

ФОРМАТИ ТА НАБОРИ ВІДКРИТИХ ДАНИХ**FORMATS AND SETS OF THE OPEN DATA**

Публічна інформація у формі відкритих даних – це дані, які будь-хто може вільно використовувати й поширювати. Дані, розміщені на офіційних державних порталах, опубліковані в текстовому (txt, rtf, pdf та ін.), графічному (gif, tiff, jpg та ін.), архівному (zip, 7z та ін.) форматі є незручними для аналізу, тому оприлюднення даних відбуватися в структурованих машиночитаних форматах (csv, xml, json тощо).

Для того щоб класифікація відкритих даних стала зручною та зрозумілою Т.Бернерс-Лі розробив п'ятизіркову модель відкритості даних, згідно з якою ступінь відкритості даних зростає від першої зірки до п'ятої залежно від формату представлення даних: одна зірка – не обов'язково вимога машинного читання даних; дві зірки – додається вимога до машиночитання даних; три зірки – додається вимога до відкритих форматів; чотири зірки – додається вимога до форматів, щоб вони задовольняли відкриті стандарти RDF; п'ять зірок – присвоюються даним, представленим за моделлю RDF і включеним у світову хмару пов'язаних відкритих даних.

Відкриті дані надаються у вигляді наборів даних (dataset), які оприлюднюються в мережі Інтернет на Єдиному державному веб-порталі відкритих даних та офіційних сайтах установ. Під набором даних звичайно розуміють організовану колекцію даних (data). Набір даних являє собою електронний документ (e-document, е-документ), який містить відкриті дані та складається із структурованої сукупності однорідних значень (записів), включає поля даних та метадані про них. Головною вимогою структурованої сукупності є однорідність значень, яка полягає у збіганні типів усіх значень, які розташовані у одному стовпчику.

Можна виділити такі характерні складові набору даних: заголовок документа, до якого входять дані розпорядника інформації, назва документу та дата, на яку складався документ тощо; структура набору даних у вигляді параметрів таблиці; безпосередньо набір даних. Набори даних згруповані в каталог. Каталог даних подібний до бібліотечного каталогу, який описує доступні книги та їх розташування. Але записи даних містять зібрані дані з різних галузей та регіонів, міністерств, державних установ і підрозділів, що включається у загальнодержавні записи метаданих. Серед цих даних є статичні та динамічні дані. Тож статичні дані оновлюються лише за наявності змін або після визначеного циклу, а динамічні потребують постійного оновлення. В ідеалі ці дані надаються у відкритому форматі, який можна прочитати за допомогою різних мов програмування.

У доповіді буде розглянуто формати відкритих даних та структуру наборів відкритих даних. Основною метою презентації наборів відкритих даних є створення їх додаткової вартості у вигляді інформаційних продуктів.

Література.

1. Ю. Б. Пігарєв. Електронне урядування та електронна демократія / Ю. Б. Пігарєв // Частина 11: Доступ до публічної інформації – 2017. – С. 34–55
2. Карпенко О. В. Цифрове врядування: імперативи реалізації в Україні / О. В. Карпенко // Актуальні проблеми державного управління : зб. наук. пр. ОРІДУ НАДУ. – 2017. – Вип. 3 (71). – С. 57–61