

Open data[1].....	2
У науці.....	3
У владі.....	3
Аргументи за і проти.....	4
Data Life Cycle.....	6
Вступ до DDF[1].....	7
Унікальні пари ключових значень.....	8
Дані зберігаються в DataPoints.....	8
Concepts, dimensions та indicators.....	9
Dimensions як ключ DataPoint.....	9
Indicators.....	10
Entities.....	10
Entity domains, properties та entities.....	10
Concepts та concept properties.....	11
Типи концепцій.....	12
Концепція композиції.....	12
Концепція композицій та ролі.....	13
Великі композиції.....	14
Вкладання композицій.....	15
Метадані.....	17
Мова та переклади.....	17
Ідентифікатор мови.....	17
Синоніми.....	17
Порівняння вітчизняної та зарубіжних Open data[1].....	19
Ділові правила і приписи.....	20
Призначення системи управління даними.....	20
Діаграма юзкейсів.....	20
Сценарій отримання інформації.....	21
Сценарій входу користувача в обліковий запис.....	22
Сценарій вводу нових даних користувачем.....	23
Сценарій відкату версії.....	23
Використані джерела.....	24

Open data[1]

Open data[1] - це ідея про те, що деякі дані повинні бути вільно доступними кожному, щоб вони використовувались та перевидавались за власним бажанням, без обмежень через авторські права, патенти та інші механізми контролю. Цілі руху open data[1] подібні до інших "відкритих" рухів, таких як open source, open hardware, open content, open education, open educational resources, open government, open knowledge, open access, open science, і open web. Як це не парадоксально, зростання руху open data[1] паралельне підвищенню прав інтелектуальної власності. Філософія open data[1] була давно затверджена (наприклад, в Мертонській традиції науки), але термін "open data[1]" сам по собі недавній, він набуває популярності з появою Інтернету та World Wide Web та, особливо, з запуском ініціатив уряду про Open Data[1], таких як Data.gov, Data.gov.uk та Data.gov.in.

Open data[1], також можуть бути linked data; коли вони є, це пов'язано з open data[1]. Однією з найважливіших форм open data[1] є відкриті державні дані, що є формою open data[1], створених правлячими державними установами. Значущість відкритої влади припадає на те, що вона є частиною повсякденного життя громадян, аж до найбільш рутинних/земних завдань, які начебто далекі від влади.

Поняття open data[1] не є новим; але формалізоване визначення є відносно новим. Одне визначення - "відкрите визначення", яке можна підсумоване твердженням, що "частина даних відкрита, якщо хто-небудь може вільно використовувати, повторно використовувати та перерозподіляти його — питання лише до вимог щодо атрибутів та розповсюдження. Інші визначення, включаючи "Open data[1] - це дані, до яких будь-хто може отримати доступ, використовувати або поділитися ", мають більш доступну коротку версію визначення, але посиляються на формальне визначення.

Open data[1] можуть включати в себе нетекстовий матеріал, такий як карти, геноми, канали, хімічні сполуки, математичні та наукові формули, медичні дані та практику, біологічну науку та біорізноманіття. Проблеми часто виникають тому, що вони комерційно цінні або можуть бути об'єднані в цінну роботу. Доступ до даних або їх повторне використання контролюють організації, як державні, так і приватні. Контроль може здійснюватися через обмеження доступу, ліцензії, авторські права, патенти та збори за доступ та повторне використання. Захисники відкритих даних стверджують, що ці обмеження суперечать загальному благу і що ці дані повинні бути доступними без обмежень або зборів. Крім того, важливо, щоб дані були повторно використані без необхідності отримання додаткового дозволу, хоча типи повторного використання (наприклад, створення похідних творів) можуть контролюватися ліцензією.

Створювачі даних часто не вважають необхідністю визначення умов володіння, ліцензування та повторного використання; замість того, вважають що не накладання авторського права робить дані загальнодоступними. Наприклад, багато вчених не розглядають оприлюднені дані, що впливають з їхньої роботи, як ті, якими потрібно контролювати та розглядати акт публікації в журналі як неявний випуск даних у загальний доступ. Проте відсутність ліцензії ускладнює визначення статусу набору даних і може обмежувати використання даних, що пропонуються в дусі "Open". Через цю невизначеність державні чи приватні організації також можуть збирати ці дані, захищати їх авторськими правами, а потім перепродувати їх.

Open data[1] може бути взята з будь-яких джерел.

У науці

Концепція відкритого доступу до наукових даних була інституційно створена з утворенням системи World Data Center, під час підготовки до Міжнародного геофізичного року 1957-1958 років. Міжнародна рада наукових спілок (зараз Міжнародна рада з науки) керує декількома світовими центрами даних з мандатом, щоб мінімізувати ризик втрати даних та максимізувати доступ до даних. Незважаючи на те, що рух open-science-data давно передував Інтернету, наявність швидкої та повсюдної мережі значно змінила контекст Open Science data, оскільки публікація або отримання даних стала набагато дешевшою та менш часозатратною. Проект "Геном людини" був основною ініціативою, яка демонструвала здатність open data[1]. Вона побудована на так званих Бермудських принципах, в якій сказано: "Усі відомості про геномну послідовність людини повинні бути вільно доступними та відкритими для того, щоб заохочувати дослідження та розвиток та максимізувати свою користь суспільству". Новітні ініціативи, такі як Консорціум структурних геномів, показали, що підхід до open data[1] також може бути використаний продуктивно у контексті промислових досліджень та розробок. У 2004 році міністри науки всіх країн Організації економічного співробітництва та розвитку (ОЕСР), до складу якої входять найбільш розвинені країни світу, підписали декларацію, яка, по суті, заявляє, що всі державні архівні дані повинні бути оприлюднені. Після запиту та інтенсивної дискусії з установами, що надають дані, в державах-членах, ОЕСР опублікував у 2007 році керівні принципи ОЕСР для доступу до даних досліджень, отриманих від державного фінансування, як рекомендацію про невідгідне законодавство.

Приклади відкритих даних у науці:

Проект Dataverse Network - програмне забезпечення архівного сховища, що сприяє обміну даними, постійне цитування даних та відтворюване дослідження. data.uni-muenster.de — open data[1] про наукові артефакти з Університету Мюнстера, Німеччина. Запущений в 2011 році. linkedscience.org/data - відкриті наукові набори даних, кодовані як Linked Data. Запущений в 2011 році.

У владі

Для державних open data[1] існує цілий ряд різних аргументів. Наприклад, деякі адвокати стверджують, що надання доступу державної інформації громадськості, може полегшити державну прозорість, підзвітність та участь громадськості. "Open data[1] можуть бути потужною силою для громадської підзвітності - вона може полегшити аналіз, обробку та об'єднання існуючої інформації, що дозволяє провести новий рівень громадської перевірки". Уряди, які надають можливість публічного перегляду даних, можуть допомогти громадянам займатися державними секторами та "додавати значення до цих даних". Деякі вважають, що відкриття офіційної інформації може підтримувати технологічні інновації та економічне зростання, дозволяючи третім особам розробляти нові види цифрових додатків та послуг. Кілька національних урядів створили веб-сайти для розповсюдження частини даних, які вони збирають. Це концепція спільного проекту

муніципального уряду щодо створення та організації культури для open data[1] або державних open data[1]. Крім того, на інших рівнях уряду створені веб-сайти open data[1]. Є багато державних установ, які просувають Open data[1] в Канаді. Data.gov перераховує сайти з 40 держав США та 46 міст США та округів з веб-сайтами для надання open data[1]. На міжнародному рівні, Організація Об'єднаних Націй має веб-сайт з open data[1], який публікує статистичні дані держав-членів та агенцій ООН, а Світовий банк опублікував низку статистичних даних стосовно країн, що розвиваються. Європейська Комісія створила два портали для Європейського Союзу: портал Open data[1] EU, який надає доступ до відкритих даних від установ ЄС, агенцій та інших органів та порталу PublicData, який надає набори даних від місцевих, регіональних та національних державних органів по всій Європі. У жовтні 2015 року Відкрите Урядове Партнерство започаткувало International Open data Charter - набір принципів та найкращих практик для випуску державних open data[1], офіційно схвалених сімнадцятьма урядами країн, штатів та міст під час Глобального саміту OGP у Мексиці.

Аргументи за і проти

Дебати щодо Open data[1] все ще розвиваються. Найкращі відкриті урядові програми спрямовані на розширення можливостей громадян, допомогу малим підприємствам або створення цінності іншим позитивним, конструктивним способом. Відкриті державні дані - це лише шлях до вдосконалення освіти, вдосконалення державної політики та створення інструментів для вирішення інших проблем у реальному світі. Хоча багато аргументів було прийнято категорично, наступне обговорення аргументів за та проти open data[1] підкреслює, що ці аргументи часто сильно залежать від типу даних та його потенційного використання. Аргументи, подані від імені Open data[1], включають в себе наступне:

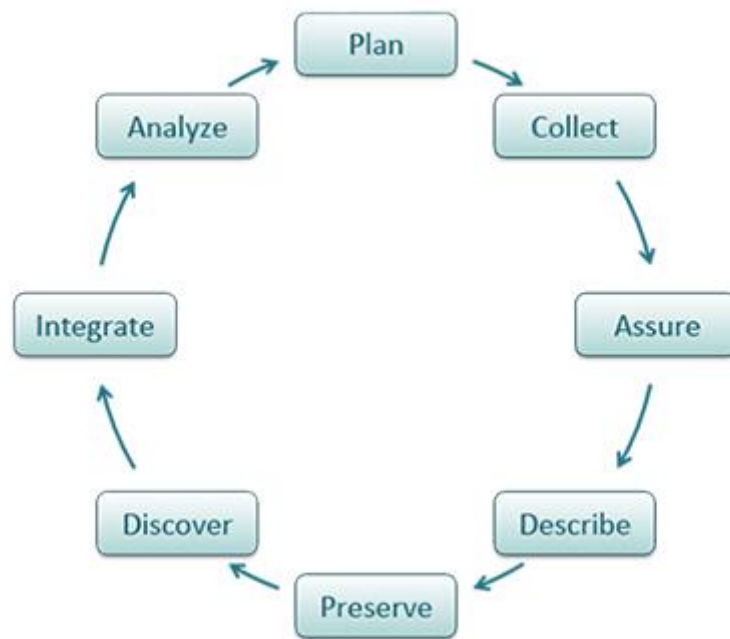
- "Дані належать людям". Типовими прикладами є геноми, дані про організми, медична наука, екологічні дані, що слідують за Організацією Конвенцією.
- Державні кошти були використані для фінансування роботи, і тому вона повинна бути загальнодоступною.
- Це було створеною державною або урядовою установою (це є загальноприйнятим у національних лабораторіях США та державних установах)
- Факти не можуть бути юридично захищеними авторським правом.
- Спонсори дослідження не отримують повного значення, якщо отримані дані не є вільно доступними.
- Обмеження на повторне використання даних створюють антикомітет.
- Дані необхідні для плавного процесу ведення громадської діяльності людини є важливим стимулом для соціально-економічного розвитку (охорона здоров'я, освіта, економічна продуктивність тощо).
- У наукових дослідженнях швидкість відкриття прискорюється шляхом кращого доступу до даних.
- Забезпечення open data[1] допомагає боротися з "зриванням даних" та гарантувати збереження даних наукових досліджень з часом.

Загалом вважається, що фактичні дані не можуть бути захищені авторським правом. Проте видавництва часто додають довідки про авторське право (часто забороняють повторне використання) до наукових даних, що супроводжують публікації. Можливо, незрозуміло, чи фактичні дані, вбудовані в повний текст, є частиною авторського права. Попри те, що людська абстракція фактів з паперових публікацій зазвичай прийнята як законна, часто це передбачає обмеження на вилучення машини роботами. На відміну від Open Access, де групи видавців заявили про свою стурбованість, open data[1], як правило, оскаржуються окремими установами. Їх аргументи обговорювалися менше в публічному дискурсі, і в цей час менше цитат, на які покладаються. Аргументи проти надання всіх даних, доступних у форматі Open data[1], включають наступне:

- Державне фінансування не може використовуватися для дублювання або заперечення діяльності приватного сектора (наприклад, PubChem).
- Уряди повинні бути відповідальними за ефективне використання грошей платника податків: якщо для агрегування даних використовуються державні кошти, і якщо дані принесуть комерційні (приватні) пільги лише невеликій кількості користувачів, то користувачі повинні відшкодувати урядам витрати на надання даних.
- Дохід, отриманий шляхом публікації даних, може бути використаний для покриття витрат на створення та/або поширення даних для того, щоб розповсюдження могло тривати безперервно.
- Дохід, отриманий за допомогою публікації даних, дозволяє некомерційним організаціям фінансувати інші види діяльності (наприклад, видавнича освіта суспільства підтримує суспільство).
- Уряд надає певну легітимність для певних організацій, щоб відшкодувати витрати (NIST в США, Ordnance Survey у Великобританії).
- Запитання про конфіденційність можуть вимагати, щоб доступ до даних містився лише для певних користувачів або для піднабору даних.
- Збирання, очищення, управління та поширення даних - це, як правило, трудомісткі та/або витратні процеси, - хто забезпечує ці послуги, повинен отримувати справедливую винагороду за надання цих послуг.
- Спонсори не отримують повного значення, якщо їхні дані не використовуються належним чином - іноді це потребує управління якістю, поширення та брендингові зусилля, які найкраще досягати шляхом стягнення плати з користувачів.
- Часто цільові кінцеві користувачі не можуть використовувати дані без додаткової обробки (аналіз, додатки тощо). Якщо хто-небудь має доступ до даних, жоден з них не може стимулювати інвестування в обробку, необхідну для полегшення даних (типовими прикладами є біологічні, медичні та екологічні дані).

Data Life Cycle

Життєвий цикл даних забезпечує високий рівень огляду етапів, пов'язаних з успішним управлінням та збереженням даних для використання та повторного використання. Існує декілька версій життєвого циклу даних з відмінностями, пов'язаними з варіаціями практик у різних доменах або спільнотах. Життєвий цикл даних [DataONE\[3\]](#) був розроблений командою Leadership DataONE[3] у співпраці з більш широкою спільнотою DataONE[3] і побудований на моделі життєвого циклу, запропонованої Національним науковим фондом у оригінальній заявці DataNet. Це служить базовою основою для розробки інструментів, послуг та навчальних матеріалів DataONE[3].



Життєвий цикл даних DataONE[3] містить вісім компонентів:

- План: опис даних, які будуть складені, і те, як даними можна керувати та як вони будуть доступні протягом усього їхнього терміну служби
- Збір: спостереження проводяться або вручну, або за допомогою датчиків чи інших інструментів; дані поміщаються в цифрову форму
- Гарантія: якість даних забезпечується через випробування та перевірки
- Опис: дані точно та ретельно описані, використовуючи відповідні стандарти метаданих
- Збереження: дані надсилаються до відповідного довгострокового архіву (тобто центру обробки даних)
- Відкриття: потенційно корисні дані, розташовані та отримані разом із відповідною інформацією про дані (метадані)
- Інтеграція: дані з неоднорідних джерел об'єднуються в єдиний однорідний набір даних, які можна легко аналізувати
- Аналіз: дані аналізуються

Деякі дослідження можуть використовувати лише частину життєвого циклу; наприклад, проект, що включає мета-аналіз, може зосереджуватися на етапах "Відкриття", "Інтеграція" та "Аналізу", тоді як проект, орієнтований на первинний збір та аналіз даних, міг би обійти кроки Відкриття та Інтеграція. Крім того, інші проекти можуть не відповідати лінійному шляху, наведеному тут, або може знадобитися декілька оборотів циклу. Крім того, деякі вчені або команди (наприклад, ті, хто займається моделюванням та синтезом) можуть створювати нові дані у процесі пошуку, інтеграції, аналізу та синтезу існуючих даних.

Життєвий цикл даних слугує навігаційним інструментом для бази даних BestONE Practics DataONE[3], що полегшує користувачам виявлення рекомендацій щодо ефективної роботи з їхніми даними на всіх етапах життєвого циклу даних. Програмне забезпечення в рамках Toolkit Investigator розроблено для підтримки дослідників на кількох етапах життєвого циклу даних. DataONE[3] розробила модулі освіти для управління даними, щоб допомогти педагогам та дослідникам в їх навчанні та самостійному навчанні.

Вступ до DDF[1]

DDF[1] - це модель даних для спільної гармонізації багатовимірної статистики.

- Модель даних DDF[1] - це модель даних, яка описує спосіб організації даних і визначає як частини даних відносяться один до одного.
- Гармонізація DDF[1] може бути використана для гармонізації даних, тобто він може поєднувати дані з різних джерел в інтегровані, послідовні та однозначні набори даних. DDF[1] підтримує практичний робочий процес, який забезпечує простоту обслуговування та постійно зростаючу колекцію порівнянних даних.
- Спільна робота DDF[1] - це загальна модель даних в Open Numbers. Open Numbers - це ініціатива скоординована та гармонізації світових даних.
- Багатовимірна статистика DDF[1] призначений для зберігання статистики з кількома параметрами. Наприклад, населення за країною, на рік, за статтю, за віковою групою.

DDF[1] використовується для визначення наборів даних. Набір даних - це сукупність послідовних, пов'язаних даних, які складаються з окремих елементів, але ними можна маніпулювати як одне ціле використовуючи один комп'ютер. Кожен набір даних DDF[1] визначає п'ять збірок даних: Concepts, Metadata, Entities, Datapoints та Synonyms.

1. **Datapoints** містять багатовимірні дані.

Населення чоловіків у Німеччині в 2015 році склало 293958.

2. **Entities** містять одномірні дані.

Офіційне довге ім'я Німеччини - Федеративна Республіка Німеччина.

3. **Metadata** містять додаткову інформацію про дані.

Джерелом для всіх даних про населення Німеччини в 2015 році (як чоловіків, так і жінок) є центральне бюро статистики в Німеччині, а точність становить 95%.

Джерелом даних, що Гонконг є китайською провінцією, є китайський уряд.

4. **Concepts** містять інформацію про змінні в наборі даних (наприклад, заголовки стовпців у табличному форматі).

Населення вимірюється в людях і має бути більше 0.

5. **Synonyms** містять синоніми, які використовуються для ідентифікації об'єктів та понять в інших наборах даних. Вони використовуються для гармонізації наборів даних з різних джерел до одного узгодженого набору даних.

У країні Німеччина є синоніми "ger", "deu", "de", "Bundesrepublik Deutschland" і "Federal Republic of Germany".

Dataset обов'язково повинен мати Concepts та можливо мати DataPoints, Entities, Metadata чи Synonyms.

Унікальні пари ключових значень

Кожна з цих колекцій складається з пар основних значень. Кожна пара ключових значень унікальна для всього набору даних.

Конкретний приклад: в одному наборі даних існує лише одне значення для: населення Німеччини у 2015 році, офіційне довге ім'я Німеччини, одиниця виміру для населення або джерело даних про всі дані населення Німеччини у 2015 році.

Дані зберігаються в DataPoints

Дані в DDF[1] зберігаються в парах ключових значень, які називаються точками даних. Ключ містить два або більше параметрів, тоді як значення складається з одного показника.

Нижче наведено таблицю:

1. DataPoints з ключем: країна та рік, значення: government_type.
2. DataPoints з ключем: країна та рік, значення: population.

	key: Dimensions		values: Indicators	
Concepts	country	year	government_type	population
	Germany	1940	dictatorship	3954858

DataPoints	Germany	2015	parliamentary democracy	50495854
	Sweden	2015	parliamentary democracy	9504847
	

Concepts, dimensions та indicators

Розглянемо коротко ці поняття (детальніше в наступних параграфах).

- **Concept** це все, що може бути заголовком стовпця в таблиці.
У таблиці вище, країна, рік і тип уряду – все це поняття.
- **Dimension** є ключем у парі ключ-значення DataPoint. У статистиці поняття ключ називається незалежною змінною.
У таблиці вище, країна та рік є розмірами, які є ключем до показника.
- **Indicator** це значення у парі ключ-значення DataPoint. У статистиці цей показник буде називатися залежною змінною, оскільки її значення залежить від ключової.
У таблиці вище тип уряду є показником

Dimensions як ключ DataPoint

Dimensions - це поняття, що складають ключ у точках даних. DataPoints в DDF[1] часто використовуються для зберігання змін протягом багатьох років та географічних розташувань.

Більше того, в ключ може входити будь-яка кількість dimensions (але більше одного), якщо це є необхідним для пошуку чи зберігання потрібної інформації.

key: Dimensions				value: Indicators
country	year	gender	age_group	population
Sweden	2015	male	0-4	485 055
Sweden	2015	female	0-4	483 538
Sweden	2015	male	5-9	495 494
...

Indicators

Сукупність **indicators** може бути розширено, щоб надавати більше даних для певного набору параметрів.

У наведеній нижче таблиці дані щодо населення та очікуваної тривалості життя надаються для вимірів країни та року

key: Dimensions		value: Indicators	
country	year	population	life expectancy
Sweden	2015	9 838 480	81.48
Sweden	2014	9 644 864	81.38
Norway	2015	5 165 000	80.69
...	

Indicator не повинен бути безперервним чи навіть числовим. У таблиці нижче показано поняття Тип уряду як indicator, який є дискретним та категоричним.

key: Dimensions		value: Indicators
country	year	government_type
Germany	1940	dictatorship
Germany	1990	parliamentary democracy
...

Disclaimer: this is a simplified version of Germany's transformation during the last century

Entities

Entity domains, properties та entities

У наведеній нижче таблиці країна є доменом організації. Full_name, широта та довгота є властивостями. Швеція, Норвегія та Росія є entities.

	key: Entity domain	values: Properties		
concept	country	full_name	latitude	longitude
entities	sweden	Kingdom of Sweden	59.3500	18.0667

	norway	Kingdom of Norway	61.0000	8.0000
	russia	Russian Federation	60.0000	90.0000

Кожен entity представлений унікальним ідентифікатором у стовпці домену об'єкта. Цей ідентифікатор унікальний, але тільки в межах домену організації. Ідентифікатори entity можуть містити лише символи літери/цифри та символи підкреслення. У кожного entity є властивості, які описують даний об'єкт чи характеризують його певним чином.

У entities може бути вказана ієрархія, яка буде вказана у колонці узагальнення (що вказують на відношення даного об'єкта до більш загальної групи).

Членство об'єкта до групи об'єктів позначається як властивість - <множина об'єктів>, з логічним значенням.

key: Entity domain	value: Properties		
geo	name	is--country	is--region
afr	Africa	false	true
eur	Europe	false	true
swe	Sweden	true	false
...

Concepts та concept properties

Concept - це заголовки таблиці у будь-якому місці набору даних DDF[1]. Concept properties це дані, які дають інформацію про concept. Спочатку загляньте в поняття-властивості - це домен та суб'єкт організації, що задають перелік, а властивості тип_концепції, домен і узагальнення.

Будь-яка властивість будь-якого concept може бути визначена аналогічним чином в DDF[1]. Всі concepts, що використовуються в data-set, повинні бути перераховані, і будь-яке властивість цих понять може бути визначено в цьому переліку

Типи концепцій

DDF[1] має ряд заздалегідь визначених типів концепцій. Концепція дається типом, встановлюючи властивість тип_концепції поняття. DDF[1] також дозволяє визначати власні типи концепцій.

Заздалегідь визначені типи концепцій:

1. Рядок (із набору символів)
2. Міра (числове значення)
3. Інтервал (між двома числовими значеннями)
4. Домен об'єкта
5. Множина об'єктів
6. Роль (сукупність, яка містить цілком ті самі сутності, що й інший об'єкт)
7. Композиція
8. Дата і час
9. Спеціальні типи (власновизначений тип)

Концепція композиції

Композиція являє собою комбінацію з двох або більше інших concepts. Наприклад, народження людини може складатися з міста і року, що завжди відбувається разом.

concept	concept_type	domain
person	entity_domain	
city	entity_domain	
year	time	
birth	composite	[city, year]

person	name	birth.city	birth.year
jodo	John Doe	sto	1985
jado	Jane Doe	bos	1957

city	name
sto	Stockholm
bos	Boston

Концепція композицій та ролі

Інший приклад - міграційний потік. Це завжди відбувається з одного місця в інше. У цьому випадку нам потрібні ролі, щоб запобігти дублюванню заголовків таблиці. Ролі також додають сенс до використання "міста".

concept	concept_type	domain
migration	measure	$x > 0$
city	entity_domain	
from_city	role	city
to_city	role	city
flow	composite	[from_city, to_city]

flow.from_city	flow.to_city	year	migration
sto	bos	2015	38575
bos	sto	2015	745

city	name
sto	Stockholm
bos	Boston

У наведеному вище прикладі, якщо ми не використовували ролі, ми матимемо два заголовки flow.city. Це означає, що ви не знаєте, який стовпець є джерелом і яке

призначення. Також буде порушено правило, що всі заголовки стовпців повинні бути унікальними.

Великі композиції

Останній приклад - коли один показник може містити кілька під-показників. Прикладом з дикої природи є data-set IMHE Global Burden of Disease. Цей data-set має 4 показники: смертність, yll, yld та daly. Кожен з них має кількість, співвідношення та відсоткові значення. Більш того, всі ці люди знову мають середнє, нижнє і верхнє значення. Це становить загальну суму 9 під-показників для кожного показника. Нижче показано, як складні поняття можуть допомогти в структуруванні цих даних.

concept	concept_type	domain
gbd_indicator	composite	[nm_lower, nm_mean, nm_upper, rt_lower, rt_mean, rt_upper, pc_lower, pc_mean, pc_upper]
death	role	gbd_indicator
yld	role	gbd_indicator
daly	role	gbd_indicator
yll	role	gbd_indicator
number	measure	$0 \leq x$
nm_upper	role	number
nm_mean	role	number
nm_lower	role	number
rate	measure	$0 \leq x$
rt_lower	role	rate
rt_mean	role	rate
rt_upper	role	rate
percent	measure	$0 \leq x \leq 1$

pc_lower	role	percent
pc_mean	role	percent
pc_upper	role	percent
location	entity_domain	
name	string	

location	year	death.nm_mean	death.nm_lower	...	yll.pc_upper
1	2015	49858	7567656	...	0.48
1	2015	46346	645746	...	0.54

location	name
1	Global

Вкладання композицій

Вище вказані не єдиний ідеальний спосіб моделювання набору даних GBD. Можливі багато інших способів структурування. Методи, що використовують дані, залежать від конкретної структури, яку вони використовуватимуть. Інший вибір полягає в тому, щоб вкладати складні поняття, як нижче:

concept	concept_type	domain
gbd_indicator	composite	[nm, rt, pc]
death	role	gbd_indicator
yld	role	gbd_indicator
daly	role	gbd_indicator
yll	role	gbd_indicator

nm	composite	[nm_lower, nm_mean, nm_upper]
rt	composite	[rt_lower, rt_mean, rt_upper]
pc	composite	[pc_lower, pc_mean, pc_upper]
number	measure	$0 \leq x$
nm_upper	role	number
nm_mean	role	number
nm_lower	role	number
rate	measure	$0 \leq x$
rt_lower	role	rate
rt_mean	role	rate
rt_upper	role	rate
percent	measure	$0 \leq x \leq 1$
pc_lower	role	percent
pc_mean	role	percent
pc_upper	role	percent
location	entity_domain	
name	string	
year	time	

location	year	death.nm.nm_mean	death.nm.nm_lower	...	yll.pc.pc_upper
1	2015	49858	7567656	...	0.48
1	2014	46346	645746	...	0.54

location	name
1	Global

Метадані

Метадані - це дані про дані. Наприклад: чоловіче населення Швеції у 2015 році становить 99483857. Це дані. Метадані для цих даних будуть його джерелом (статистика народонаселення ООН), точністю (5%) та примітками. Жіноче населення Швеції у 2015 році може мати інше джерело, точність та примітку. Те саме для тих же даних за 2016 рік. Однак поля можуть містити однакові метадані.

Як і у випадку з точками даних, об'єктами та поняттями, заголовки стовпців метаданих є поняттями і повинні бути визначені в наборі даних як поняття з поняттям типу.

Мова та переклади

Всі рядки в data-set DDF[1] ПОВИННІ бути доступними принаймні на одній мові. Вони можуть бути доступними на більшості мов. Принаймні однієї мовою повинен бути завершеним і альтернативною мовою, якщо інші мови мають неповні переклади.

Ідентифікатор мови

DDF[1] набори даних повинні визначати мову оригіналу або переклади, використовуючи ідентифікатор мови відповідно до технічного звіту Unicode # 35.

UTR # 35 дуже близький до стандарту RFC (BCP 47) "Теги для ідентифікації мов". Стандарт Unicode має вбудовану додаткову сумісність. Перетворення ідентифікатора мови Unicode в тег мови BCP 47 є простим. Цей документ містить трохи більше інформації про стандарти. Більш дружні огляди, подібні до цього, часто будуть досить точними, щоб допомогти вибрати правильний тег мови.

Синоніми

Синонім - це рядок, який може ідентифікувати поняття чи сутність, присутні в наборі даних. Колекція синонімів містить сутність ідентифікаторів концепцій з їх синонімами.

Наприклад. У країні, департамент deu має синоніми "ger", "de" та " *Federal Republic of Germany* ".

Синоніми використовуються для пошуку ідентичних понять та сутностей в інших наборах даних для семантичної гармонізації.

Синоніми відрізняються від перекладів. Переклади призначені для споживання людьми, для подання даних користувачам. Синоніми є для гармонізації наборів даних. Тим не менш, ви можете використовувати переклади, щоб збільшити набір синонімів для гармонізації.

Первинний ключ у цій колекції є складним: як поняття / сутність, так і його синонім разом утворюють ключ. Концепція, що використовується для синонімів, повинна називатись «синонімом» Концепція, що використовується для концепції / сутності, **ПОВИННА** бути «концепцією» концепцій або домену об'єктів або множини об'єктів (наприклад, географічного чи гендерного) для суб'єктів. Можна додати додаткові властивості, такі як джерело синоніму.

Синоніми повинні бути унікальними в колекції домену або концепцій організації. Наприклад, "Конго" можна використовувати тільки для однієї з двох країн "Республіка Конго" та "Демократична Республіка Конго". Але "Конго" можна використовувати як синонім як для країни, так і для компанії, якщо вони є двома різними доменами об'єкта.

<i>key</i>		<i>value</i>
country	synonym	source
deu	Deutschland	German name
deu	Germany	English name
deu	de	ISO 3166-1 alpha-2
deu	Federal Republic of Germany	Full english name
deu	276	ISO 3166-1 numeric
swe	Sweden	
swe	se	ISO 3166-1 alpha-2
...

Порівняння вітчизняної та зарубіжних Open data[1]

Сервіс "<https://datacatalog.worldbank.org/>" надає інформацію про економічні та соціальні дані країн світу.

Сервіс "<https://www.gapminder.org/data/>" надає різноманітну інформацію про країни світу.

Сервіс "<https://data.gov.ua/dataset>" надає різноманітну інформацію про Україну.

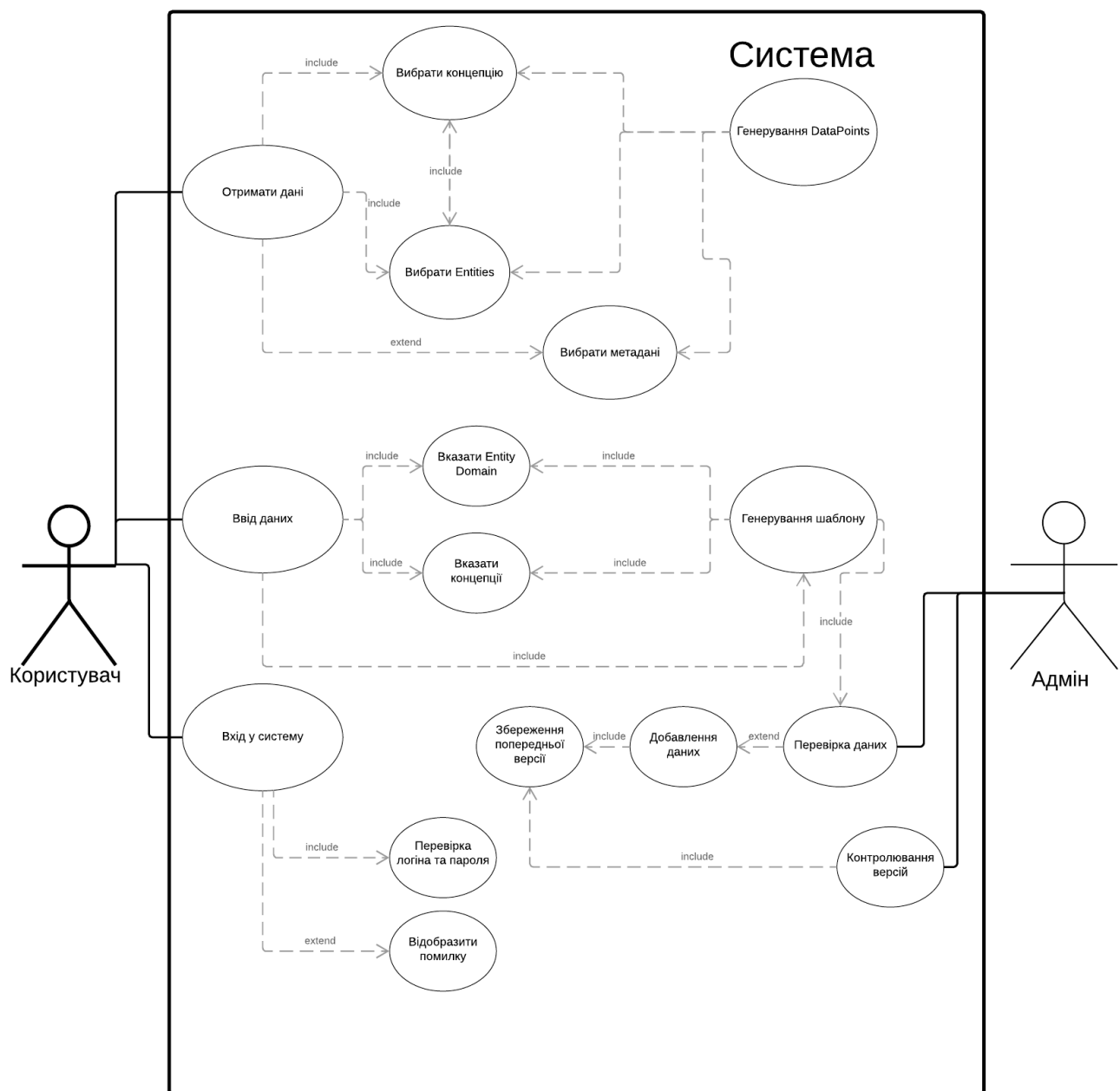
	gapminder.org	data.worldbank.org	data.gov.ua
Графіки та діаграми on-line	+	+	-
Додавання даних фізичними особами	-	-	+
Обмеження щодо використання даних	-	+	-
Перевірка даних	-	-	+
Накладання графіків	+	+	-
csv, xml, excel	+	+	+
Інтероперабельність даних	+	+	-
Прикладний програмний інтерфейс	Доступні лише користувацькі версії: індексація	Індексація, проектування, фінансові данні	Індексація

Ділові правила і приписи

Призначення системи управління даними

Функціональність системи призначена для автоматизації об'єднання результатів пошуку даних із Open Data на стороні сервера, а не користувача, та економія часу при роботі із даними із Open Data.

Діаграма юзкейсів



Сценарій отримання інформації

Назва:	Отримання інформації на основі запиту сформульованого користувачем
Учасники:	Користувач, система
Передумови:	
Результат:	Data Points, які складені на основі запиту користувача.
Виключні ситуації:	
Основний сценарій:	<ol style="list-style-type: none"> 1. Користувач вибирає концепт Entity domain. 2. Користувач вибирає Entity domains. 3. Система залишає для вибору Indicators лише ті концепти Properties, що є спільними для всіх концептів Entity domain. 4. Система генерує концепти Entity domains, які мають хоча б один спільний концепт Properties. 5. Користувач переходить до кроку 1, якщо бажає уточнити свій вибір, або переходить до наступного кроку для вибору концепту Indicators . 6. Користувач вибирає концепт Indicators. 7. Система генерує композицію концептів, на основі запиту. 8. Система повертає Data Point.

Сценарій входу користувача в обліковий запис

Назва:	Вхід користувача
Учасники:	Користувач, Система
Передумови:	<ol style="list-style-type: none"> 1. Користувач натискає кнопку [Реєстрація]. 2. Система пропонує форму «Реєстрація нового користувача». 3. Користувач заповнює поля: логін, пароль, прізвище, ім'я, email, сфера діяльності. 4. Користувач підтверджує правильність введених даних. 5. Система перевіряє введені дані. 6. Користувач стає повноправним користувачем системи.
Результат:	Користувач зайшов в обліковий запис і може вводити нові данні для подальшої модерації.
Виключні ситуації:	Користувач відхилив вхід
Основний сценарій:	<ol style="list-style-type: none"> 1. Користувач натискає кнопку [Вхід]. 2. Система перевіряє існування даного користувача і правильність введеного паролю. 3. Якщо Користувача не існує Система вибиває помилку і пропонує пройти реєстрацію. 4. Якщо Користувач існує, але пароль введено неправильно, Система вибиває помилку і пропонує повторити спробу входу. 5. Система підтверджує правильність введених даних і користувач заходить в свій обліковий запис.

Сценарій вводу нових даних користувачем

Назва:	Ввід даних
Учасники:	Користувач, Адміністратор, Система
Передумови:	Користувач увійшов у свій обліковий запис.
Результат:	Додано нові дані.
Виключні ситуації:	Користувач або Адміністратор відмінили додавання нових даних.
Основний сценарій:	<ol style="list-style-type: none"> 1. Користувач натискає кнопку [Додати дані]. 2. Користувач вибирає необхідний Entity Domain. 3. Користувач вибирає періоди часу (н-д у роках). 4. Користувач вибирає необхідну концепцію або вказує власну. 5. Система генерує шаблон множини DataPoints. 6. Користувач заповнює необхідний стовпець. 7. Користувач натискає кнопку [Підтвердити]. 8. Адміністратор перевіряє запропоновані дані і при необхідності відхиляє їх. 9. Система додає нові записи до бази даних

Сценарій відкату версії

Назва:	Повернення до попередньої версії даних.
Учасники:	Адміністратор, система.
Передумови:	Нововведені дані виявились неточними або потребують доробки.
Результат:	Надати системі правильні та надійні дані.
Виключні ситуації:	
Основний сценарій:	<ol style="list-style-type: none"> 1. Адміністратор заходить у меню редагування. 2. Адміністратор вибирає необхідну версію із списку всіх версій даних. 3. Адміністратор натискає кнопку [Застосувати]. 4. Система надалі використовує ту версію, яку вказав Адміністратор.

Використані джерела

1. <https://docs.google.com/document/d/1Cd2kEH5w3SRJYaDcu-M4dU5SY8No84T3g-QINSW6pIE/edit#heading=h.5h1e33vzhdlu>
2. https://en.wikipedia.org/wiki/Open_data
3. <https://www.dataone.org/data-life-cycle>
4. <https://www.lucidchart.com>