

[Мовець 1]

Управління даними досліджень або науковими даними. Власне, управління даними досліджень – це така система заходів для того, щоб можна було ефективно управляти даними протягом усього життєвого циклу. Від того, як створюються дані або збираються дані, як вони організовані, зберігаються, в тому числі в довгостроковому аспекті, як вони повторно використовуються після того, як уже завершено дослідження.

Основна мета – РДМ – це Research Data Management. Власне, термін дуже розповсюджений і ця аббревіатура дуже часто використовується, ви часто її знайдете і в українських текстах іноді навіть не перекладаються, так і використовують РДМ – Research Data Management. Основна задача – це забезпечити, щоб дані були стійкими, доступними для пошуку, безпечними, зрозумілими для використання.

Власне, оцей життєвий цикл, тому що управління дослідницькими даними відбувається протягом усього життєвого циклу даних. Значить, розуміння життєвого циклу по-різному його структурують, по-різному його прописують. Життєвий цикл даних починається, звісно, від планування, тому що треба визначити види даних і процес управління ними.

Наступний етап – це збір даних або створення даних, в залежності від того, що у вас за дослідження. Відповідно, це можуть бути експерименти, вимірювання, опитування, моделювання, або вже пошук даних, їх збір в результаті вже проведених інших досліджень. Наступний етап – це обробка даних, первинна очистка, структурування, створення даних у форматі, який придатний для їх подальшої обробки і, можливо, виконання зміни і так далі.

Далі – аналіз даних. Використовуються методи та інструменти для того, щоб аналізувати їх, інтерпретувати. Також наступний етап – це документування даних.

Він включає якраз їх опис, підготовку метаданих за тими стандартами, які задані передусім науковою організацією, але так само і дослідник самостійно може якісь параметри важливі для нього визначити. Зберігання даних – це визначення того, яким чином забезпечене не просто зберігання, але й захищеність даних в стабільних системах. Поширення і публікація – це надання доступу даних для інших дослідників і суспільства, звісно, з урахуванням тих правил, які передбачаються в ліцензії, яку ви вже собі виберете для публікації і поширення цих даних.

Власне, на стадії поширення публікацій даних потрібно застосовувати обов'язково принципи FAIR. Також в життєвий цикл наукових даних входить їх повторне використання, це використання їх в наступних дослідженнях, в освіті, використання іншими дослідниками так само. Ще один етап – це довготривале збереження даних, яким чином, де вони будуть зберігатися, як буде забезпечуватись їх захист.

План управління дослідницькими даними – це такий дуже важливий документ, який описує, яким чином буде відбуватись управління науковими даними протягом його усього життєвого циклу. А навіщо він потрібен, такий план? Ну, по-перше, це управлінська позиція.

З точки зору управління і організації процесу має бути план. Будь-яке управління починається, звісно, з плану, який задає рамку, що буде робитись, як буде робитись, хто це буде робити. Але дуже важливий момент – це змістовне наповнення.

Навіщо це потрібно? Тобто, що воно дає запровадження отакої практики? Справа в тім, що наукові дані, результати досліджень, фактично можуть діяти як ресурс, який приносить додану вартість.

Що це означає? Це означає, що отримані результати досліджень є основою для наступних досліджень, для впровадження в практику результатів, для використання не тільки бізнесом, але й в державному управлінні, місцевому самоврядуванні. Це означає можливість співпраці, розширення можливостей наукових в результаті такої співпраці.

Тобто, поступ в науці, поступ в практиці, в нашому людському житті, розвиток бізнесу, управлінських систем і так далі. Тобто, якщо є план управління даними, це створює таке інституційне середовище і задає таку рамку, коли ці дані не перетворюються на капітал. Вони можуть працювати і приносити якийсь додатковий зиск, додаткову вартість.

Якщо цього не робити, залишається ресурс, якимось там воно організовано і все. Але для того, щоб воно працювало і приносило користь, додаткову вартість, додану вартість, потрібно забезпечити всі передумови. План управління дослідницькими даними, ось він для цього і існує.

Звісно, такі ефекти не спостерігаються водночас і залежать від кількісних змін взагалі в розповсюдженні практики управління дослідницькими даними серед наукових установ. І тоді це загалом дасть спільний ефект, загальний ефект, коли дані перетворюються на додану вартість. Забула ще сказати користь для дослідника, який стає видимим, результати якого достовірні, вони цитовані, їх можуть використовувати, і це дозволяє розвиватись не тільки практиці, бізнесу, науковій організації, але й самому досліднику.

Історичні витоки, звідки взялися ці плани управління даними? З того, що я знайшла, це були розмови на рівні тих, хто працює з даними, я маю на увазі, збирає, систематизує, зберігає, тобто бібліотекарі. Масштаби отриманої інформації настільки великі, що яким чином ними управляти, зберігати, класифікувати, давати доступ і так далі, це питання було для бібліотекарів.

І у 2000 році це дуже активне обговорення, масштабне обговорення цього питання. І в 2006 році були перші рекомендації з боку ряду наукової бібліотеки, центрів досліджень, які якраз зіштовхнулися з оцими великими обсягами наукових даних. І в 2011 році Національний науковий фонд у Штатах розробив дата-менеджмент-план DMP.

І він це зробив, це обов'язково, вірніше не розробив, а зробив його обов'язковою частиною, обов'язковою вимогою всіх заявок, які подаються на фінансування. І ось ця практика розробки планів управління даними, дата-менеджмент-план, набула свого розповширення і зараз фактично навіть є вимогою майже, здається, в усіх європейських конкурсах, ну, дуже багатьох. Навіщо потрібні?

План управління даних – це інструкція, як передоверити файли на комп'ютері на результат, який побачать всі, які розуміють, будуть цитувати і, можливо, також використовувати. Дата-менеджмент-план або план управління даними – це передумова для того, щоб якісно управлялись дані, їх створення або збирання, організація зберігання, доступ, повторне використання і так далі. Зниження ризику втрати даних – дуже важлива позиція, яка забезпечується завдяки тому, що продумані інструменти, яким чином зберігати їх і як робити резервні копії, як часто не зберігати і так далі.

Також в планах управління даними передбачена позиція, яка дозволяє імплементувати принципи FAIR. Крім того, план управління даними забезпечує прозорість і відтворюваність досліджень, підтримуючи принципи відкритої науки і підвищення довіри суспільства, підвищення цінності для наукової спільноти і для бізнесу, для управлінських практик. Також використання таких планів сприяють прогресу в різних галузях знань і зменшують, запобігають дублювання різних суспільств, різних науковців, різних команд.

Ну і план управління даними – це вимога грантів і журналів дуже часто, вже навіть журнали вимагають плани управління даними. Хто розробляє плани управління даними? Передусім

організації, наукові установи або університети, вони встановлюють внутрішні стандарти для всіх наукових дослідницьких даних.

Яким чином? Створюючи або розробляючи положення про план управління даними. Ось вони якраз і задають оцю рамку щодо того, як управляти даними в межах установи.

І ці положення формують певну культуру роботи з даними всередині установи. В рамках плану управління даними установи визначається, хто в цій установі повинен відповідати і хто повинен складати вже безпосередній план. Це можуть бути керівники НДР, це можуть бути команди проєктні або групи дослідників.

Це можуть бути навіть індивідуальні дослідники, які формують план управління даними в рамках тих проєктів, які вони виконують. Звичайно, ось ці особи, на кого покладено обов'язок о тим документам, який ухвалює установа, віднесе відповідальність за формування дата менеджмент план для проєкту. Якщо мова йде про просто грантовий проєкт якийсь, де немає в організації, немає ухваленого документу, ухваленого положення про план управління даними, то в таких умовах все рівно керівник проєкту або менеджер проєкту або той, хто очолює консорціум, той і забезпечує і несе відповідальність за підготовку такого плану управління даними.

Так, які інструменти? Звісно, для того, щоб скласти план управління даними, то можна звернутися до цифрових інструментів, які існують для полегшення цієї роботи. Найбільш відомі є DMP-Online і DMP-Tool і Argos.

Argos – це інструмент європейський, який є в OpenAIR, один із сервісів, і там можна зареєструватися і формувати план ще й на спільних засадах, в колаборації з іншими членами команди чи консорціуму для того, щоб визначити всі параметри, які мають бути в управлінні даними. Ось OpenAIR, якщо ви відкриваєте сервіси OpenAIR, ось цей Argos Managed Data, він є, він дозволяє укласти дата менеджменту плану. Значить, якогось загального стандарту, що включається в цей дата менеджмент план, його немає.

Існують рекомендації різного роду, існує досвід інших якихось установ, які вже розробили. І те, що тут пропонується ось на слайді, це лише загальна така рамка, загальна рекомендація, що має включатися в план управління дослідницькими науковими даними. Це загальні відомості про ці дані, тема, назва проєкту, хто керівник, хто автори, якась анотація проєкту, за рахунок чого він виконується, які строки виконуються.

Також те, яким чином будуть збиратись дані, які методи будуть використовуватись, які типи даних будуть, які формати файлів, який обсяг даних. Також в плані управління даними визначається етика використання, отримання і використання даних. Наразі дуже часто стоїть вимога етичного дозволу на використання даних, і журнали навіть можуть вимагати висновку етичної комісії, якщо мова йде про використання якихось персональних чи чутливих даних.

Чи буде надаватись, до речі, інформація учасникам досліджень, якщо збираються персональні дані. Також, окрім того, як вони будуть отримуватися, ця інформація, ці дані, хто буде мати доступ до файлів, наприклад, в короткому періоді, в довгостроковому періоді, на яких правах? На правах перегляду чи на правах редагування, опрацювання цих файлів і так далі.

Тобто, все це стосується етики отримання та використання даних в проєкті. Також, для належного управління даними треба вирішувати питання інтелектуальної власності. Кому належать дані, хто буде використовувати і так далі.

Потім, збереження даних і їх безпека. Які правила збереження даних, які правила найменування файлів, папок, як відбуваються резервне копіювання, які заходи захисту мають бути. Також, поширення і обмін даними.

Де буде розміщуватись, в якому репозиторії, які ідентифікатори будуть присвоєні, в умовах якого доступу, закритого, відкритого, якогось обмеженого. Період ембарго. Ембарго – це той період, коли дані недоступні.

Наприклад, якщо є припринт, то вони можуть бути недоступні, поки не будуть опубліковані. Або ще з якихось інших причин, на якийсь період часу, закритий доступ до даних, а потім він буде відкритий. Але до ООПС вже наявний.

На цьому слайді ви бачите рекомендації Міністерства освіти і науки України. Це рекомендація, це не є обов'язковим. В рекомендаціях Міністерства науки і освіти задані позиції, які можуть бути використані в такому плані.

Я не буду зачитувати, тому що більшість позицій були проголошені. Оскільки вам презентація буде доступна, ну або ви можете відкрити самостійно ці рекомендації Міністерства освіти, прочитати, ознайомитися з ними більш детально. І знову ж таки, рекомендації Міністерства освіти, яким чином організувати управління дослідницькими даними в закладах вищої освіти та наукових установах.

Також з ними можна з документом ознайомитись, подивитись. Тут викладені основні позиції, включені в ці рекомендації. Як перетворити дані на актив, на актив, який приносить додану вартість?

Значить, наукова установа, вона визначає ролі і відповідальність. Вона приймає положення про план управління даними і задає необхідні вимоги до публікації, до наборів даних, які забезпечують їхню видимість. Якщо немає мета даних і цифрового ідентифікатору, то дані не будуть видимими однозначно і це має бути забезпечено обов'язково.

Значить, керівник проекту або очільник НДР або менеджер проекту, він відповідає за розробку плану управління даними. Визначає всі параметри, які ми говорили раніше. Також є дата стюарти, така нова посада фактично, можливо не посада поки що, але ті фахівці, які фактично є редакторами видимості даних, можна їх так назвати, це ті фахівці, які не тільки розміщують дані, але і допомагають і забезпечують можливість їх видимості і FAIR даних, щоб були дотримані принципи під час організації цих даних і розміщення їх в репозиторії.

Звісно, можуть залучатися і ті команди установи, і юрист в науковій установі, якщо виникають відповідні питання, пов'язані з організацією даних і з етичним їх використанням, чи наданням права доступу, чи з інтелектуальною власністю. Вони можуть залучатись, якщо такі особи є. Значить, який порядок приблизно роботи?

Автор готує опис даних за тим мінімумом, який може забезпечити їх видимість і довіру додати до даних. Це все, що стосується метаданих, це обов'язкове використання...

[Мовець 5]

Ні, я слухаю зараз семінар по дату, от це я хотів спросити, не слухаю.

[Мовець 1]

Микрофони можна?

[Мовець 5]

Відкрита наука.

[Мовець 1]

А можна попросити, так. Перепрошую. Так, дивіться, автор готує метадані, обов'язкове інформацію про себе, і інформацію про всіх авторів з їх орсідами, також оркід, також ідентифікатор для установи, ліцензія, режим доступу, в який буде регламентовано все це, також описовий файл з описуванням того, що це за набір даних.

І також обов'язково проводиться перевірка на те, чи відповідає опис даних, їх підготовка, їх формат феопринципам. Власне, я на цьому зупинюсь, я ще просто дивлюсь на час, мені б не хотілося перебрати, оскільки вам будуть детальніше розповідати далі про і репозиторії, і про те, яким чином готувати метадані, для того, щоб можна було їм оперувати таким чином, що вони йшли на користь і дозволяли розвиватися і вашій науковій установі, і авторами.

[Мовець 4]

Лідія, дуже дякуємо. Чиє питання у аудиторії до Лідії?

[Мовець 3]

У мене чисто технічне питання таке, якщо дані вже перевірені, розміщені, і виникає потреба для їх оновлення, чи це технічно важко?

[Мовець 1]

Сергій Якович включається, тому що технічні питання знає Сергій Якович.

[Мовець 2]

Давайте я відповім. Я буду казати тільки про тот репозиторій, який мені відповідає. Це програма забезпечення Dataverse, і там є можливість зробити іншу версію ваших даних.

Тобто там є така версійність, тобто дої не змінюються, самі набори даних не змінюються, ви додаєте додаткову інформацію. Тобто версія початкова і версія один, тобто це можливо зробити, і вам про це завтра розкаже Наталя Денисюк.

[Мовець 3]

Добре, дякую дуже.

[Мовець 1]

Можна простежувати історію і бачити, що ви і коли робили, і вдосконалювали.

[Мовець 4]

Ще питання є? Якщо нема, то я передаю слово Сергію Яковичу.