

Кременчуг, 2017

Открытые данные и их применение

определение, текущее состояние, применение, проблемы

IT-Перспектива

Київ, 2017

Что обсуждаем?

1

Понятие открытых данных

2

Открытые данные в Украине

3

Приложения на открытых данных

4

Опыт разработки приложений

5

Проблемы открытых данных

Определение



Публичная информация в форме открытых данных -

это публичная информация в формате, который позволяет ее автоматическую обработку электронными средствами, свободный и бесплатный доступ к ней, а также ее дальнейшее использование.

Принципы открытых данных



Открытость по умолчанию



Своевременные и комплексные



Доступные и используемые



Интероперабельные и сравнимые



Использование для оптимизации
управления и привлечение граждан



Для инклюзивного развития и инноваций

Государственные открытые данные

- 1 Данные должны быть полными
- 2 Данные должны быть первичными
- 3 Данные должны быть актуальными
- 4 Данные должны быть доступными
- 5 Данные должны быть машиночитаемыми
- 6 Доступ должен быть недискриминационным
- 7 Формат данных должен быть непроприетарным
- 8 Данные должны распространяться под свободной лицензией

Открытые данные в Украине

14.10.2014



Появление понятия открытых данных в “Антикоррупционной стратегии Украины на 2014-2017 годы”

01.02.2015



Про поддержке ПРООН разработан законопроект про открытые данные

09.04.2015



Законопроект принят Верховной Радой

21.10.2015



Принято постановление КМУ про наборы данных

23.10.2015



Запущен Национальный портал открытых данных











Открытые данные в Украине

06.10.2016

“Вітаю! Щойно у Мадриді офіційно оголошено про приєднання України до Міжнародної Хартії відкритих даних. ”



Законодавельство

-  [ЗУ «Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних»](#)
-  [ЗУ «Про доступ до публічної інформації»](#)
-  [ЗУ «Про захист персональних даних»](#)
-  [ЗУ «Про внесення змін до статті 28 Бюджетного кодексу України щодо доступу до інформації про бюджетні показники у формі відкритих даних»](#)
-  [ЗУ «Про відкритість використання публічних коштів»](#)
-  [Закон «Про внесення змін до деяких законів України у сфері державних закупівель щодо приведення їх у відповідність із міжнародними стандартами та вжиття заходів з подолання корупції»](#)
-  [ЗУ «Про публічні закупівлі»](#)
-  [Постановление КМУ від 21.10.2015 р. №835 «Про затвердження Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних»](#)
-  [Постановление КМУ від 30.11.2016 р. №867 «Деякі питання оприлюднення публічної інформації у формі відкритих даних»](#)
-  [ЗУ «Про зпобігання корупції»](#)

Где искать открытые данные?

Наиболее простой способ найти данные



open data great britain



Где искать открытые данные?

Государственные данные можно найти на таких ресурсах:



Единый государственный веб-портал открытых данных



Открытые данные Верховной Рады Украины



Электронные декларации на сайте НАПК



Результаты ВНО на сайте УЦ оценки качества образования

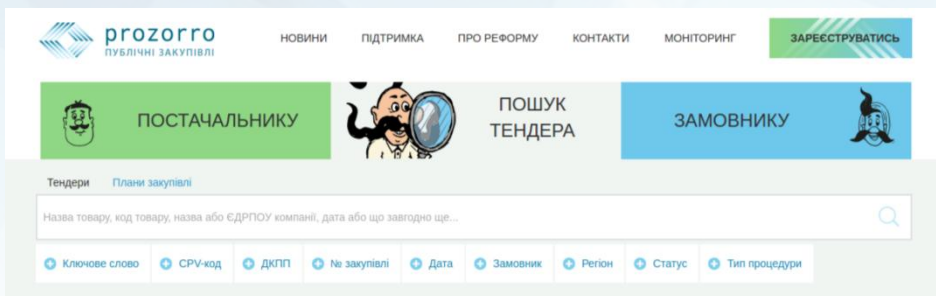


Электронная система публичных закупок



Официальный портал публичных финансов Украины


Приложения



The screenshot shows the ProZorro website interface. At the top, there is a navigation bar with the ProZorro logo and links for "НОВИНИ", "ПІДТРИМКА", "ПРО РЕФОРМУ", "КОНТАКТИ", "МОНІТОРИНГ", and "ЗАРЕЄСТРУВАТИСЬ". Below this, there are three main sections: "ПОСТАЧАЛЬНИКУ" (green), "ПОШУК ТЕНДЕРА" (white), and "ЗАМОВНИКУ" (blue). A search bar is present with the placeholder text "Назва товару, код товару, назва або ЄДРПОУ компанії, дата або що завгодно ще...". Below the search bar, there are several filter buttons: "Ключове слово", "CPV-код", "ДКПП", "№ закупівлі", "Дата", "Замовник", "Регіон", "Статус", and "Тип процедури".

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

[Детальніше](#)

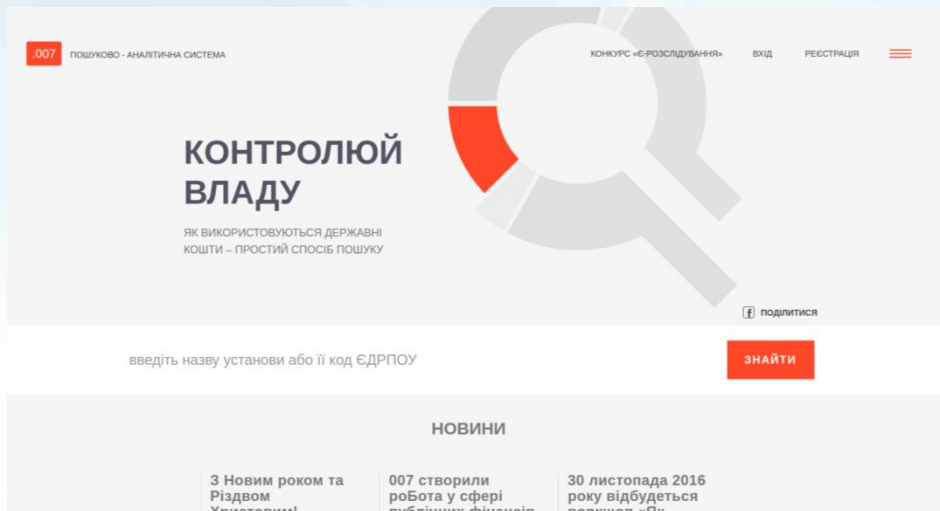


ОНЛАЙН-СИСТЕМА
ДОЛАЄ
ХАБАРМЕНА!
ПРОЗОРИ ДЕРЖТЕНДЕРИ

Система
государственных
закупок **Prozorro** и
связанные проекты:

- RIALTO
- Dozorro

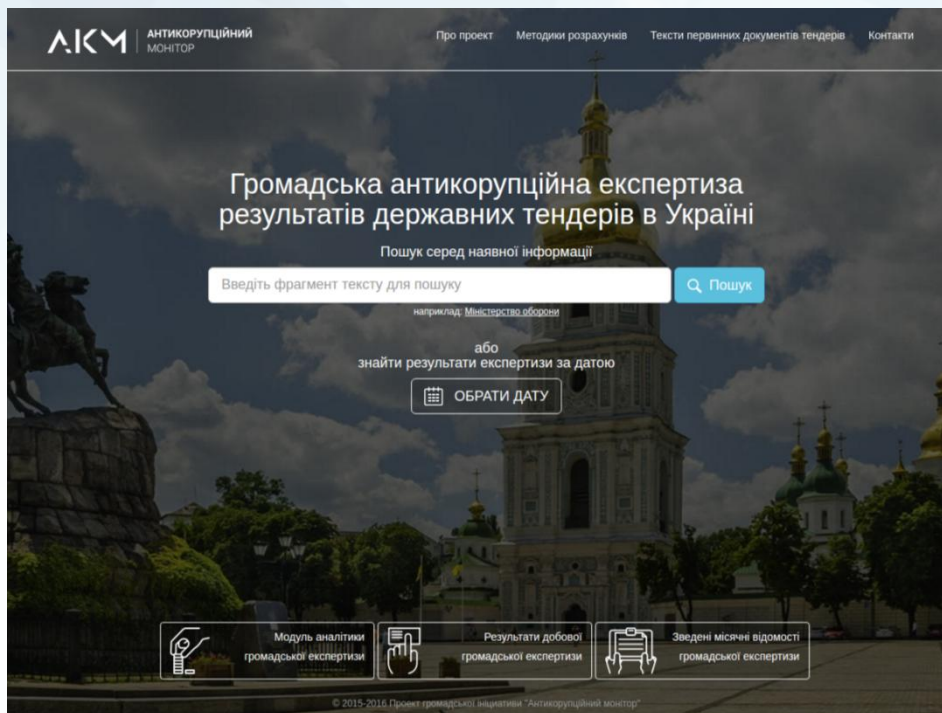
Приложения



Поисково-аналитическая система 007

Антикоррупционная платформа, цель которой - упростить поиск и визуализацию открытых данных про финансы в Украине.

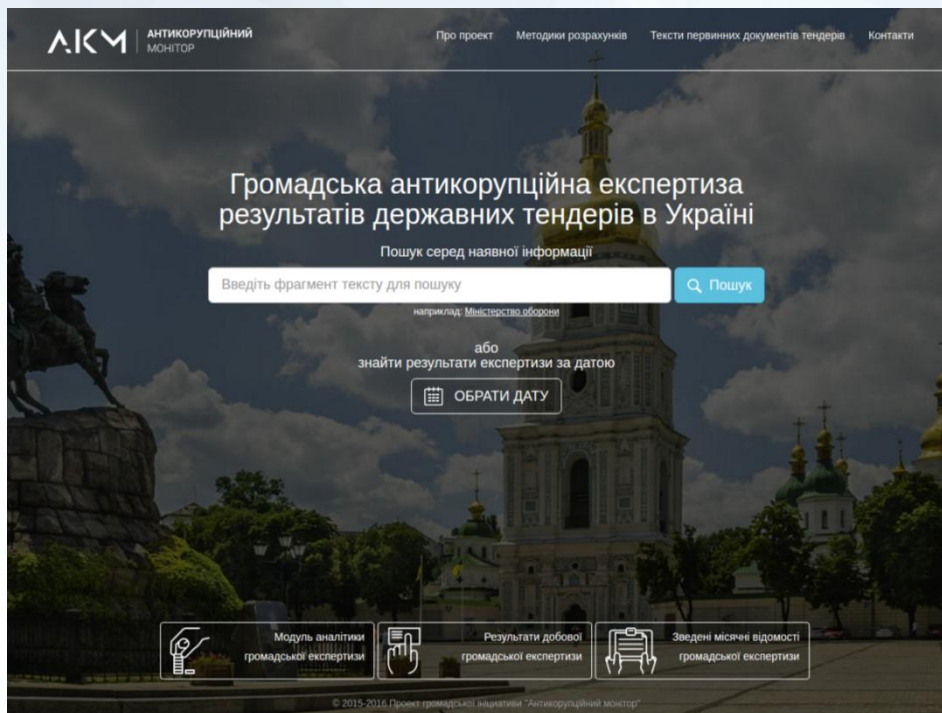
Приложения



Антикоррупционный монитор

Платформа позволяет провести аудит всех победителей государственных тендеров методами кредитного скоринга.

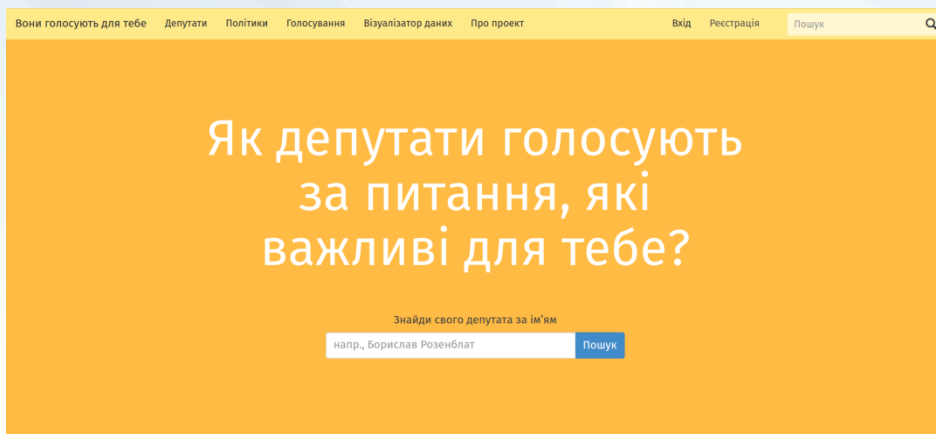
Приложения



Антикоррупционный монитор

Платформа позволяет провести аудит всех победителей государственных тендеров методами кредитного скоринга.

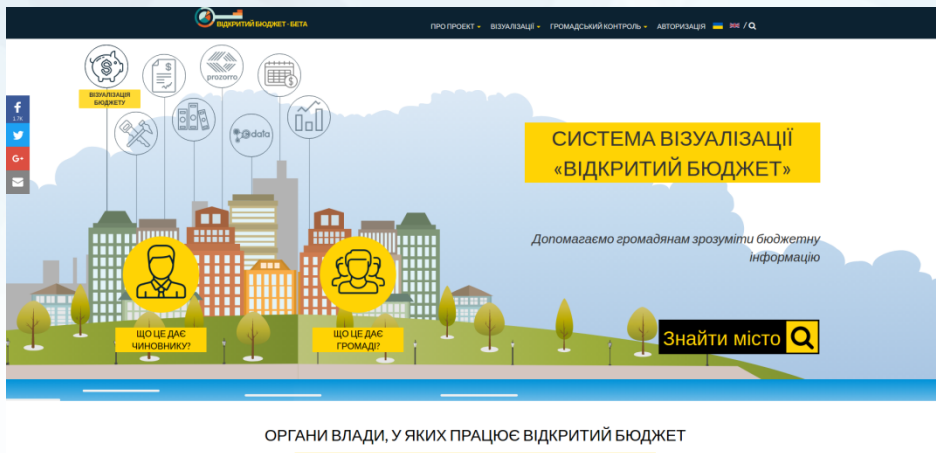
Приложения



Rada For You

Проект гражданской сети “Опора” позволяет следить за голосованиями и позицией народных депутатов Украины.

Приложения



Открытый бюджет

Система визуализации, разработанная общественной организацией “Эйдос”. Помогает гражданам разобраться в местном бюджете.

Чатботы



Open DataBot

Осуществляет поиск информации о юридических лицах в реестрах и судебных решениях



Open Declaration

Осуществляет поиск информации по декларациям чиновников



Казна Дніпра

Позволяет узнать о работах, которые местная власть проводила по конкретному адресу



Prozorro Bot

Уведомляет о новых торгах на Prozorro по заранее заданной системе фильтров



Zheglov Bot

Осуществляет поиск по физическим лицам, совершившим правонарушение

Опыт разработки

Преимущества разработки чатботов:

- Быстрый старт
- Готовая инфраструктура коммуникации с пользователем
- Нет необходимости в создании GUI
- Ориентированность на одну простую задачу
- Возможность быстро достичь специфической аудитории

Опыт разработки



Доступ через
API

Алгоритм

1. По кодам ЕГРПОУ распорядителей бюджетных средств запрашивается вся информация о транзакциях.
2. Фильтруется информация, где в назначении платежа упоминается адрес
3. При помощи регулярных выражений проводится очистка данных
4. По запросу пользователей производится поиск данных по адресу в базе и выдается результат

Опыт разработки



Проблемы

1. Невозможность получить архив данных, только отдельные записи через API
2. Нестабильная работа портала
3. Отсутствие tidy data: в поле payment_description входят КЭКР, ВПКРКМБ, назначение, номер договора
4. Намеренное использование в записях на украинском языке латинских букв и спецсимволов
5. Переименования улиц

Опыт разработки



Доступ через
API

Алгоритм

1. Каждые 5 минут алгоритм забирает все обновления торгов с Prozorro
2. Для каждого пользователя проверяется соответствие новых торгов установленным фильтрам
3. Если торг подходит по фильтру, пользователь получает уведомление

Опыт разработки



Проблемы

1. Торг имеет различные стадии, при этом не каждый торг имеет дату создания – сложно отслеживать является ли торг новым и неочевидно, после какого этапа его можно удалять
2. Поля описания и названия торгов используются произвольно
3. Информация об организациях загружается с площадок, а не единого реестра
4. Сложности пользователей с настройкой фильтров

Глобальные проблемы

Проблема

Революция в области больших данных не оправдала высоких ожиданий

Решение

По-прежнему настаивать на раскрытии данных, объяснять, какую пользу открытые данные уже приносят обществу, активно обсуждать успешные кейсы

Глобальные проблемы

Проблема

Открытые данные не являются бесплатными

Стоимость открытых данных

Открытые данные здорового человека

Арно Сагуге, технический директор New York University's GovLab указал на факторы скрытой стоимости открытия данных:

- Затраты на запуск, если данные хранятся в унаследованной компьютерной системе, которая требует переформатирования;
- Расходы, связанные с поддержкой актуальности а качества данных
- Судебные издержки, связанные с соблюдением законодательства об открытых данных;
- Расходы на проблемные ситуаияя, например, публикация непубличной информации;
- Затраты на связь с общественностью, которые могут возникнуть из-за критики на основе публичных данных

Факт: в Калифорнии за три года было **потрачено 756 000\$**
на раскрытие данных

Стоимость открытых данных

Открытые данные курильщика

- “Data.gov.ua был разработан ... за счет средств Международного фонда “Відродження”
- “Microsoft ... были спонсором нашего первого хакатона про открытые данные”
- “Очень дорого стоит содержание. Сейчас портал в бесплатном облаке Microsoft”
- “Стоимость содержания реальных государственных данных в облаке будет стоить около миллиона гривен в год. Содержание серверов будет еще дороже”
- “Доступ через API, и это большая нагрузка на базу. Нужно масштабироваться, а сделать это на серверах сложно.”

Глобальные проблемы

Проблема

Открытые данные не являются бесплатными

Решение

Работа над приоритезацией наборов данных, разработка решений, которые позволят автоматизировать процессы, связанные с публикацией данных

Глобальные проблемы

Проблема

Открытые данные не устраняют информационное неравенство, а увеличивают его, предоставляя преимущество техническим специалистам

Решение

Организация образовательных мероприятий, хакатонов, конференций. Создание дополнительного инструментария по визуализации и обзору данных.

Глобальные проблемы

Проблема

Проблемы, связанные с персональными данными. Эффект мозаики – когда сопоставление данных из разных источников позволяет идентифицировать личность

Решение

Экспертная работа с данными перед публикацией, выработка рекомендаций для распорядителей данных

Глобальные проблемы

Проблема

Хорошие решения в области открытых данных не лежат на поверхности. Для качественных приложений используется 5-10 датасетов

Решение

Стимулировать публикацию данных, требовать соблюдения стандартов при публикации данных, добиваться интероперабельности данных

Глобальные проблемы

Проблема

Отсутствие метаданных, без которых датасеты могут терять какой бы то ни было смысл

Решение

Активно взаимодействовать с распорядителями данных, указывать на проблемы и ошибки при публикации данных

Глобальные проблемы

Проблема

Закон Гудхарта: если мы концентрируемся на каком-то конкретном показателе результата, то такой показатель теряет смысл, так как его начинают эксплуатировать

Пример Закона Гудхарта

ТОП-наборы данных по мнению пользователей
Национального портала открытых данных:

- Звіт про використання бюджетних коштів ДФС, зокрема, за окремими бюджетними програмами ДПІ у м. Краматорську ГУ ДФС у Донецькій області – **33 балла**
- Організаційна структура Вінницької міської ради – **31 балл**
- Організаційна структура Департаменту інформаційних технологій Вінницької міської ради – **19 баллов**

Глобальные проблемы

Проблема

Закон Гудхарта: если мы концентрируемся на каком-то конкретном показателе результата, то такой показатель теряет смысл, так как его начинают эксплуатировать

Решение

Участие в комьюнити по открытым данным, независимые целостные экспертные оценки

Глобальные проблемы

Проблема

Поиск данных на портале открытых данных довольно тяжелое занятие. Публикуется много «мусора»

Мусор в даних

Названия датасетов:

- "Інформація з фотографіями щодо об'єктів незавершеного будівництва, які можуть бути задіяні для інвестицій"
- "Наказ Департаменту освіти і науки Миколаївської обласної державної адміністрації від 27.03.2017 № 266 ""Про проведення обласних змагань з судномодельного спорту серед учнівської молоді та юнацтва (в класах моделей до 600 мм та швидкісних радіокерованих моделей)»"", зареєстрований в Головному територіальному управлінні юстиції у Миколаївській області 07.04.2017 за № 131/2712"
- "Звіт за 3 квартал 2016 року"
- "Баланс 2017 р. "
- "Рішення сесії"

Глобальные проблемы

Проблема

Поиск данных на портале открытых данных довольно тяжелое занятие. Публикуется много «мусора»

Решение

Принимать участие в разработке стандартов по публикации данных, активно участвовать в оценке наборов

Глобальные проблемы

Проблема

Существующие решения для публикации данных были созданы для администрирования наборов, но не для прямого использования

API Национального портала

```
▼<result>
  <title>15-UFOP</title>
  <created>18.04.2017 18:17</created>
  <filemime>application/zip</filemime>
  <format>zip</format>
  <filesize>313334146</filesize>
  ▼<url>
    http://data.gov.ua/sites/default/files/media/document/1013/18.04.2017/15-ufop.zip
  </url>
</result>
```

Глобальные проблемы

Проблема

Существующие решения для публикации данных были созданы для администрирования наборов, но не для прямого использования

Решение

Для ценных наборов данных необходим переход к полноценным API-интерфейсам

Глобальные проблемы

Проблема

Отсутствие стандартов и интероперабельности данных. Например, идентификатор в данных о предприятиях, имеющих налоговый долг – название предприятия

Решение

Разработка стандартов публикации данных, нормализация баз данных

Глобальные проблемы

Проблема

Сопротивление чиновников дополнительной работе, извращенное толкование законодательства в сфере открытых данных

Толкование законодательства

«Інформація, яка міститься в кількох документах і не може бути зібрана **без значних інтелектуальних зусиль**, не відповідає критеріям «відображеності та задокументованості» і не є публічною»

Ответ ГФС на запрос о количестве адвокатов в Украине

Глобальные проблемы

Проблема

Сопротивление чиновников дополнительной работе, извращенное толкование законодательства в сфере открытых данных

Решение

Работа с общественным мнением, направленная на реформы в области открытых данных

Глобальные проблемы

Проблема

Связанные данные (5 звезд) более трудоемкие в создании и использовании. Даунтайм LOD-сервисов составляет 1,5 дней в месяц, что непригодно для коммерческих проектов

Решение

На данный момент для работы будет достаточно данных на уровне 3 звезд

Кременчуг, 2017

Вопросы?

ІТ-Перспектива

Київ, 2017