

Цифрові інструменти оцінки даних на відповідність принципам FAIR

Чуканова С.О., канд. пед.н., Сертифікований тренер Google for Education

chukanovaso@ukma.edu.ua

FAIR принципи

FAIR – це абревіатура, що розшифровується як **Findable, Accessible, Interoperable, Reusable**:

Findable

Відшукуваність

Accessible

Доступність

Interoperable

Сумісність (інтероперабельність)

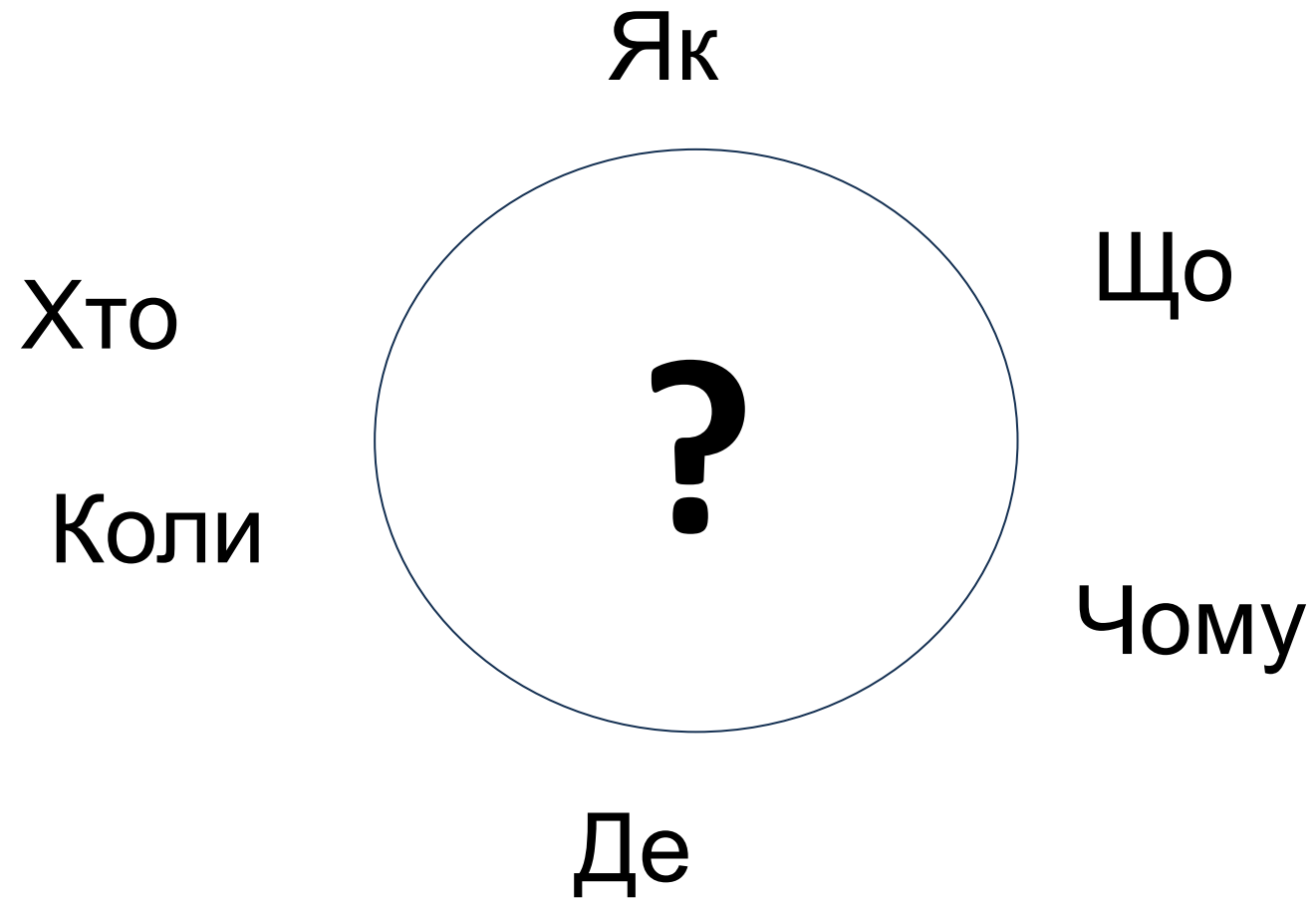
Reusable

Повторне використання
(багаторазовість)

FAIR принципи

Найкращий спосіб привести дані у відповідність до цих принципів – це використання репозитарію, що відповідає принципам FAIR та стандартам на кшталт Core Trust Seal. Відповідний репозитарій – це вже більша половина підготовки

Findable



Findable

1. Перший крок – переконатись, що ваші дані можуть знайти як люди, так і машини. Найкращий спосіб – це скористатись FAIR-сумісним репозитарієм, визнаним у вашій галузі. Якщо ваша установа не має власного репозитарію, ви можете скористатись загальними, напр. Dryad, Figshare, Zenodo.
2. Упевніться в тому, що метадані, які описують ваші дані є придатними до опрацювання людиною та машиною. Для цього потрібно застосувати схему стандарту Dublin Core чи Data Cite. Якщо ви використовуєте репозитарій для даних, то він скоріш за все, послуговується визнаними стандартами метаданих, тож при депонуванні, ви просто маєте заповнити відповідні поля. Наявність сталого ідентифікатора є обов'язковою.
3. Для повного опису даних варто використовувати стандарт метаданих, специфічний за дисципліною. Постійний ідентифікатор – ключовий аспект FAIR.

Findable

- (Мета)даним присвоюється унікальний постійний ідентифікатор
- Дані описують за допомогою розширених метаданих
- Метадані мають чіткий та прозорий ідентифікатор даних, що вони описують
- (Мета)дані зареєстровані чи індексуються у пошуковому ресурсі (репозитарій: специфічний за дисципліною, міждисциплінарний, інституційний)
 - Критерії вибору репозитарію: сертифікація, надання постійних ідентифікаторів, доступ до даних, чіткі умови використання та депонування



Findable

● Легко знайти людям та комп'ютерними системам

● Базується на обов'язковому описі матаданих, що сприяє видимості потрібних пакетів даних

Accessible

До даних потрібно отримати доступ. Якщо ви використовуєте репозитарій, специфічний за дисципліною або загальний – він послуговується стандартними протоколами, що підтримують онлайн доступ.

FAIR ≠ Open

Різні рівні доступу: safe heaven (авторизація), авторизований, автентифікація. В записі метаданих потрібно прописати варіанти доступу (statement), відповідну ліцензію, тривалість зберігання (проте метадані зберігаються завжди)

Accessible

- (Мета)дані мають властивість відновлюватись за допомогою їх ідентифікатора з використанням стандартизованого комунікативного протоколу
- Протокол відкритий, вільний, універсального застосування
- Протокол дозволяє аутентифікацію авторизацію за потреби
- Метадані доступні, навіть тоді, коли дані вже не доступні
- Доступ : відкритий, обмежений, закритий
- Автентифікація та авторизація: приватність, цілісність, доступність

Метадані

- Дані про дані
- Структурована інформація (незалежні дані) про інші дані чи ресурси та їх характеристики, що додається вручну або автоматично. Важливо для машинного зчитування.
- Метадані: змістові (автор, опис – важливо для цитування), адміністративні (власник, головний дослідник тощо – контекст проекту) ,технічні (дата запису, геодані – створюються автоматично),структурні (юніти, метод збору тощо – інсайт).



Accessible

● Довготривале зберігання, щоб до даних можна було легко доступитись та/чи завантажити .

● Чітко визначена ліцензія використання та умови доступу на рівні метаданих або контенту даних

Interoperable

Щоб ваші дані можна було повторно використовувати, забезпечте їхню інтеграцію з іншими даними так, щоб їх можна було використати через застосунки чи здійснити процеси аналізу, збереження, обробки та довготривалого використання.

Уникайте використання пропрієтарного ПЗ при створенні даних. Звертайте увагу на вимоги репозитаріїв щодо бажаних форматів файлів.

Для забезпечення автоматизації процесу знаходження та інтероперабельності пакетів даних важливо застосовувати контрольовані сховища, щоб переконатись, що ваші дані добре структуровані та використовуються визначені моделі даних.

Interoperable

- (Мета)дані використовують формальну, доступну, зрозумілу, широковживану лексику для відтворення знання
- (Мета)дані використовують вакабуляри, що відповідають принципам FAIR
- (Мета)дані містять якісні посилання на інші (мета)дані
- Мають бути у загальноприйнятому форматі
- Визначені стандарти (зрозумілість, інтероперабельність, інтерпретованість даних)
- Авторитетні файли слугують унікальній ідентифікації людей, установ, фондів тощо
- Контрольований словник (лексика) : класифікація – для розподілу об'єктів , тезаурус – лінгвістична упорядкована колекція термінів та їх співвідношення.



Interoperable

- Дані готові до поєднання з іншими пакетами даних, зібраних людьми або пристроями.

Reusable

- Щоб збільшити потенційне повторне використання ваших даних, переконайтесь, що вони та пов'язані з ними метадані добре описані так, що їх можна відтворити або поєднати у різні набори.
- Визначте ліцензію, якщо нульова, то пропишіть Creative Commons. Дані про ліцензію потрібно внести в опис метаданих, те саме стосується ПЗ для обробки даних (можна використати License Selector).
- Потрібно подати детальну документацію: опис методології дослідження, файл, одиниця, рівень змінних, щоб третя сторона могла зрозуміти ваші дані.
- Пояснити, які рядки в Excel є чутливими, чи є окремий документ зі скороченнями та аббревіатурами. Також потребує опису використане ПЗ. Краще поєднати ваші дані з відповідними результатами – це забезпечить контекстну рамку для розуміння, яким чином ви дійшли таких висновків.
- Ізольовані дані часто не мають сенсу, не забувайте приєднувати результати в ORCID.
- Включіть інформацію про цитування ваших даних (автор, назва, видавець, дата, ідентифікатор).

Reusable

- (Мета)дані детально описані. Мають атрибуції авторства
- (Мета)дані мають чіткі та доступні ліцензії використання
- (Мета)дані мають детальний опис
- (Мета)дані відповідають основним стандартам галузі
- Ліцензування даних : Creative Commons, Open Data Commons, Software Licences



Reusable

- Дані готові до використання у майбутніх дослідженнях та до подальшої обробки з використанням обчислювальних методів

DMP та FAIR

DMP описує рівень і формат даних, які ви генеруєте, збираєте або повторно використовуєте під час дослідження. Він також описує як ви будете ними управляти та обмінюватись під час проекту та після його завершення у довгостроковій перспективі.

Якщо робити дані відповідними до принципів FAIR в ретроспективі, то це може бути дорожче та важче, ніж спланувати це наперед

FAIR Data Assessment Tools

ARDC Fair Data self-assessment tool (Australian Research Data Commons) оцінює дані на відповідність принципам FAIR та дає поради з покращення цієї відповідності. Онлайн опитувальник на основі пов'язаних між собою запитань згрупованих відповідно до кожного з принципів FAIR.

DANS FAIRdat tool визначає наскільки вже існуючі дані відповідають принципам FAIR. Онлайн опитувальник.

EUDat FAIR Data Checklist – pdf чеклист

F-UJI Automated FAIR Data Assessment Tool – Оцінка F-UJI базується на 16 із 17 основних показників оцінки об'єктів FAIR, розроблених у рамках FAIRsFAIR, і кожен з яких відповідає частково або повністю принципу FAIR.

F-UJI дотримується існуючих веб-стандартів і найкращих практик служб вирішення PID і використовує зовнішні реєстри та ресурси, такі як re3data ,Datacite API, SPDX License List, RDA Metadata Standards Catalog , Linked Open Vocabularies (LOV)



Практика



ARDC Fair Data self-assessment tool



5 Star Data Rating Tool



DANS FAIRdat tool



F-UJI Automated FAIR Data Assessment Tool