МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

ФАКУЛЬТЕТ ІНФОРМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ

КАФЕДРА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ

Документація до проекту

“Система управління даними”

з дисципліни “[Бази даних-1.](http://wiki.kpi.ua/index.php/%D0%91%D0%B0%D0%B7%D0%B8%20%D0%B4%D0%B0%D0%BD%D0%B8%D1%85-1.%20%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F%20%D0%B1%D0%B0%D0%B7%20%D0%B4%D0%B0%D0%BD%D0%B8%D1%85_(20402040)" \o "Бази даних-1. Проектування баз даних)

[Проектування баз даних](http://wiki.kpi.ua/index.php/%D0%91%D0%B0%D0%B7%D0%B8%20%D0%B4%D0%B0%D0%BD%D0%B8%D1%85-1.%20%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F%20%D0%B1%D0%B0%D0%B7%20%D0%B4%D0%B0%D0%BD%D0%B8%D1%85_(20402040)" \o "Бази даних-1. Проектування баз даних)”

Виконала:

студентка гр.ІП-54

Чиж Ю.М.

Номер залікової книжки**:** *5420*

Перевірив:

[доц. Болдак А. О.](http://rozklad.kpi.ua/Schedules/ViewSchedule.aspx?v=0d3b4492-f3b7-42c5-a4fe-1dbaa2d8ef8e)

Київ – 2017

**Запити зацікавлених осіб**

1. **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 використовуючи метадані. Для терміну *метадані* немає єдиного формального визначення. Навпаки, існують різні визначення цього терміну. Ось просте і популярне формулювання: *метадані* — це дані про дані. Цей термін в широкому сенсі 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 (data lifecycle)*, розуміють:

Створення даних (Data creation)

* дизайн дослідження
* управління даними плану (формати, зберігання і т.д.)
* згода плану для спільного використання
* знаходження існуючих даних
* збір даних (експеримент, спостереження, вимірювання, моделювання)
* створення метаданих

Обробка даних (Data processing)

* введення даних, оцифровка, транскрибування, переклад
* перевірка, очищення даних
* анонімізація даних в разі потреби
* опис даних
* управління та зберігання даних

Аналіз даних (Data analysis)

* інтерпретування даних
* отримання даних
* створення результатів досліджень
* визначення автора публікації
* підготовка даних для збереження

Збереження даних (Data preservation)

* 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 (Data Access)

* поширення даних
* контролювання доступу
* встановлення авторських прав
* розділення даних

Повторне використання даних (Data reuse)

* подальші дослідження
* нові дослідження
* ретельне вивчення результатів

[[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. Поставлене завдання містить специфічні задачі, які не можуть розглядатись тільки у межах одного стандарту (Дублінське ядро, наприклад).

*Дублінське ядро (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 (data management)* — це розробка і виконання архітектур, політики, практики і процедур з метою ефективного управління життєвим циклом інформації за потреби того чи іншого підприємства. [[Data management](http://searchdatamanagement.techtarget.com/definition/data-management)]
* *Набір даних (dataset)* являє собою сукупність деяких даних. Найчастіше набір даних відповідає змісту однієї таблиці бази даних, або однієї матриці статистичних даних, де кожен стовпець таблиці являє собою певну змінну, а кожен рядок відповідає певному члену набору даних, про який йде мова. Дані, установлені значення для кожної із змінних (таких як розмір об'єкта) існують для кожного елемента набору даних. Кожне значення відоме як точка прив'язки. Набір даних може містити дані для одного або більше членів, що відповідає числу рядків. Набір даних це термін, який також може бути використаний більш вільно, щоб звернутися до даних у якості тісно пов'язаних таблиць, які відповідають певному експерименту або події. Прикладом такого типу є набори даних, зібрані космічними агентствами, які здійснюють експерименти з інструментами на борту космічних апаратів. Деякі набори даних є настільки великими, що традиційні додатки для обробки даних не задовольняють поставлені задачі для роботи з великими обсягами даних. [[Data set](https://en.wikipedia.org/wiki/Data_set)]
* *Великі дані (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. Альтернативне визначення називає *великими даними* феноменальне прискорення нагромадження даних та їх ускладнення. 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)), проте 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 |
| data.gov.ua | + | + | + | + | + | + |
| data.worldbank.org |  |  |  | + | + | + |
| chnppopen.wdc.org.ua |  |  | + | + | + | + |

Таблиця 1.1 – Створення даних

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

Таблиця 1.2 – Обробка даних

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Аналіз даних** | | | | | |
|  | інтерпретація даних | отримання даних | створення результатів досліджень | автор публікації | підготовка даних для збереження |
| data.gov.ua | + | + |  | + |  |
| data.worldbank.org | + | + | + | + | + |
| chnppopen.wdc.org.ua | + | + | + |  | + |

Таблиця 1.3 – Аналіз даних

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

Таблиця 1.4 – Збереження даних

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Надання доступу до даних** | | | | |
|  | поширення даних | контроль доступу | встановлення авторських прав | поділ даних |
| data.gov.ua | + | + | + | + |
| data.worldbank.org | + | + | + | + |
| chnppopen.wdc.org.ua |  | + |  | + |

Таблиця 1.5 – Надання доступу до даних

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

Таблиця 1.6 – Повторне використання даних

1. **Сценарії виконання прецедентів**

**ID: IC001**

**Назва: Вхід**

**Учасники:** Користувач, система

**Передумови:** Користувач має намір ввійти в систему

**Результат:** Користувач в системі

**Основний сценарій:**

1. Користувач натискає відповідну кнопку для входу
2. Система відображає форму для авторизації користувача
3. Користувач вводить у відповідні поля свій логін та пароль
4. Система ідентифікує введені користувачем дані
5. Система автентифікує користувача

**Виключні ситуації:**

1. Введені користувачем дані невірні

**ID: IC002**

**Назва: Попередня реєстрація**

**Учасники:** Користувач, система.

**Передумови:** Користувач має намір зареєструватись в системі.

**Результат:** Створено неактивний профіль користувача.

**Основний сценарій:**

1. Користувач натискає відповідну кнопку для реєстрації.
2. Система відображає форму для реєстрації користувача.
3. Користувач вибирає опцію реєстрації.
4. Користувач заповнює форму своїми персональними даними, а саме:

* Прізвище, ім’я, ім’я по батькові;
* Контактна інформація;
* Логін;
* Пароль.

1. Система створює профільв неактивному стані.
2. Система надсилає на пошту, вказану користувачем, лист-підтвердження**,** що містить посилання з унікальним ідентифікатором, який розпізнається як ключ для активізації профілю.

**Виключні ситуації:**

1. Користувач не заповнив обов'язкові поля в формі для реєстрації.
2. Користувач вказав невірну електронну пошту.
3. Користувач вказав ненадійний пароль.
4. Користувач вказав існуючий логін.

**ID: IC003**

**Назва: Перегляд потрібного набору даних**

**Учасники:** Користувач, система.

**Передумови:** Користувач має намір переглянути певний набір даних.

**Результат:** Показано потрібний набір даних.

**Основний сценарій:**

1. Відвідувач обирає на потрібний набір даних.
2. Система відображає сторінку з обраним набором даних.

**Виключні ситуації:**

1. Набір даних відсутній.

**ID: IC004**

**Назва: Редагування профілю**

**Учасники:** Зареєстрований користувач, система.

**Передумови:** Зареєстрований користувач має намір редагувати профіль.

**Результат:** Профіль відредаговано.

**Основний сценарій:**

1. Зареєстрований користувач натискає кнопку для редагування власного профілю.
2. Зареєстрований користувач редагує дані профілю.
3. Зареєстрований користувач натискає кнопку підтвердження редагування.

**ID: IC005**

**Назва: Активація профілю**

**Учасники:** Модератор, система.

**Передумова:** система повідомила модератору про наявність неактивних профілів.

**Результат:** Профіль набув статусу активного.

**Основний сценарій:**

1. Модератор натискає на кнопку для візуалізації списку неактивних профілів.
2. Система формує та відображає список неактивних профілів.
3. Модератор переглядає один із профілів.
4. Модератор натискає кнопку для переводу профілю в активний стан.
5. Система переводить профіль в активний стан.
6. Система надсилає лист-повідомлення про активацію профілю на електронну пошту зареєстрованого користувача.

**Виключні ситуації:**

1. Модератор не схвалює активацію профілю.

**ID: IC006**

**Назва: Деактивація профілю**

**Учасники:** Модератор, система.

**Передумови:** Модератор має намір деактивувати профіль.

**Результат:** Профіль набув статусу неактивного.

**Основний сценарій:**

1. Модератор виявив профіль зареєстрованого користувача, який не відповідає правилам.
2. Модератор натискає на кнопку для деактивації профілю зареєстрованого користувача.
3. Модератор вказує причину та термін деактивації.
4. Профіль деактивований.
5. Система надсилає лист-повідомлення про деактивацію профілю на електронну пошту зареєстрованого користувача.

**ID: IC007**

**Назва: Перегляд профілю**

**Учасники:** Модератор, система.

**Передумови:** Модератор бажає переглянути профіль зареєстрованого користувача.

**Результат:** Відображено профіль.

**Основний сценарій:**

1. Модератор натискає кнопку для перегляду профілю зареєстрованого користувача.
2. Система відображає модератору дані профілю.

**ID: IC008**

**Назва: Редагування профілю**

**Учасники:** Модератор, система.

**Передумови:** Модератор має намір редагувати профіль зареєстрованого користувача. **Результат:** Профіль відредаговано.

**Основний сценарій:**

1. Модератор натискає кнопку для редагування профілю.
2. Модератор редагує дані профілю.
3. Модератор натискає кнопку підтвердження редагування.

**ID: IC009**

**Назва: Створення нового набору даних**

**Учасники:** Модератор, система.

**Передумови:** Модератор має намір створити новий набір даних.

**Результат:** Набір даних створено.

**Основний сценарій:**

1. Модератор натиснув на кнопку для створення нового набору даних.
2. Система відображає форму для створення нового набору даних.
3. Модератор заповняє необхідні метадані та вказує джерело набору даних.
4. Система опубліковує новий набір даних.

**Виключні ситуації:**

1. Модератор не заповнив обов'язкові поля.

**ID: IC0010**

**Назва: Редагування набору даних**

**Учасники:** Модератор, система.

**Передумови:** Модератор має намір редагувати набір даних.

**Результат:** Набір даних відредаговано.

**Основний сценарій:**

1. Модератор натиснув на кнопку для редагування набору даних.
2. Система відображає форму для редагування набору даних.
3. Модератор редагує набір даних або метадані.
4. Система опубліковує відредагований набір даних.

**ID: IC0011**

**Назва: Вихід**

**Учасники:** Зареєстрований користувач, система.

**Передумови:** Зареєстрований користувач має намір вийти з профілю.

**Результат:** Здійснено вихід з профілю.

**Основний сценарій:**

1. Зареєстрований користувач натискає кнопку для виходу з власного профілю.
2. Система завершує поточний сеанс роботи.
3. **Бізнес сценарії**

**ID: UC 001**

**Назва: Сценарій обробки заявки на внесення нового набору даних користувачем**

**Учасники:** Власник даних, система, модератор.

**Передумови:** Користувач має намір створити новий набір даних.

**Результат:** Внесено новий набір даних.

**Основний сценарій:**

1. Власник даних отримує список вимог для внесення нового набору.

2. Власник даних заповнює форму реєстрації набору даних.

3. Власник відправляє заявку для її підтвердження.

4. Система реєструє заявку.

5. Система обробляє заявку.

6. Система надає модератору системи інформацію щодо перевірки поданої заявки.

7. Модератор згідно отриманої інформації відправляє власнику набору даних відповідь щодо заявки.

8. Власник одержує відповідь.

9. Система переводить заявку в стан «підтверджено».

10. Власник одержує відповідь.

**Виключні ситуації***:*

1. Власник набору даних отримав негативну відповідь.
2. Модератор отримує питання щодо реєстрації (UC 003).

**ID: UC 002**

**Назва: Сценарій використання даних користувачем**

**Учасники:** Користувач, система.

**Передумови:** Користувач має намір використати доступні дані.

**Результат:** Необхідні дані використані.

**Основний сценарій:**

1. Користувач відкриває список всіх наборів даних.

2. Користувач відфільтровує дані через доступні йому метадані.

3. Користувач здійснює пошук по назві/тегам.

4. Після знайдення шуканого набору даних користувач заходить на його сторінку для отримання наступних метаданих:

* + Дата додавання.
  + Автори/Власники.
  + Кількість завантажень.

5. Користувач переходить за посиланнями, які ведуть до завантаження набору даних.

**Виключні ситуації:** відсутні.

**ID: UC 003**

**Назва: Сценарій обробки питань від користувачів**

**Учасники:** Користувач, модератор.

**Передумови:** У користувача виникли питання.

**Результат:** Отримано відповідь на задані питання.

**Основний сценарій:**

1. Користувач надсилає модератору питання.
2. Модератор отримує список заданих йому питань.
3. Модератор надсилає на пошту користувача відповіді на задані питання.

**Виключні ситуації:** відсутні.

1. **Use-case діаграма**

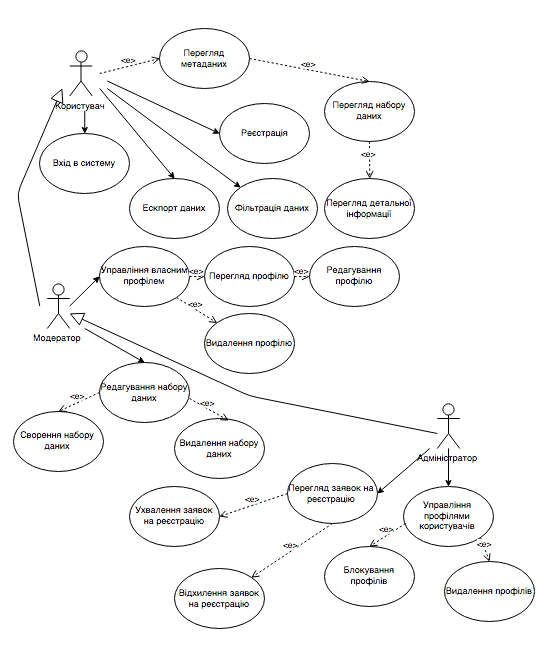
****

Рисунок 4 – Use-case діаграма

1. **Діаграма сутностей**

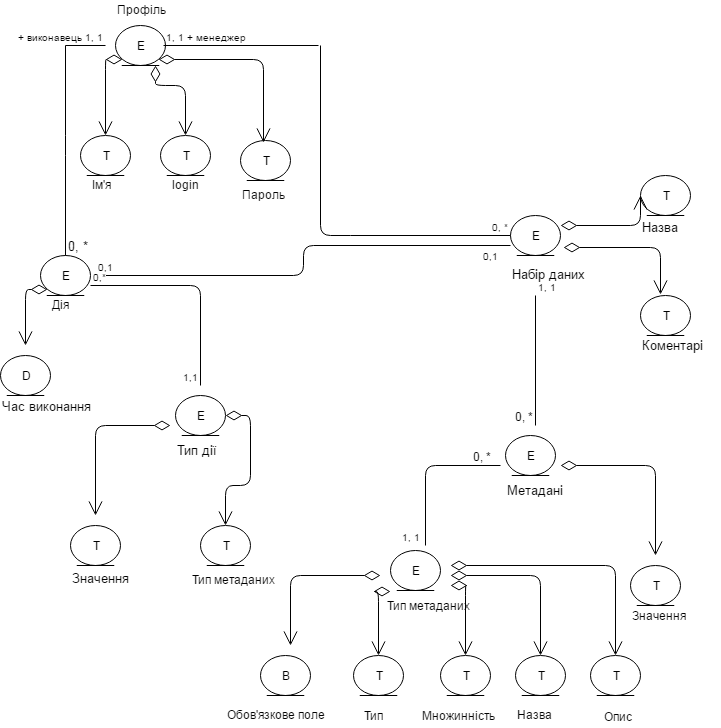
****

Рисунок 5 – Діаграма сутностей

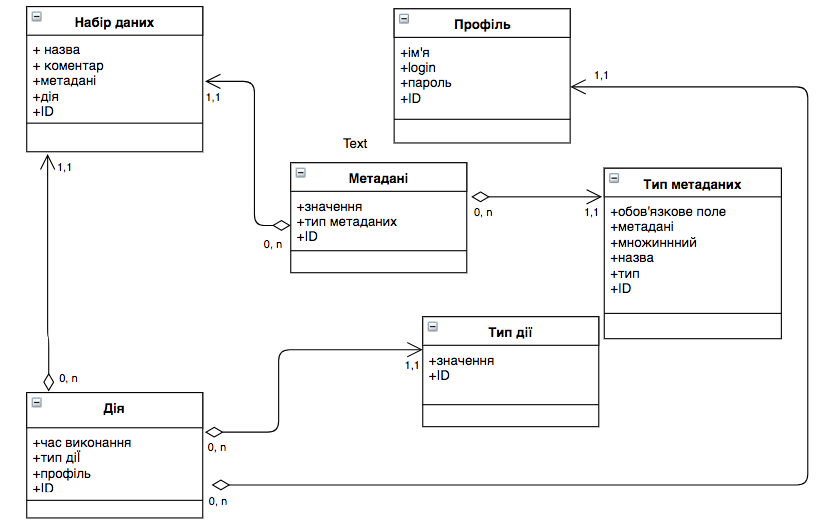
1. **Діаграма класів **

Рисунок 6 – Діаграма класів

1. **Реляційна база даних**

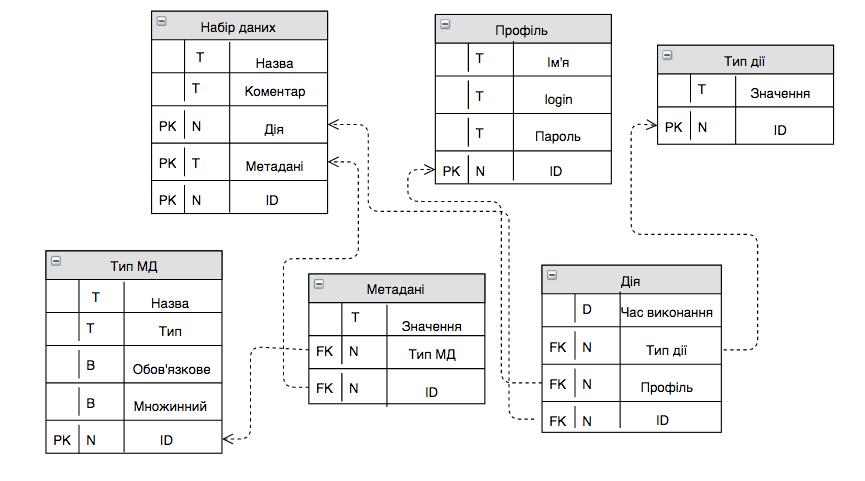
****

Рисунок 7 – Реляційна база даних