## Data sets

**Набір даних** — колекція однотипних даних, що застосовується в задачах машинної обробки даних.

Найчастіше набір даних відповідає змісту однієї [таблиці бази даних](https://uk.wikipedia.org/wiki/%D0%A2%D0%B0%D0%B1%D0%BB%D0%B8%D1%86%D1%8F_(%D0%B1%D0%B0%D0%B7%D0%B8_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85)" \o "Таблиця (бази даних)) або статистичній [матриці даних](https://uk.wikipedia.org/w/index.php?title=%D0%9C%D0%B0%D1%82%D1%80%D0%B8%D1%86%D1%8F_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85_(%D0%B1%D0%B0%D0%B3%D0%B0%D1%82%D0%BE%D0%B2%D0%B8%D0%BC%D1%96%D1%80%D0%BD%D0%B0_%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0)&action=edit&redlink=1" \o "Матриця даних (багатовимірна статистика) (ще не написана)), де кожна з колонок таблиці містить однорідні значення, а кожен з рядків таблиці відповідає певному члену набору даних.

Наприклад, набір даних про квіти може містити назву різновиду, розміри пелюсток, яскравість забарвлення тощо.

Термін *набір даних* також використовується при визначенні даних в сукупності тісно пов'язаних таблиць, зображень тощо, що описують результати конкретного експерименту або подію. Прикладом такого типу є набори даних, зібрані космічними агентствами, що виконують експерименти з приладами на борту космічного зонда, або світлини, передані з космосу.

Окремі набори даних широко використовуються в академічних колах як тестові набори, що підтверджують результати наукових досліджень.

Окремі набори даних є відкритими для використання, інші надаються за, звичайно символічну, плату.

https://uk.wikipedia.org/wiki/%D0%9D%D0%B0%D0%B1%D1%96%D1%80\_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85

## Сервіси для Open Data

* [LocalData](http://localdata.com/case-detroit.html): мобільний додаток, який дозволить організованим групам жителів долучатися до відкритого розгляду документів місцевою владою, пропонувати свої рішення, повідомляти про нагальні проблеми тощо.
* [Open Data Impact Map](http://opendataimpactmap.org/): геопросторова платформа, на якій відмічені спільноти з усього світу, які вже використовують та розвивають проекти з відкритими даними. Підійде для пошуку натхнення і корисних ідей, оскільки дозволяє побачити роботу вже реалізованих задумів.
* [SmartAppCity](http://www.smartappcity.com/en/): іспанська розробка, що базується на використанні відкритих даних та партнерстві державних органів і приватних підприємств. Останні мають змогу поширювати свою інформацію, важливу для жителів, через міські сервіси відкритих даних. Працює в Мадриді.

### Прозорість і підзвітність

* [Clear My Record](https://clearmyrecord.codeforamerica.org/#learn_more_section): безкоштовний інструмент для браузеру, який об’єднує інформацію про кримінальні злочини з різних офіційних джерел в один репозиторій. Місцеві органи можуть вільно шукати об’єднані дані про попередньо вчинені злочини жителями певного міста, оновлювати та актуалізувати їх.
* [DataHaven](http://www.ctdatahaven.org/): сервіс для збору та обробки відкритих даних для кращого управління справами окремо взятого регіону — штату Коннектикут. Візуалізація даних допомагає, наприклад, виявити різний рівень доходу місцевого населення, який впливає на зміну економіки в окремих містах штату.
* [Young Europeans](http://ec.europa.eu/eurostat/cache/infographs/youth/index_en.html): використовує відкриті дані від Eurostat щоб створювати демографічні скетчі з даними про європейську молодь віком від 16 до 29 років. Хороший інструмент для тих, хто бажає моніторити економічні можливості для свого покоління у сусідніх країнах.

## Інфраструктура і транспорт

* [CoAXs](http://coaxs.mit.edu/job-map/): сервіс для створення мап з використанням відкритих даних про зайнятість населення. Дослідження нових вакансій, закритих позицій або ж динаміки зміни цих показників дозволяє вивчати економічний розвиток певного регіону. З використанням CoAXs, який, до речі, розробили в MIT, раніше були створені мапи Бостону, Лос-Анджелесу та інших мегаполісів.
* [Citi Bike NYC](https://www.citibikenyc.com/system-data): найвідоміший приклад аналізу даних про велосипедну інфраструктуру міста. Пропонує відкриті дані різноманітної тематики. На його основі навіть було [створено](https://cloud.google.com/blog/big-data/2016/12/analyzing-nyc-biking-data-with-google-bigquery) мапу найпопулярніших міських маршрутів Нью-Йорку за гендерною характеристикою: окремо серед велосипедистів і велосипедисток. Показовий приклад того, як відкриті дані можуть виявити такі характеристики міста, про які ми і не підозрювали.

### Інтернет речей для Smart City

* [Autolib](https://www.autolib.eu/en/our-commitment/partners/): паризький сервіс для сумісних поїздок, в якому задіяні електромобілі. Програма збирає дані від усіх сенсорів, датчиків, розташованих по місту, і планує маршрути. Розповсюдження сервісу допоможе зменшити рівень шуму в столиці Франції та скоротити обсяг шкідливих викидів в повітря.

### Співучасть громадян

* [BA Accessible](http://www.buenosaires.gob.ar/aplicacionesmoviles/ba-accesible): мобільний додаток для жителів Буенос-Айресу, Аргентина, що надає інформацію про доступність міських об’єктів жителям з особливими потребами. Через додаток можна направити повідомлення місцевій владі про необхідність облаштувати додаткові конструкції в громадських місцях, міських будівлях тощо.
* [Mapping Police Violence](https://mappingpoliceviolence.org/): продукт на основі відкритих даних, що з’явився в США у відповідь на численні випадки протистояння населення з правоохоронними органами. Відкриті державні дані повинні не лише містити багато інформації, але мати різнопланову і всеосяжну базу даних. Тільки тоді сервіси, як перераховані в цій підбірці, так і ті, що працюють в Україні, матимуть справді важливий соціальний ефект.

<https://www.imena.ua/blog/open-data-services/>

### Портали відкритих даних в Україні

* Офіційний портал [публічних фінансів України](http://www.spending.gov.ua/) https://spending.gov.ua/
* Електронна система публічних закупівель [ProZorro](http://www.prozorro.gov.ua/) https://www.prozorro.gov.ua/
* Єдиний державний [веб-портал відкритих даних](http://www.data.gov.ua/) https://data.gov.ua/
* Офіційне [інтернет-представництво НБУ](http://www.bank.gov.ua/) https://www.bank.gov.ua/control/uk/index
* Верховна рада України: [відкриті дані](http://opendata.rada.gov.ua/" \t "_blank) https://opendata.rada.gov.ua/
* Державна [Казначейська служба України](http://www.treasury.gov.ua/) https://www.treasury.gov.ua/ua

**Якими бувають дані:**  
  
1. Метадані (Metadata) (поняття вже детально описав Стас)  
2. Референс-дані (Reference data) - відносно рідко змінні дані, які визначають значення конкретних сутностей, які використовуються при виконанні операцій в рамках всього підприємства. До таких сутностей найчастіше ставляться: валюти, країни, одиниці виміру, типи договорів / рахунків.  
3. Майстер-дані (Master data) - це базові дані, які визначають бізнес-сутності, з якими має справу підприємство. До таких бізнес-сутностей зазвичай відносяться клієнти, постачальники, продукція, послуги, договори, рахунки, пацієнти, громадяни і т.п. Крім інформації безпосередньо про ту чи іншу майстер-сутність, в майстер-дані входять взаємозв'язку між цими сутностями і ієрархії. Часто замість терміна «майстер-дані» використовують термін «нормативно-довідкова інформація».  
4. Транзакційні дані (Transactional data) - це дані, які утворилися в результати виконання підприємством будь-яких бізнес-транзакцій. Наприклад, для комерційного підприємства: продажу продуктів і послуг, закупівлі, надходження / списання грошових коштів. Зазвичай такі дані базуються в системі управління ресурсами підприємства (ERP) або інших галузевих системах. Природно, транзакційні системи широко використовують майстер-дані при виконанні транзакцій.  
5. Історичні дані (Historical data) - це дані, які включають в себе історичні транзакційні і майстер-дані. Найчастіше такі дані акумулюються в ODS і DWH системах і служать для вирішення різних аналітичних завдань і підтримки прийняття управлінських рішень.

https://habr.com/ru/post/324148/