Проект: EXCELLCITY: Повишаване на научния и иновационен капацитет в интелигентния град чрез изграждане на връзки (ТУИНИНГ) BG-RRP-2.005-0003

ПЛАН ЗА УПРАВЛЕНИЕ НА ДАННИ

Доклад за 3.5

РАБОТЕН ПАКЕТ 3

Бенефициент: Университет по библиотекознание и информационни технологии (УниБИТ)

Структурата за наблюдение и докладване: Изпълнителна агенция „Програма за образование“

Автори на документа	Александър Шикаланов и Денис Чикуртев Подпис_____ Дата: _____
Ръководител на РП …	…….. Подпис_____ Дата: _____

Одобрено от Координационния съвет на проекта	Дата: _____

Версия на документа	2
Дата на версията	20.04.2025

Финансирано от Европейския съюз NextGenerationEU

ИЗПЪЛНИТЕЛНА АГЕНЦИЯ „ПРОГРАМА ЗА ОБРАЗОВАНИЕ“

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА

Контрол на документа

Номер на версията	Автор	Дата на версията	Статус	Одобрено от
1	Александър Шикаланов	20.02.2024	Първа чернова за вътрешен преглед	Денис Чикуртев
2	Денис Чикуртев	20.04.2025	Финална версия	Евгения Ковачева

Контрол на промените

Номер на променената версия	Раздели за промяна	Обобщение на причините за промяна	Автор	Нова версия
0.1	Всичко	Вътрешен преглед		

Разпределение

Всички версии на този документ се разпространяват до следните лица:

Съдържание

1. Принципи при изработване на плана за управление на данни

Разработваният План за управление на данните (Data Management Plan – DMP) е базиран на насоките на Европейската комисия, които описват жизнения цикъл на управлението на данните, които ще бъдат събирани, обработвани и/или генерирани в рамките на проекта, както и подхода за тяхното правилно управление съгласно принципите FAIR.

Този документ ще бъде регулярно актуализиран и ще предоставя подробна информация относно данните (набори от данни, други резултати от изследванията) и тяхното управление в съответствие с Общия регламент за защита на данните (GDPR), както и с другото приложимо законодателство на ЕС и местните разпоредби. Описаната методология на проекта отговаря на принципа *да не се нанасят значителни вреди*, съгласно член 17 от Регламент (ЕС) № 2020/852.

Дейностите по проекта следват принципа *да не се нанасят значителни вреди*, въведен от Европейската комисия, тоест, когато е осъществимо, подходящо и пропорционално, в съответствие със съответните секторни правила, те ще бъдат реализирани така, че да постигнат заложените цели без значителна вреда за екологичните цели, свързани със смекчаване изменението на климата, адаптация към изменението на климата, устойчивото използване и защита на водните и морските ресурси, прехода към кръгова икономика, предотвратяването и контрола на замърсяването, както и защитата и възстановяването на биологичното разнообразие и екосистемите.

Следвайки етично ориентиран и фокусиран върху човека подход към изкуствения интелект и възприемайки най-новите методологии за преодоляване на разликите между разработването на надежден изкуствен интелект и техническите и методологични практики, Excel City използва подхода Z-Inspection®. Този подход позволява задълбочена оценка и валидиране на разработената AI рамка, като същевременно гарантира, че всички създадени компоненти са в съответствие с европейските насоки за надежден изкуствен интелект. Това се осъществява чрез разработването на социално-технически сценарии, които извеждат на преден план всички етични проблеми и напрежения, следени чрез ясни количествени показатели.

В допълнение към това, схемите на базите данни са проектирани, като се вземат предвид съответните законодателни аспекти (Акт за изкуствения интелект, Закон за данните и управлението на данните, Общия регламент за защита на данните (GDPR) и др.), предоставяйки рамка, която отговаря на настоящите изисквания. Оценката на въздействието върху поверителността и защитата на данните (DPIA), както и Оценката на социалното въздействие (SIA), са ключови процеси през целия жизнен цикъл на проекта. Финалните резултати се оценяват и валидират, за да се гарантира, че са в съответствие с европейския подход към изкуствения интелект.

Данните се съхраняват на сървър на УниБИТ в PostgreSQL/MySQL база данни, с използване на Keyrock за оторизация:

Актуализиран Пълен План за Управление на Събрани Данни от FIWARE Система (Университетски Сървър, PostgreSQL/MySQL, Keyrock). Описанието на плана включва база данни PostgreSQL, но се взимат в предвид използване и на MySQL по аналогичен ред.

2. Определяне на Целите и Обхвата (Без Промяна):

- Основни цели:
    - Анализ на енергийната ефективност на сградата за идентифициране на възможности за оптимизация и намаляване на консумацията.
    - Анализ на комфортна в сградата за осигуряване на здравословна и благоприятна учебна среда.
    - Анализ на моделите на велосипедни разходки за подобряване на инфраструктурата, планиране на събития или насърчаване на устойчив транспорт.
    - Предоставяне на достъп до анонимизирани или обобщени данни на изследователи и други оторизирани лица след заявка.
- Типове данни:
    - Енергийна ефективност: Консумация на електроенергия (обща и по зони/уреди), данни от сензори за температура, влажност, осветеност, данни от HVAC системи и други свързани параметри.
    - Велосипедни разходки: GPS координати, времеви печати, продължителност на пътуванията, изминато разстояние (ако е налично от устройствата на потребителите).
- FIWARE Компоненти: (Зависят от конкретна FIWARE архитектура, но могат да включват):
    - IoT Agents (за свързване на сензори и устройства).
    - Context Broker (Orion LD) за управление на контекстни данни.
    - Data Handling компоненти (като Cygnus за персистиране в PostgreSQL).
- Обхват на плана: Включва събиране, съхранение в PostgreSQL на университетския сървър, управление на достъпа чрез Keyrock, обработка, анализ, запазване, архивиране и изтриване на данните.
- Заинтересовани страни:
    - Екип по поддръжка на сградния фонд.
    - Екип по информационни технологии на университета.
    - Изследователски екипи.
    - Администрация на университета.
    - Студенти (възможно за анонимизирани обобщени данни).

3. Събиране и Съхранение на Данни (Конкретизация за PostgreSQL):

- Събиране: Данните се събират от крайните устройства и сензори чрез IoT Agents и се управляват от Context Broker. Cygnus или друг FIWARE компонент за персистиране и е

конфигуриран да записва данните в PostgreSQL базата данни на университетския сървър.

- Съхранение (PostgreSQL):
    - Структура на базата данни: проектират се схеми и таблици в PostgreSQL, които да отговарят на структурата на данните от енергийната ефективност и велосипедните разходки.
        - Таблица за енергийна ефективност: energy_data (timestamp TIMESTAMP WITH TIME ZONE, location_id TEXT, sensor_id TEXT, energy_consumption NUMERIC, temperature NUMERIC, humidity NUMERIC, light_level NUMERIC, ...). Обръща се внимание на нормализацията, използването на подходящи типове данни (например TIMESTAMP WITH TIME ZONE за времеви данни) и индексирането за ефективно търсене.
        - Таблица за велосипедни разходки: bike_rides (ride_id UUID PRIMARY KEY, user_id UUID, start_time TIMESTAMP WITH TIME ZONE, end_time TIMESTAMP WITH TIME ZONE, duration INTERVAL, start_latitude DOUBLE PRECISION, start_longitude DOUBLE PRECISION, end_latitude DOUBLE PRECISION, end_longitude DOUBLE PRECISION,[1] distance NUMERIC). Използването на UUID за уникални идентификатори е препоръчително.
        - Схема за данните свързани с проследяване наличието на замърсители на въздуха в София. Разглеждат се два модела на данните – релационен (PostgreSQL) и NoSQL с използване на MongoDB.
    - Схема на данните: Ще се документира подробно схемата на всяка таблица, типовете данни, връзките между таблиците (включително външни ключове, ако е приложимо) и ограниченията (constraints).
    - Сигурност по време на събиране и съхранение:
        - Комуникацията между FIWARE компонентите и PostgreSQL е защитена (например чрез TLS/SSL криптиране).
        - Достъпът до PostgreSQL сървъра, както и до останалите компоненти съдържащи и предоставящи данни, е ограничен само до оторизирани IP адреси и потребители.
        - Използват се силни пароли за достъп до базата данни. Обмисля се използването на role-based access control (RBAC) в PostgreSQL.
    - Интегритет на данните:
        - Прилага се валидация на данните на ниво приложение и/или база данни (например чрез CHECK constraints, NOT NULL ограничения, уникални индекси).
        - Имплементират се механизми за откриване и обработка на липсващи или невалидни данни.

- Капацитет и управление на нарастването: Ще се следи използваното дисково пространство и ще се планира разширяване при необходимост. Ще се обмислят стратегии за партициониране на таблиците (например time-based partitioning) за по-добра производителност при голям обем данни.

4. Обработка и Анализ на Данни (Без Съществени Промени):

- Обработка:
  - Почистване: Премахване на дубликати (могат да се използват DISTINCT клаузи в SQL), обработка на липсващи стойности (COALESCE функция в SQL), коригиране на грешни данни (чрез UPDATE заявки).
  - Трансформация: Преобразуване на данни в подходящи формати за анализ (например агрегиране на данни с помощта на GROUP BY клаузи и агрегатни функции).
  - Агрегация: Създаване на обобщени данни (например средна консумация на енергия за даден период, общ брой велосипедни пътувания с помощта на SQL заявки).
- Инструменти и технологии за обработка: Могат да включват SQL заявки директно към PostgreSQL, скриптове на Python с библиотеки като Psycopg2 (PostgreSQL адаптер) и Pandas, или други ETL инструменти, които поддържат PostgreSQL.
- Анализ:
  - Енергийна ефективност: Анализ на тенденции в консумацията, идентифициране на пикове и спадове, сравнение на консумацията между различни периоди или зони, откриване на аномалии (могат да се използват статистически функции в PostgreSQL или външни инструменти).
  - Велосипедни разходки: Анализ на най-популярни маршрути (пространствен анализ с PostGIS разширението на PostgreSQL може да бъде полезен), времеви модели на пътувания, средна продължителност и разстояние, идентифициране на потенциални проблеми в инфраструктурата.
- Инструменти и технологии за анализ: Могат да включват BI инструменти (например Grafana, свързан с PostgreSQL), статистически софтуер (например R, свързан с PostgreSQL), Python библиотеки (например Matplotlib, Seaborn, Scikit-learn, свързани с PostgreSQL).
- Гарантиране на качеството: Резултатите от обработката и анализа ще бъдат валидирани и сравнявани с очаквани стойности или други източници на информация. Ще се документират методите за обработка и анализ.

5. Управление на Качеството на Данните (Без Съществени Промени):

- Дефиниции за качество: (както преди)

- Измерване и наблюдение: Създадат се SQL заявки за периодично проверка на качеството на данните (например използване на `COUNT`, `AVG`, `MAX`, `MIN` с `GROUP BY` и `HAVING` клаузи за идентифициране на аномалии или непълнота). Резултатите ще се следят на табла за управление.
- Процеси за коригиране: (както преди)
- Документиране на проблеми: (както преди)

6. Сигурност и Достъп до Данни (Конкретизация за Keyrock):

- Политики за сигурност: Спазват вътрешните политики за сигурност на университета, както и специфични изисквания, свързани с обработката на данни.
- Управление на достъпа (Keyrock):
    - Keyrock е конфигуриран като Identity Manager за FIWARE платформата.
    - Потребителите и организациите се управляват в Keyrock.
    - Приложенията, които искат достъп до данните (например инструменти за анализ, BI платформи), са регистрирани като услуги в Keyrock.
    - Разрешенията за достъп до данните се базират на OAuth2 токени, издадени от Keyrock.
    - API Gateway (например Kong, ако се използва във FIWARE архитектурата) валидира тези токени преди да разреши достъп до бекенд услугите, които работят с PostgreSQL.
    - На ниво PostgreSQL, достъпът е ограничен до определени потребители, които ще имат права само за тези данни, до които са оторизирани от Keyrock. Това се постига чрез създаване на различни PostgreSQL роли и предоставяне на специфични привилегии.
- Заявка от клиента: Процесът за получаване на достъп ще включва:
    - Клиентът (например изследовател) подава заявка за достъп през определена процедура (например формуляр до ИТ отдела).
    - След одобрение, на клиента се създава акаунт в Keyrock или се предоставят необходимите роли и разрешения в съществуващ акаунт.
    - Приложението на клиента използва OAuth2 flow (например Client Credentials Grant или Authorization Code Grant), за да получи токен от Keyrock.
    - Този токен се използва при заявки към API-тата, които предоставят достъп до данните в PostgreSQL. API Gateway проверява валидността на токена и разрешенията.

Одит на достъпа: Водят се логове на ниво Keyrock за автентикация и авторизация, както и логове на ниво API Gateway за достъп до ресурсите. Могат да се следят и логовете на PostgreSQL за връзки и изпълнени заявки (с внимание към производителността).

Защита срещу загуба на данни: извършват се редовни резервни копия на PostgreSQL базата данни на сигурно място (съгласно политиките на университета). Тестват се процедурите за възстановяване на данни.

7. Защита на потребителските данни GDPR

Спазване на GDPR: Прилагат мерки за анонимизиране или псевдонимизиране на лични данни (ако има такива), преди предоставянето им за анализ или на трети страни, освен ако няма законно основание за обработка на идентифицирани лични данни. Keyrock може да помогне за управление на съгласието и правата на субектите на данни.

Защитата на потребителските данни е правно задължителен процес, който изисква съответствие с Общия регламент относно защитата на данните (GDPR) и националния Закон за защита на личните данни (ЗЗЛД). Успешният План за управление трябва да гарантира законосъобразност, прозрачност, минимална обработка и сигурност на информацията през целия ѝ жизнен цикъл. [1, 2, 3, 4]

8. Запазване и Архивиране на Данни (Съобразяване с Университетски Политики):

- Политики за запазване: (както преди)
- Архивиране: Данните могат да се архивират с помощта на вградени инструменти на PostgreSQL (например pg_dump) или други инструменти за архивиране на файлова система. Архивите трябва да се съхраняват на сигурно и отделно място.
- Възстановяване от архиви: Ще се документират процедури за възстановяване на данни от архивите с помощта на pg_restore или други подходящи методи.
- Сигурност и интегритет на архивите: Архивираните данни ще бъдат криптирани и ще се проверява тяхната цялост (например чрез контролни суми).

9. Изтриване на Данни (Съобразяване с Университетски Политики и GDPR):

- Политики за изтриване: (както преди)
- Процедури за сигурно изтриване: Данните в PostgreSQL ще бъдат изтривани чрез стандартни SQL команди за изтриване (DELETE). За пълно премахване на данните от диска (ако е необходимо за чувствителна информация), могат да се използват по-сложни техники, но това трябва да се планира внимателно и да се съобрази с университетските политики за съхранение.
- Документиране на изтриването: (както преди)

10. Роли и Отговорности (С Акцент върху Keyrock):

- Администратор на база данни (PostgreSQL): Отговаря за поддръжката, сигурността, резервните копия и производителността на PostgreSQL сървъра.
- FIWARE Администратор: Отговаря за поддръжката и конфигурацията на FIWARE компонентите, включително Keyrock и връзката с PostgreSQL (например конфигурация на Cygnus).
- Администратор на Keyrock: Отговаря за управлението на потребители, организации, приложения и разрешения в Keyrock.
- Служител по сигурността на данните: Отговаря за прилагането на политиките за сигурност и управление на достъпа (включително конфигурацията на Keyrock и PostgreSQL).
- Анализатори на данни: Отговарят за обработката и анализа на данните, като използват предоставения им достъп.
- Собственици на данните: Определят нуждите за събиране на данни и техните изисквания за достъп и използване.
- ИТ отдел на университета: Осигурява инфраструктурата и поддръжката.

11. Документация и Обучение (С Акцент върху Keyrock):

- Документация:
    - Този актуализиран план за управление на данни.
    - Схема на PostgreSQL базата данни.
    - API документация на FIWARE компонентите, използвани за събиране и персистиране на данни.
    - Документация за инсталация и конфигурация на Keyrock.
    - Процедури за получаване на достъп до данните през Keyrock.
    - Процедури за обработка и анализ на данните.
    - Политики за сигурност и запазване на данни.
    - Ръководства за потребители как да получават и използват OAuth2 токени от Keyrock за достъп до данните.
- Обучение: Ще се провеждат обучения за съответните лица относно политиките и процедурите за управление на данни, използването на Keyrock за оторизация и работата с PostgreSQL.

12. Преглед и Актуализация на Плана:

- Планът се преглежда и актуализира поне веднъж годишно или при значителни промени в системата, регулациите или нуждите на университета.

- Отговорност за прегледа и актуализацията носи определен екип или лице (например служител по сигурността на данните съвместно с ИТ отдела и FIWARE администратора).

Използвана литература:

1. GDPR, Общ регламент относно защитата на данните, https://gdpr-text.com/bg/
2. MultiLaw, Global Data Protection Law Guide , Dimitrov, Petrov & Co, https://multilaw.com/Multilaw/Multilaw/Data_Protection_Laws_Guide/DataProtection_Guide_Bulgaria.aspx
3. BFBHR, Data Protection and Cybersecurity Policy, https://bfbhr.org/en/data-protection-policy/
4. CMS (2025) Data protection and cybersecurity laws in Bulgaria, https://cms.law/en/int/expert-guides/cms-expert-guide-to-data-protection-and-cyber-security-laws/bulgaria