**Запити зацікавлених осіб**

**Постановка задачі**

У цьому документі описуються запити зацікавлених осіб, у якості яких виступають користувачі розподіленої системи, що буде організовувати повний життєвий цикл даних для міждисциплінарних досліджень.

Виходячи з цього, в такій системі повинно існувати необхідне забезпечення. Реалізувати це можливо через метадані. Для терміну МЕТАДАНІ немає єдиного формального визначення. Навпаки, існують різні визначення цього терміну. Ось просте і популярне переформулювання: метадані — це дані про дані. Цей термін в широкому сенсі слова використовується для будь-яких «даних про дані»: іменах таблиць, колонок в таблиці, програм і тому подібне.   
Метадані — це дані з загальнішої формальної системи, що описує задану систему даних.   
Існують вужчі визначення:   
Метадані це структуровані дані, що являють собою характеристики описуваних сутностей для цілей їх ідентифікації, пошуку, оцінки, управління ними.   
Метадані це набір допустимих структурованих описів, які доступні в явному вигляді і призначення яких допомогти знайти об'єкт. Це визначення використовується набагато рідше, оскільки воно концентрується на одному з призначень метаданих — пошук об'єктів, сутностей, ресурсів — та ігнорує інші призначення. [[Метадані](https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%B0%D0%B4%D0%B0%D0%BD%D1%96)]

Саме метадані будуть забезпечувати можливості реалізації усього життєвого циклу даних. Під життєвим циклом даних розуміють:

Створення даних   
•планування дослідження   
•управління даними плану (формати, зберігання і т.д.)   
•погодження плану для спільного використання   
•пошук існуючих даних  
•збір даних (експеримент, спостереження, вимірювання, моделювання)   
•виокремлення і створення метаданих   
  
Обробка даних   
• введення даних, оцифровка, транскрибування, переклад   
• перевірка, очищення даних   
• анонімізація даних в разі потреби   
• опис даних   
• управління та зберігання даних   
  
Аналіз даних   
• інтерпретація даних   
• отримання даних   
• створення результатів досліджень   
• автор публікації   
• підготовка даних для збереження   
  
Збереження даних   
• перенесення даних в кращий формат   
• перенесення даних у відповідне середовище   
• резервне копіювання та зберігання даних   
• створення метаданих і документації   
• архівація даних   
  
Надання доступу до даних   
• поширювати дані   
• контроль доступу   
• встановлення авторських прав   
• розділяти дані   
  
Повторне використання даних   
• подальші дослідження   
• нові дослідження   
• проведення науково-дослідних оглядів   
• ретельне вивчення результатів   
• вчити і вчитися   
[[data life cycle](https://www.dataone.org/data-life-cycle)] 

Метадані використовуються в кожній системі для своїх цілей. Оскільки метадані дуже різноманітні, то було придумано дещо універсальне “ключ значення” (це деяка структура в якій можна виділити ключі і значення, які під цими ключами лежать). Тому існують деякі парсери які дозволяють розпарсити метадані.

Наша система не є винятком, у нас є специфічні задачі які не можуть розглядатись тільки у межах одного стандарту (Дублінське ядро, наприклад).

Дублінське ядро (англ. Dublin Core) — словник (семантична мережа) основних понять англійської мови, призначений для уніфікації метаданих.   
Простий набір елементів метаданих Дублінського ядра (Dublin Core Metadata Element Set; DCMES) складається з 15 елементів метаданих:   
Title - назва;   
Creator - творець;   
Subject - тема;   
Description - опис;   
Publisher - видавець;   
Contributor - вніс вклад;   
Date - дата;   
Type - тип;   
Format - формат документа;   
Identifier - ідентифікатор;   
Source - джерело;   
Language - мова;   
Relation - відносини;   
Coverage - покриття;   
Rights - авторські права.   
Кваліфікований (компетентний) набір елементів метаданих Дублінського ядра, крім 15 перерахованих вище, може включати:   
Audience - аудиторія (глядачі);   
Provenance - походження;   
RightsHolder - правовласник.   
Кожен елемент опціональний і може повторюватися. Ініціатива метаданих Дублінського ядра (Dublin Core Metadata Initiative; DCMI) описала стандартні шляхи визначення елементів і заохочує використання схем кодування і словників. Не існує заздалегідь заданого порядку перерахування цих елементів. DCMI також підтримує невеликий загальний словник, який рекомендується використовувати з елементом Type (Тип) і який складається з 12 слів. Повна інформація за визначеннями елементів і відносин між ними описана в Реєстрі метаданих Дублінського ядра (Dublin Core Metadata Registry). [[Дублінське ядро](https://ru.wikipedia.org/wiki/%D0%94%D1%83%D0%B1%D0%BB%D0%B8%D0%BD%D1%81%D0%BA%D0%BE%D0%B5_%D1%8F%D0%B4%D1%80%D0%BE)] 

У роботі нам необхідно оперувати також такими термінами:

Управління даними - це розробка і виконання архітектур, політики, практики і процедур з метою ефективним чином управління життєвим циклом інформації потреби того чи іншого підприємства. [[Data management](http://searchdatamanagement.techtarget.com/definition/data-management)]   
  
 Набір даних являє собою сукупність деяких даних. Найчастіше набір даних відповідає змісту однієї таблиці бази даних, або одну матрицю статистичних даних, де кожен стовпець таблиці являє собою певну змінну, а кожен рядок відповідає певному члену набору даних про який йде мова. Дані установки значень списки для кожної з змінних, таких як зріст і вагу об'єкта, для кожного елемента набору даних. Кожне значення відомо як точки прив'язки. Набір даних може містити дані для одного або більше членів, що відповідає числу рядків. Набір даних термін також може бути використаний більш вільно, щоб звернутися до даних в сукупності тісно пов'язаних таблиць, з яких відповідає певному експерименту або події. Прикладом такого типу є набори даних, зібрані космічними агентствами, які здійснюють експерименти з інструментами на борту космічних апаратів. Набори даних, які настільки великі, що традиційні додатки обробки даних недостатньо для боротьби з ними відомі як великих обсягів даних. [[Data set](https://en.wikipedia.org/wiki/Data_set)]   
  
 Вели́кі да́ні (англ. Big Data) в інформаційних технологіях — набори інформації (як структурованої, так і неструктурованої) настільки великих розмірів, що традиційні способи та підходи (здебільшого засновані на рішеннях класу бізнесової аналітики та системах управління базами даних) не можуть бути застосовані до них[1]. Альтернативне визначення називає великими даними феноменальне прискорення нагромадження даних та їх ускладнення[1]. Важливо також відзначити те, що часто під цим поняттям у різних контекстах можуть мати на увазі як дані великого об'єму, так і набір інструментів та методів (наприклад, засоби масово-паралельної обробки даних системами категорії NoSQL, алгоритмами MapReduce, чи програмними каркасами проекту Hadoop). [[Big data](https://uk.wikipedia.org/wiki/%D0%92%D0%B5%D0%BB%D0%B8%D0%BA%D1%96_%D0%B4%D0%B0%D0%BD%D1%96)]

Існують прототипи, які реалізують повний життєвий цикл ([data.gov.ua](file:///E:\загрузки\data.gov.ua), [data.worldbank.org](file:///E:\загрузки\data.worldbank.org)), але предметом нашої розробки є не забезпечення повного життєвого циклу, а система управління даними.

Покажемо, які елементи життєвого циклу реалізовані у обраних прототипах, наведемо таблиці:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Створення даних | | | | | |
|  | |  | | --- | | планування дослідження | |  |  | |  |  | |  |  | |  |  | |  |  | | управління даними плану | погодження плану для спільного використання | пошук існуючих даних | збір даних | виокремлення і створення метаданих |
| <http://data.gov.ua/> | + | + | + | + | + | + |
| <http://data.worldbank.org/> |  |  |  | + | + | + |
| <http://chnppopen.wdc.org.ua/> |  |  | + | + | + | + |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Обробка даних | | | | |
|  | введення даних, оцифровка, транскрибування, переклад | перевірка, очищення даних | анонімізація даних в разі потреби | опис даних | управління та зберігання даних |
| <http://data.gov.ua/> | + | + |  | + | + |
| <http://data.worldbank.org/> | + | + |  | + | + |
| <http://chnppopen.wdc.org.ua/> | + | + | + | + | + |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Аналіз даних | | | | |
|  | Інтерпретація даних | отримання даних | створення результатів досліджень | автор публікації | підготовка даних для збереження |
| <http://data.gov.ua/> | + | + |  | + |  |
| <http://data.worldbank.org/> | + | + | + | + | + |
| <http://chnppopen.wdc.org.ua/> | + | + | + |  | + |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Збереження даних | | | | |
|  | перенесення даних в кращий формат | перенесення даних у відповідне середовище | резервне копіювання та зберігання даних | створення метаданих і документації | архівація даних |
| <http://data.gov.ua/> | + | + |  | + | + |
| <http://data.worldbank.org/> | + | + | + | + | + |
| <http://chnppopen.wdc.org.ua/> | + | + |  | + |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Надання доступу до даних | | | |
|  | поширення даних | контроль доступу | встановлення авторських прав | поділ даних |
| <http://data.gov.ua/> | + | + | + | + |
| <http://data.worldbank.org/> | + | + | + | + |
| <http://chnppopen.wdc.org.ua/> |  | + |  | + |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Повторне використання даних | | | | |
|  | подальші дослідження | нові дослідження | проведення науково-дослідних оглядів | ретельне вивчення результатів | вчити і вчитися |
| <http://data.gov.ua/> | + | + |  | + | + |
| <http://data.worldbank.org/> | + | + | + | + | + |
| <http://chnppopen.wdc.org.ua/> | + | + | + | + | + |