

УДК 631.47

© 2024

## АДАПТУВАННЯ ТАБЛИЦЬ ІЗ ҐРУНТОВИМИ ДАНИМИ ДО ПОДАЛЬШОГО ВИКОРИСТАННЯ В БАЗАХ ДАНИХ

*В.В. Лебедь*

*кандидат біологічних наук*

*Національний науковий центр «Інститут ґрунтознавства  
та агрохімії імені О.Н. Соколовського»*

*вул. Чайковська, 4, м. Харків, 61024, Україна*

*e-mail: swdiscover@gmail.com*

*ORCID: 0000-0002-1429-4121*

*Надійшла 04.04.2024*

**Мета.** Адаптувати існуючі таблиці з даними про ґрунти за національними та міжнародними стандартами для їх використання у базі даних «Ґрунти України». **Методи.** Застосовували методи: аналізу — для виявлення конструктивних та функціональних характеристик ґрунтових баз даних, порівняння — для визначення відмінностей у принципах побудови структури баз даних, систематизації — для узагальнення підходів до налаштування таблиць із ґрунтовими даними. **Результати.** Розглянуто конструктивні та функціональні особливості глобальних баз даних SOTER, GSM, WoSIS і по кожній із них визначено ключові положення, які можна запозичити для розроблення бази даних «Ґрунти України». Проведено перевірку структури таблиць ґрунтових даних та присвоєно формати даних полям таблиць відповідно до принципів реляційної моделі даних. Кожній таблиці бази даних присвоєні ключові поля: «Просторові атрибути» — *id* точки відбору проб, «Морфологічні атрибути» — *id* точки відбору проб та *id* генетичного горизонту, «Фізичні атрибути», «Фізико-хімічні і хімічні атрибути» — *id* місця відбору проб, *id* генетичного горизонту та *id* шару ґрунту, з якого відбирали пробу. Кодування атрибутів проводили у три способи: з використанням скороченої назви латиницею, набору буквених символів (аббревіатури) і числових кодів. **Висновки.** Проведений аналіз міжнародних баз даних дозволив виділити ключові підходи до створення, налаштування та підтримки бази даних «Ґрунти України» у майбутньому. Процес адаптування включає перевірку таблиць на відповідність принципам реляційної моделі даних, налаштування формату даних кожного поля таблиці, визначення ключових полів для забезпечення зв'язків між таблицями у базі даних та кодування назв атрибутів. Так досягається відповідність таблиць з даними про ґрунти стандартам побудови баз даних, а також проходить гармонізація зі світовими центрами ґрунтової інформації для глобального обміну даними.

**Ключові слова:** ґрунтові дані, таблиця, база даних, реляційна модель даних, поле таблиці, запис таблиці, ключові атрибути.

DOI: <https://doi.org/10.31073/agrovisnyk202411-09>

Досить часто таблиці, що містять ґрунтові дані, мають довільну форму подання. Проте таблиці, не структуровані за чіткими критеріями, швидше носять довідковий характер та дають лише часткове уявлення про набір даних. Відсутність єдиного шаблону побудови таблиць унеможливує їх подальше використання в електронних базах даних (БД) з чіткими правилами внесення інформації.

Автори праць [1, 2] виконали роботу зі стандартизації та гармонізації ґрунтових атрибутів (параметрів властивостей) і розробили відповідні таблиці. Однак форма подання таблиць для зберігання ґрунтових даних у попередніх розробках не відповідає стандартам побудови баз даних. Тому постала потреба в адаптації структури таблиць до прийняттого вигляду з використанням міжнародного досвіду.

Водночас є необхідність виконати кодування назв ґрунтових атрибутів, використовуючи оригінальні коди глобальних баз даних WoSIS, SOTER та GSM. А отже, постає потреба в аналізі конструктивних і функціональних особливостей даних інформаційних систем та встановленні можливості їх застосування у національній базі даних «Ґрунти України». Інакше кажучи, процес адаптації таблиць ґрунтових даних включає кілька етапів їх інтеграції до баз даних та ґрунтово-інформаційних систем.

**Мета досліджень** — адаптувати існуючі таблиці з даними про ґрунти до національних та міжнародних стандартів задля можливості їх використання у базі даних «Ґрунти України».

**Матеріали і методи досліджень.** Адаптування існуючих таблиць ґрунтових атрибутів до використання у сучасних базах даних потребує врахування основних конструктивних та функціональних особливостей глобальних баз даних. Під час роботи над створенням національної бази даних та її гармонізацією з міжнародними стандартами, за основу було взято глобальні інформаційні системи та бази даних SOTER — Soil and Terrain database (База даних ґрунтів і земель), WoSIS — World Soil Information Service (Світова ґрунтова інформаційна служба) та GSM — Global Soil

Map (Карта ґрунтів світу) [3, 4]. Було проведено додаткове упорядкування та змінено структуру таблиць, що мали безпосередньо використовуватись у базі ґрунтових даних. Для цього скористались одним зі способів автоматичної переорієнтації вже існуючих таблиць трансферних форм у програмі Excel — функцією «Транспозиція».

Дані, що розміщуються у таблицях, мають відповідати правилам реляційних моделей. Термін «реляційний» походить від англійського слова relation (відношення). Саме це поняття, а також принцип використання реляційних БД були запропоновані Едгаром Коддом у 1970 р. [5–7]. У загальному розумінні, в реляційній моделі все подається множиною таблиць, рядки та стовпчики яких є елементами даних. Програмні засоби реляційної системи керування базами даних (СКБД) забезпечують керування зв'язками між таблицями. Наразі реляційна модель є найпопулярнішою з усіх, що використовуються для розроблення баз даних. Змогу працювати з реляційними базами даних дають такі програми, як MS Access [8] та PostgreSQL [9].

Для позначення структурних елементів таблиць у базах даних використовується спеціалізована термінологія. Загалом існують терміни для користувачів баз даних (на рівні розуміння механізмів роботи) та для програмістів (на рівні внутрішньої фізичної структури). Так, вертикальні структури таблиць користувачі називають «стовпчики», а програмісти — «поля»; горизонтальні елементи таблиць — відповідно «рядки» та «записи».

Після упорядкування структури таблиць слід здійснити налаштування для відповідності реляційній моделі даних, яка передбачає дотримання п'яти умов.

- Усі записи в певному полі (стовпчику) мають бути одного типу. Це означає, що всі записи одного поля мають бути, наприклад, числами, датами або текстовими рядками. В одному полі не допускається змішування даних різних типів — чисел, дат, текстів тощо.

- Кожне поле повинне мати унікальне ім'я, в межах однієї таблиці (тобто

стовпчики можуть мати однакові імена в різних таблицях, але не в межах однієї).

- Таблиця не може містити однакових записів (рядків). Якщо два записи таблиці являють собою повністю ідентичну інформацію у кожному з полів, то така таблиця не є коректною.

- Порядок слідування полів не має значення і не повинен впливати на інформаційний вміст таблиці; будь-яка перестановка стовпчиків не повинна змінювати інформацію.

- Порядок записів неважливий і не повинен впливати на інформаційний вміст таблиці.

У кожній таблиці слід визначити ключі, або ключові поля — один чи групу атрибутів, що однозначно визначають решту атрибутів у записі таблиці. Слід звернути увагу на те, що ключі бувають прості, з одним атрибутом, та композитні, або складені — містять більше одного атрибута. На додачу до них у базах даних традиційно використовують Первинний ключ (*Primarykey*) та Зовнішній ключ (*Foreignkey*). Первинним ключем є атрибут, який пов'язаний з іншими атрибутами в одній таблиці та однозначно визначає їх, відповідно зовнішній ключ — це атрибут, який унікально ідентифікує рядок в іншій таблиці, створюючи таким чином зв'язок між ними.

Небажано як ключ використовувати довгі текстові записи, якісні та кількісні показники ґрунтів тощо. Вони можуть повторюватися неодноразово, що призведе до дублювання даних, а це неприпустимо у ключових полях таблиць.

З метою гармонізації з міжнародними базами даних та забезпечення відповідності стандартам назви атрибутів уніфікують за певними правилами та здійснюють їх кодування. Насамперед, назви атрибутів (заголовки полів) таблиць мають бути короткими та водночас зрозумілими для користувачів. Слід зазначити, що для формування назв атрибутів визначено певні правила і їх треба дотримуватися. До того ж усі назви атрибутів мають бути написані виключно латиницею, малими літерами, наприклад: *horizon*. Якщо назва складається із кількох слів, то між

ними ставиться нижнє підкреслення, тобто пробіли не допускаються, наприклад: *soil\_profile*.

Найпростішим для внесення у заголовки полів атрибутів є код у вигляді числа. Нумерація завжди повинна бути послідовною. Однак відлік може починатися як від одиниці (1, 2, 3...), так і від будь-якого додатного числа. Оптимальною буде нумерація з чотирьох чисел у межах певного діапазону для кожної таблиці бази даних, наприклад для першої таблиці — 1001, 1002, 1003..., для другої — 2001, 2002, 2003... і т.д. Однак для розуміння того, який саме ґрунтовий параметр визначає та чи інша цифра, необхідно створити окрему таблицю-довідник — за її допомогою оператор набору даних матиме змогу розшифрувати числові значення.

Для позначення атрибутів у вигляді абрєвіатури зазвичай використовують перші літери кожного слова в назві атрибута, інколи кілька літер. Наприклад, *Organic carbon* (вміст органічного вуглецю) позначається як *ORGC*. Для стандартизації таких позначень у базах даних прийнято використовувати коди з чотирьох літер. Таке позначення атрибутів сприймається краще, ніж числові коди, адже, знаючи назву параметра англійською мовою, досить легко визначити, що означає кожна абрєвіатура. Проте для більшості користувачів знадобиться довідник з кодами атрибутів та їхніми повними назвами.

Найкраще сприймаються скорочені назви атрибутів латиницею, користувачі бази даних у цьому разі не потребують додаткової інформації для розшифровки специфічних кодів. Однак така форма подання атрибутів застосовується нечасто через досить велику кількість символів, що призводить до надлишковості бази даних.

**Результати досліджень.** На першому етапі дослідження було проаналізовано конструктивні та функціональні особливості глобальних баз даних — *SOTER* [10–12], *GSM* [13] та *WoSIS* [14, 15]. Щодо кожної з них визначено ключові положення, які можна запозичити для розроблення національної ґрунтової бази даних.

Серед позитивних конструктивних та функціональних характеристик БД SOTER (рис. 1) відмічено такі:

- атрибутивна частина БД включає основні властивості ґрунтів та рельєфу, які є важливими для агроекологічного моніторингу;
- дані із цієї БД можна використовувати для проведення широкого кола досліджень — оцінювання прояву різних видів деградації у ґрунтовому покриві та вразливості ґрунтів до забруднення, для моделювання запасів ґрунтового органічного вуглецю тощо.

Національна база даних ґрунтів України має враховувати широкий спектр ґрунтових атрибутів, які використовуються у БД SOTER.

Під час розбудови бази даних на національному рівні, звісно, були запозичені й окремі ключові принципи функціонування БД GSM. Йдеться, зокрема, про діджиталізацію, зберігання в повному обсязі метаданих, інформації про методи й одиниці вимірювання.

Що стосується БД WoSIS (рис. 2), то у ній для використання у національній ґрунтовій базі даних було запозичено такі ключові особливості:

- вимоги стосовно стандартизування вхідної інформації (стандартна назва показника, стандартний опис методу аналізування, стандартні одиниці вимірювання тощо);
- дані провайдерів повністю забезпечуються усім необхідним набором метаданих, що гарантує однозначність щодо походження, типу й якості даних, ідентифікації авторів інформації;
- запобігання можливості дублювання профілів у разі об'єднання кількох баз і наборів даних;
- стандартизовані атрибути (показники властивостей ґрунту).

У процесі розробки та створення національної бази даних «ґрунти України» використано найкращі характеристики та функції кожної із проаналізованих міжнародних баз даних. Водночас БД WoSIS визначено як еталонну для організації бази даних ґрунтових профілів (точкових даних параметрів властивостей ґрунту).

Далі створені раніше таблиці ґрунтових атрибутів [1] було приведено до вигляду, придатного для використання у сучасних базах даних. Це стосується насамперед структури таблиць, порядку розташування стовпчиків та рядків, формату даних.

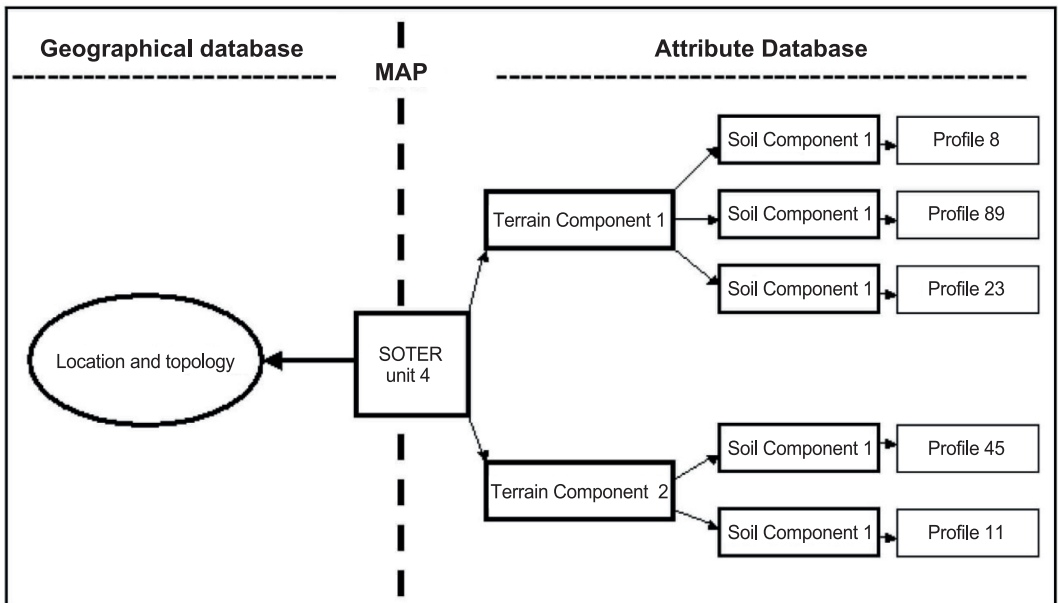
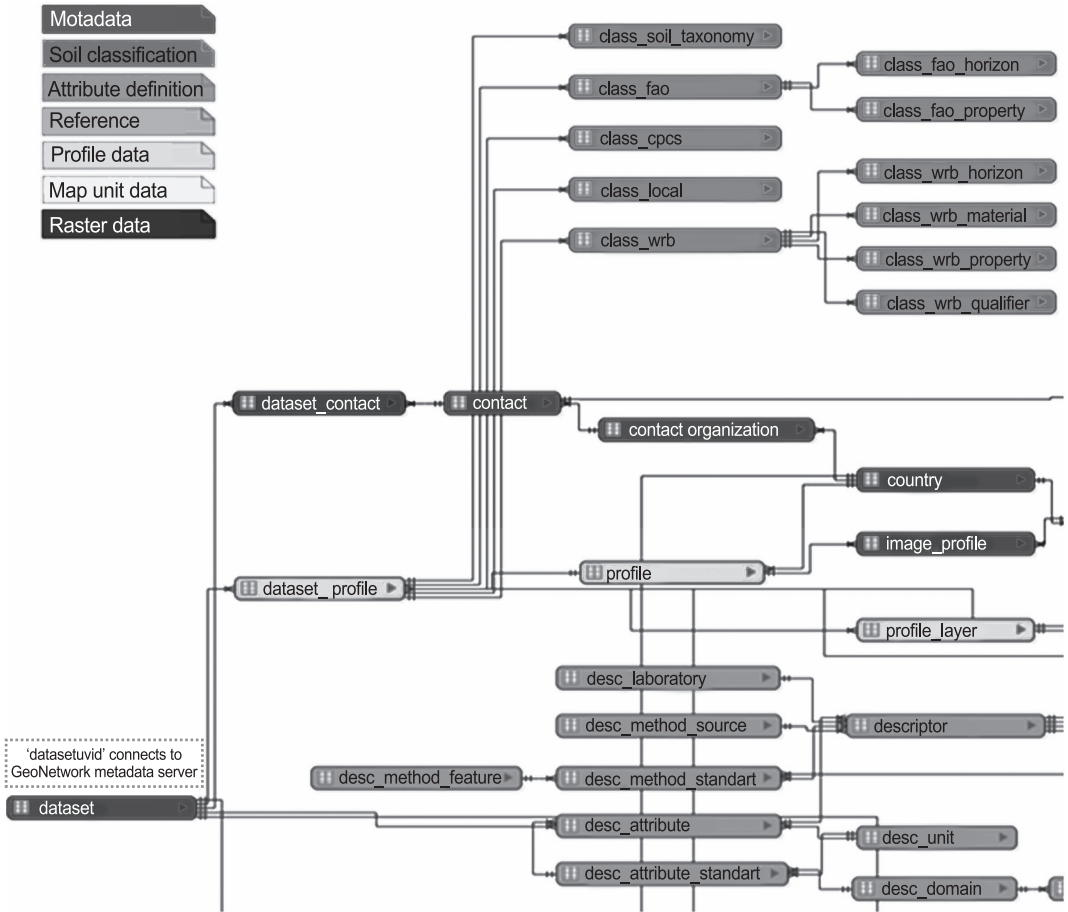


Рис. 1. Схематичне подання картографічних одиниць та їх атрибутивних даних у БД SOTER [10]



**Рис. 2.** Фрагмент структури та схеми зв'язків між таблицями у моделі бази даних WoSIS [14]

Зокрема, таблиці було налаштовано шляхом їх перевірки на відповідність поняттю «відношення» за п'ятьма умовами реляційної моделі даних. Відповідно до цих умов проведено присвоєння форматів даних кожному полю таблиць. Для атрибутів, параметри яких відображаються дробовими числами, встановлено кількість знаків після коми, що дає змогу стандартизувати всі записи у відповідних полях таблиці.

Також налаштовано перевірку на предмет дублювання даних у таблицях з використанням функції «Видалити дублікати» у програмі Excel — вона виявляє дублікати завдяки розширеним параметрам сортування даних. Що стосується порядку

полів та записів у таблицях, то їх розміщення регулюється виключно через сортування даних та не впливає на змістове наповнення таблиць.

Кожній таблиці бази даних були присвоєні ключові атрибути (рис. 3). Якщо їх не визначити, то це буде просто набір таблиць з даними і жодна система керування базами даних (СКБД) не зможе з ними працювати.

Слід зауважити, що кожне ключове поле має заголовок id, до якого додається просте число або набір із кількох літер. Так, для бази даних «Ґрунти України» у таблиці «Просторові атрибути» Первинним ключем обрано id точки відбору проб — розрізу (профілю),

| Просторові атрибути        |                     | Морфологічні атрибути      |                     |                         |                              |                       |                      |
|----------------------------|---------------------|----------------------------|---------------------|-------------------------|------------------------------|-----------------------|----------------------|
| <i>id атрибута</i>         | id1008              | <i>id атрибута</i>         | id1008              | id2001                  | id2002                       | id2003                | id2004               |
| <i>Назва атрибута</i>      | № профілю (розрізу) | <i>Назва атрибута</i>      | № профілю (розрізу) | № генетичного горизонту | Індекс генетичного горизонту | Верхня межа горизонту | Нижня межа горизонту |
| <i>Позначення атрибута</i> | number_soil_sample  | <i>Позначення атрибута</i> | number_soil_sample  | genetic_horizon_number  | genetic_horizon_index        | horizon_upper_limit   | horizon_lower_limit  |

| Фізичні, фізико-хімічні, хімічні атрибути |                     |                         |                              |                       |                      |  |   |
|---|---------------------|-------------------------|------------------------------|-----------------------|----------------------|--|---|
| <i>id атрибута</i>                        | id1008              | id2001                  | id2002                       | id2003                | id2004               | id3001                                       | id3002                                      |
| <i>Назва атрибута</i>                     | № профілю (розрізу) | № генетичного горизонту | Індекс генетичного горизонту | Верхня межа горизонту | Нижня межа горизонту | Верхня межа шару ґрунту, з якого взято пробу | Нижня межа шару ґрунту, з якого взято пробу |
| <i>Позначення атрибута</i>                |                     | genetic_horizon_number  | genetic_horizon_index        | horizon_upper_limit   | horizon_lower_limit  | layer_upper_limit                            | layer_lower_limit                           |

Рис. 3. Визначені ключові атрибути у таблицях бази даних

#### Система кодування атрибутів (фрагмент)

| Назва українською                      | Скорочена назва латиницею | Абревіатура | Числовий код |
|--|---------------------------|-------------|--------------|
| Тип відбору проби                      | sample_type               | SAMT        | 1009         |
| Назва ґрунту                           | soil_name                 | SOIN        | 1046         |
| Потужність гумусованої частини профілю | humus_profile_thickness   | HPRT        | 1053         |
| Тип структури                          | structure_type            | STRT        | 2008         |
| Наявність грубих фрагментів (> 2 мм)   | coarse_fragments          | COFR        | 2016         |

прикопки, свердловини. З ним пов'язана решта атрибутів для однозначної ідентифікації параметрів властивостей конкретного ґрунту в конкретному місці відбору. Відповідно в усіх інших таблицях id точки відбору проб буде Зовнішнім ключем.

У таблиці «Морфологічні атрибути» до id точки відбору проб додано ще один ключовий атрибут — id генетичного горизонту. В нашому випадку це вже буде композитний ключ, оскільки у ґрунтового профілю завжди є кілька горизонтів, що слід неодмінно враховувати.

У таблицях «Фізичні атрибути» та «Фізико-хімічні і хімічні атрибути» до ключів id місця відбору проб та генетичного

горизонту додається ключ id шару ґрунту, з якого відбирали пробу. Це пов'язано з тим, що у межах генетичного горизонту може бути кілька шарів відбору проб.

Отже, упорядкування структури таблиць для збору даних відповідно до стандартів реляційної моделі та налаштування форматів даних і ключових полів дає змогу використовувати ці таблиці у базі даних «Ґрунти України».

Після налаштування структури таблиць було проведено кодування їхніх атрибутів за трьома способами: з використанням скорочених назв латиницею, наборів буквених символів (аббревіатур) та числових кодів (таблиця).

## Висновки

Часто таблиці з ґрунтовими даними потребують адаптування до використання в базах даних та інформаційних системах. Проведений аналіз міжнародних баз даних дав змогу виділити ключові підходи до створення, налаштування та підтримки бази даних «Ґрунти України» у майбутньому. Процес адаптування передбачає перевірку таблиць на відповідність принципам реляційної

моделі даних, налаштування формату даних кожного поля таблиці, визначення ключових полів для забезпечення зв'язків між таблицями у базі даних та кодування назв атрибутів. Так досягається відповідність таблиць з даними про ґрунти стандартам побудови баз даних, що важливо для гармонізації зі світовими центрами ґрунтової інформації та глобального обміну даними.

### Lebed V.

NSC «O.N. Sokolovskyi Institute for Soil Science and Agrochemistry Research», 4 Chaikovska Str., Kharkiv, 61024, Ukraine; e-mail: swdiscover@gmail.com; ORCID: 0000-0002-1429-4121

### Adaptation of soil data tables for further use in databases

**Goal.** To adapt existing soil data tables according to national and international standards for their use in the «Soils of Ukraine» database.

**Methods.** The following methods were used: analysis — to identify the structural and functional characteristics of soil databases; comparison — to determine differences in the principles of building the database structure; systematization — to generalize approaches to setting up soil data tables. **Results.** The structural and functional features of the global databases SOTER, GSM, and WoSIS were considered and key provisions were identified for each of them that can be borrowed for the development of the «Soils of Ukraine» database. The structure of the soil data tables was checked and data formats were assigned to the table fields under the principles of the relational data model. Each database table was assigned key fields: «Spatial

attributes» — sampling point id, «Morphological attributes» — sampling point id and genetic horizon id, «Physical attributes», «Physicochemical and chemical attributes» — sampling point id, genetic horizon id, and soil layer id from which the sample was taken. Attribute coding was carried out in three ways: using an abbreviated name in Latin, a set of letter symbols (abbreviations), and numerical codes. **Conclusions.** The analysis of international databases allowed us to identify key approaches to the creation, configuration, and maintenance of the «Soils of Ukraine» database in the future. The adaptation process included checking tables for compliance with the principles of the relational data model, setting the data format of each table field, defining key fields to provide relationships between tables in the database, and encoding attribute names. In this way, the compliance of tables with soil data to database construction standards was achieved, as well as harmonization with world centers of soil information for global data exchange.

**Key words:** ground data, table, database, relational data model, table field, table record, key attributes.

DOI: <https://doi.org/10.31073/agrovisnyk202411-09>

## Бібліографія

1. Лактіонова Т.М., Соловей В.Б., Лебедь В.В. Систематизований перелік гармонізованих атрибутів бази даних «Ґрунти України». Інструкція для формування бази даних; за наук. ред. С.А. Балюка, Т.М. Лактіонової та М.М. Мірошниченка. Харків: ФОП Бровін О.В., 2020. 48 с.

2. Соловей В., Лебедь В., Лактіонова Т. Науково-методичні основи функціонування Українського ґрунтового інформаційного центру. *Bulletin of Agricultural Science*. 2022.

Т. 100, № 9. С. 26–33 doi: 10.31073/agrovisnyk202209-03

3. Лактіонова Т.М., Бігун О.М., Накісько С.Г., Уваренко К.Ю. Конструктивні та функціональні особливості провідних світових баз ґрунтових даних. Аналітичний огляд. *Агротехніка і ґрунтознавство*, 2020. № 89. С. 4–17 doi: 10.31073/acss89-01

4. Лебедь В.В. Досвід найвідоміших ґрунтових інформаційних систем світу. Аналітичний

огляд. *Агрохімія і ґрунтознавство*. 2023. Вип. 94. С. 54–61. doi: 10.31073/acss94-06

5. Пасічник В.В., Резніченко В.А. Організація баз даних та знань: підручник для ВНЗ. Київ: Видавнича група ВНУ, 2006. 384 с.

6. Зацерковний В.І., Бурачек В.Г., Железняк О.О., Терещенко А.О. Геоінформаційні системи і бази даних: монографія. Ніжин: НДУ ім. М. Гоголя, 2014. 492 с.

7. Вовкодав О.В., Ліп'яніна Х.В. Сучасні інформаційні технології: навч. посібн. Тернопіль, 2017. 500 с.

8. Шичула О.М. Система керування базами даних *MS Access*: навч. посібн. Київ: ІПДО, 2017. 177 с.

9. Павловський В.І., Петрашенко А.В., Победа Д.В. Бази даних та засоби управління. Практикум. [Електронний ресурс]: навч. посібн. Київ: КПІ ім. Ігоря Сікорського, 2021. 112 с.

10. *Batjes N.H.* A taxotransfer rule-based approach for filling gaps in measured soil data in primary SOTER databases (ver. 1.1). Wageningen :

ISRIC World Soil Information. 2005. 44 p.

11. *Van Engelen V.W.P., Dijkshoorn J.A.* (eds.). *Global and National Soils and Terrain Digital Databases (SOTER)*. Wageningen: ISRIC World Soil Information. 2013. 198 p.

12. *SOTER data model*. Electronic data. Wageningen: ISRIC World Soil Information, 2012–2018. URL: <https://www.isric.org/soter-data-model> (дата звернення 20.03.2024)

13. *GLOBALSOILMAP.NET*. Electronic data. Wageningen: ISRIC World Soil Information, 2008–2018. URL: <http://www.globalsoilmap.net> (дата звернення 20.03.2024)

14. *Ribeiro E., Batjes N.H., Oostrum van A.J.M.* World Soil Information Service (WoSIS) — Towards the standardization and harmonization of world soil data. Wageningen: ISRIC World Soil Information, 2018. 166 p.

15. *Batjes N.H., Ribeiro E., Oostrum A. van et al.* WoSIS: providing standardised soil profile data for the world. *Earth Syst. Sci. Data*. 2017. V. 9. P. 1–14.