 2015 році Кабінет міністрів України прийняв постанову, яка визначає порядок відкриття даних та перелік, що складається з понад 300 першочергових наборів, багато з яких є повноцінними реєстрами. Фактично, зараз функціонують більше 100 відкритих державних реєстрів. Насправді даних набагато більше.

Тільки один сайт data.gov.ua (Єдиний державний портал відкритих даних, про який ми писали в однойменній статті) пропонує понад 14 тисяч наборів даних, але більшість з них не структуровані (записи в текстовому форматі, які складаються з декількох полів, та відскановані документи в форматі зображення і т. п.), тому доступ до інформації ускладнений, а можливість автоматичної обробки практично зведена нанівець.

2. Придатні формати open dataфайлів та open data-проекти

Вибір формату набору найчастіше залежить від програмного забезпечення, в якому зберігаються дані. Формати структурованих даних умовно можна розділити на табличні (XLS, XLSX, ODS, CSV) та ієрархічні (JSON та XML).

Формати електронних таблиць (XLS, XLSX, ODS). Електронні таблиці (Microsoft Excel, LibreOffice Calc, Google Таблиці) — це найбільш поширене програмне забезпечення для роботи з наборами даних.

CSV (англ. comma separated values — значення, що розділені комою). Цей формат використовується для обміну даними між різним програмним забезпеченням. Для збереження CSV-файлів необхідно використовувати Google Таблиці або LibreOffice Calc. CSV-файли з Microsoft Excel не відповідають загальноприйнятому формату, оскільки використовують специфічне кодування операційної системи (Windows-1251 замість UTF-8) та розділювач крапку з комою замість коми.

JSON та XML. На відміну від табличних форматів, формати JSON та XML слугують для збереження ієрархічних даних. Часто такі формати використовуються для передачі даних в інтернеті. Наприклад, можна легко отримати звітність та договори зі spending.gov.ua й оприлюднити їх на data.gov.ua. Також ієрархічно структуровані дані у Prozorro, Інспекційному порталі, Єдиному державному реєстрі юридичних осіб, фізичних осіб-

підприємців та громадських формувань. Для того, щоб обробляти та аналізувати такі набори, потрібно мати навички програмування.

API, Application Programming Interface (з англ. інтерфейс прикладного програмування). У контексті оприлюднення відкритих даних термін «API» означає спеціальний функціонал вебсервісів, що дає можливість оприлюднювати датасети автоматично (без участі розпорядника). Для отримання даних через API вебсервісів найчастіше використовуються запити у форматі гіперпосилань. Останні 8 цифр у посиланні — це код ЄДРПОУ розпорядника. Відповідно, змінюючи код ЄДРПОУ, можна отримувати дані нових розпорядників.

Найбільш відомі open data-проекти що забезпечують прозорі процеси управління інфраструктурою країни та допомагають налагодити взаємодію між державою та громадськістю

7 red lines: відстежування кожного вагону.

Щодня по коліям «Укрзалізниці» ходять десятки тисяч вагонів. Якщо реалізувати технологію відстежування кожного вагону, можна буде контролювати де він знаходиться, з якою швидкістю їде, наскільки сильно завантажений.

«Еквиток»: мікрокредитування та автоматичний викуп проїзного

Як відомо, пасажирські і приміські перевезення є основним джерелом збитків «Укрзалізниці». По-перше, це зумовлено низькими і дотаційними цінами на квитки, по-друге — застарілою моделлю комунікації із пасажирями.

«Розумна країна»: залиш свій відгук про отримані послуги

Кожен, хто хоч раз користувався послугами державних органів, знає, що далеко не завжди там надають якісний сервіс. Проект «Розумна країна» дозволяє будь-якому громадянину залишити свій відгук про якість обслуговування держкомпаній і держустанов — «Укрпошти», «Укрзалізниці», «Укравтодору» та інших.

3. Рейтинг Global Open Data Index. Портали відкритих даних України

Рейтинг Global Open Data Index складається з оцінки відкритості 16 різних сфер — від сфери державних фінансів, сфери державних витрат та державних закупівель до інформації про власників землі, результатів виборів усіх рівнів та державної статистики.

Найбільш схвальні оцінки, за даними дослідження, Україна отримала за відкритість державного бюджету та доступу до державних реєстрів. Також дослідники відзначили відкритість даних статистики, закупівель, національного законодавства та державних витрат. Найменш відкритими виявилися дані про землевласників та дані про природне середовище.

Єдиний державний веб-портал відкритих даних створено на виконання Закону України "Про доступ до публічної інформації" та постанови Кабінету Міністрів України від 21 жовтня 2015 р. № 835 "Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних".

Портал призначений для забезпечення надання доступу до публічної інформації у формі відкритих даних та передбачає доступ до інформації органів влади з можливістю її наступного використання.

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.

Цілі створення Порталу:

- забезпечення своєчасного розміщення органами влади інформації, яка підлягає оприлюдненню, а також будь-яких інших даних, що відповідають визначенню публічної інформації, у формі відкритих даних;
- оприлюднення та регулярне оновлення розпорядником інформації відкритих даних на Порталі;
- забезпечення для всіх користувачів інформаційного простору спільних правил щодо оприлюднення інформаційних матеріалів у формі відкритих даних;
- забезпечення своєчасного розміщення повної та достовірної інформації;
- забезпечення ефективних двосторонніх комунікацій і каналів зворотного зв'язку.

Вимоги до структури наборів даних, оприлюднених на Порталі:

- структура набору відкритих даних включає опис складу (елементів) набору даних, їх формат, параметри та призначення.
- структура набору відкритих даних оприлюднюється у форматах XSD, JSON, CSV або інших аналогічних форматах.

При оприлюдненні наборів даних використовуються такі формати:

Тип даних	Формат даних
Текстові дані	TXT, RTF, ODT, DOC(X), PDF (з текстовим змістом, нескановане зображення), (X)HTML
Структуровані дані	RDF, XML, JSON, CSV, XLS(X), ODS, YAML
Графічні дані	GIF, TIFF, JPG (JPEG), PNG

Тип даних	Формат даних
Відеодані	MPEG, MKV, AVI, FLV, MKS, MK3D
Аудіодані	MP3, WAV, MKA
Дані, розроблені з використанням програми Macromedia Flash	SWF, FLV
Архів даних	ZIP, 7z, Gzip, Bzip2
Геопросторові дані	GeoTIFF, SHP, DMF, MID/MIF, DXF, XML, GeoJSON, GPX, LOC, ARINC, AIXM.

Висновки

Відкриті дані є важливим інструментом для розвитку прозорості, інновацій та громадського контролю. В Україні існують численні платформи та проекти, які сприяють використанню відкритих даних у різних сферах. Подальший розвиток цієї галузі залежить від підвищення якості даних, регулярного оновлення та активної участі користувачів.

Література

1. Жабін А. І. Технологічні трансформації в бібліотеці: від бібліографування до наукометрії. Наукові праці Національної бібліотеки України ім. В. І. Вернадського. 2015. Вип. 41. С. 346–360.
2. Інформаційні технології та інновації в економіці, управлінні проектами і програмами : монографія / за заг. ред В. О. Тімофєєва, І. В. Чумаченко. Харків: ХНУРЕ, 2016. 402 с.
3. Інформаційні технології та технічні засоби навчання : навч. посіб. для студ. ВНЗ / О. П. Буйницька ; Київський ун-т імені Бориса Грінченка. Київ : Центр учб. л-ри, 2018. 240 с.
4. Клімушин П. С., Спасібов Д. В. Інноваційні сервіси відкритих даних для забезпечення ефективного функціонування е-уряду. Актуальні проблеми державного управління. 2017. № 1 (51). С. 1–8.
5. Копанєва В. О. Наукова комунікація: від відкритого доступу до відкритої науки. Бібліотекознавство. Документознавство. Інформологія. 2017. № 2. С. 35–45.
6. Електронна система Open Data URL: <https://opendatabot.ua>