

СВІТОВИЙ ДОСВІД ЗІ СТВОРЕННЯ ЕЛЕКТРОННИХ ГАЗЕТИРІВ ГЕОГРАФІЧНИХ НАЗВ

© Бєлова О., 2010

Рассмотрены электронные газеттеры географических названий четырёх стран: США и Канады, Великобритании, Венгрии, а также изображение газеттеров ЮАР и Австралии; изложены их параметры. Описана сущность международного стандарта ISO 19112 “Географическая информация – Пространственная привязка с помощью географических идентификаторов” касательно создания электронных газеттеров географических названий объектов. Предложено решение проблемы интеграции различных ведомственных данных в электронных газеттерах в соответствии с этим стандартом.

The geographical names electronic gazetteers of four countries: USA and Canada, Great Britain, Hungary, and the portrayal of Australian and SAR gazetteers, are considered; their parameters are shown. The main point of the regulations of International standard ISO 19112 “Geographic information – Spatial referencing by geographic identifiers” concerning geographical names electronic gazetteers creation, is described. The solving of the problem for different departmental data integration in electronic gazetteers is proposed.

Постановка проблеми. На сучасному етапі соціально-економічного розвитку України важливе значення має застосування геоінформаційних технологій для розв’язання низки задач у землевпорядній, картографо-геодезичній, картографічній, екологічній, навігаційній галузі тощо. Сучасні геоінформаційні ресурси мають багатогалузеве походження, тому їх створення є досить трудомістким, оскільки використовується велика кількість об’єктів, що необхідно закоординувати та розмістити в єдиному геоінформаційному просторі. Але при створенні такого геоінформаційного простору виникає низка проблем, пов’язана з тим, що геоінформаційні ресурси створюють за відомчим принципом, застосовуючи різноманітні дані: не всі вони є закоординованими, деякі з них реєструються з використанням різних картографічних масштабів, проекціях у різних програмно-технологічних комплексах, у різних форматах обміну даними тощо. Це, своєю чергою, призводить до неузгодженості інформації, дублювання трудомістких і дорогих робіт зі збору даних, що тільки збільшує витрати і знижує якість інформації для прийняття рішень в управлінні територіями. Дуже складно, іноді навіть неможливо інтегрувати геопросторові дані, одержані з різних джерел та різних баз даних і ГІС [9].

Вирішує цю проблему уніфікація геопросторових даних в єдину систему для забезпечення виконання геоінформаційного аналізу. Це здійснюється за допомогою електронних газетирів, де дані з різних джерел зводяться в єдину систему. Такий підхід дасть змогу зменшити витрати на підтримку в базових наборах геопросторових даних атрибутів обов’язкового постійного зберігання, а також забезпечить інтеграцію даних з численних інформаційних ресурсів і, отже, створення різних тематичних геопросторових даних на усій сукупності відомостей, доступних з різних державних реєстрів та баз даних. Досягається найбільша цілісність, достовірність, інформативність і тематична різноманітність геопросторових даних за рахунок їх сумісності з першоджерелами галузевих інформаційних ресурсів [9].

Аналіз останніх досліджень і публікацій. Дослідження багатьох авторів з питання інтеграції даних узагальнено у міжнародному стандарті ISO 19112 “Географічна інформація – Просторова прив’язка з допомогою географічних ідентифікаторів”.

Постановка завдання. Метою публікації є дослідження проблеми вибору оптимального електронного газетира географічних назв, який би відповідав міжнародному стандарту ISO 19112 “Географічна інформація – Просторова прив’язка з допомогою географічних ідентифікаторів”.

Виклад основного матеріалу. Газетир – це електронний географічний довідник інформації про місцезнаходження та їх назви з геокодованими координатами. Він використовується разом із картою або повним атласом. Зазвичай містить інформацію стосовно географічного складу країни, регіону або континенту, так само, як і відомчі дані та інформацію про фізичні об’єкти [20].

Огляд електронних газетирів топонімів. Електронний газетир США та Канади

Ця база даних поєднує два продукти: ZIPList5 Geocode та ZIPList Canada Geocode, в єдину базу даних ZIP-кодів, яка охоплює продукт обох країн – США та Канади, в єдиний продукт. Файл цієї бази даних вносить кожний активний ZIP-код і поштовий код для США та Канади, зокрема широту, довготу, назву міста, назву країни/провінції, код регіону, назву штату (для США), часову зону та світлий час доби. Ці дані можна використовувати у популярних комерційних програмах баз даних, таких як MS Access, dBase, Paradox, FoxPro тощо (табл. 1) [14].

Таблиця 1

Параметри запису газетира США та Канади [14]

Поле	Тип	Довжина
Назва міста	Змінна довжина	30 символів
Код штату/провінції	Фіксована довжина	2
ZIP/поштовий код	Фіксована довжина	6 символів + пробіли
Код регіону	Фіксована довжина	3 цифрових знаків
Назва штату (для США)	Змінна довжина	25 символів
Часова зона	Змінна довжина	5 символів (EST, CST, MST, etc.)
Світлий час доби	Фіксована довжина	1 (“Y” or “N”)
Країна	Фіксована довжина	1 (“C”, or “U”)
Широта	Фіксована довжина	7 символів: nn.nnnn
Довгота	Змінна довжина	7 або 8 символів: nnn.nnnn
Тип ZIP-коду (для США)	Фіксована довжина	1 символ (P, U, M, або пробіл)
Попередня територія вибірки штату (FIPS) (для США)	Фіксована довжина	5 цифрових знаків

Поле “Назва міста” містить одну назву міста для певного ZIP/поштового коду. Ця база даних містить лише один запис для кожного активного ZIP/поштового коду та НЕ вносить будь-якої неофіційної назви міста. Якщо потрібно отримати неофіційну назву міста, потрібно звернутися до ZIPList Can-USA Max, що містить всю інформацію у цій базі даних, а також неофіційні назви міст та інші потрібні дані.

Деякі назви провінцій Квебека містять спеціальні діакритичні знаки.

Поле “Штат/Провінція” містить стандартне двосимвольне скорочення штату або провінції.

Поле “ZIP/поштовий код” містить лише дійсні ZIP/поштові коди, визначені Поштовою службою США та Поштовою корпорацією Канади. ZIP/поштові коди, які вийшли з ужитку та які “застаріли”, не входять до цієї бази даних.

Код Регіону – це “найбільш використаний” код регіону для одного поштового коду. Якщо множину кодів регіону застосовують до того одного поштового коду, то до списку вноситься лише перший варіант коду регіону, повторні значення кодів регіону до списку не вносяться.

Поле “Назва штату” визначає найвживанішу назву штату для п’ятирозрядного ZIP-коду. Це поле застосовується лише до записів ZIP-коду США. У всіх канадських записах це поле пусте.

Поле “Часова зона” завжди містить позначення стандартної часової зони для певного поштового коду. Якщо цей код входить до утворення, яке міститься у “Світлому Часі Доби” протягом літніх місяців, напроти поля “Світлий Час Доби” позначено “Y”. В іншому випадку – позначено “N”.

Поле “Країна” містить єдиний символ, що визначає, чи запис стосується до США (“U”), чи Канади (“C”).

Поля “Широта” і “Довгота” містять географічні координати в градусах з точністю до чотирьох знаків, надаючи загальний рівень точності до 100 футів. Для більших канадських міст вибрано широту й довготу, застосовуючи перші три символи поштового коду (відомі як “Попередня територія вибірки” “Forward Sortation Area або FSA”), що відповідає рівню п’яти-розрядного ZIP-коду в США.

Поле “Тип ZIP-коду” (для США) визначає спеціальні ZIP-коди, які використовують для поштових адрес, компаній або для військових цілей.

Код попередньої території вибірки штату (FIPS-код) (для США) – це п’ятирозрядне числове поле, що має унікальне числове визначення, встановлене для певного штату правлінням США. Це поле та Назва Штату визначають найвживаніші назви штатів для п’яти-розрядного ZIP-коду [14].

Проілюструємо картографічне зображення газетира географічних назв з прив’язкою до карти на прикладі результату пошуку північної частини м. Торонто серед населених пунктів Канади (рис. 1).



Рис. 1. Фрагмент картографічного зображення результату пошуку північної частини м. Торонто серед населених пунктів Канади [17]

Електронний газетир Великобританії та Ірландії

Поштові коди Великобританії створено з двох частин. Перша частина – це зовнішній код, що визначає офіс доставки пошти, друга частина – внутрішній код, який визначає пункт доставки, коли стоїть поряд із назвою або номером будинку. Приклад запису газетира Великобританії та Ірландії в базі даних надано в табл. 2 [16].

Таблиця 2

Приклад запису газетира Великобританії та Ірландії в базі даних [16]

Частина	Значення	Номер
SW	Територія поштового коду	124 у Великобританії
SW1A	Район поштового коду	Приблизно 20 районів на територію
SW1A 2	Сектор поштового коду	Приблизно 3,000 адрес на сектор
SW1A 2AA	Територія поштового коду	Приблизно 15 адрес на одиницю.

Цей газетир має три основні функції: надає повний індекс географічним назвам Великобританії; прив’язує кожну назву до її історичної країни; прив’язує кожну назву до набору адміністративних територій. Основні риси цього газетира:

- індекс географічної назви;

- історичні назви країн Великобританії;
- адміністративні території;
- потреба фіксованої прив'язаної системи координат.

Мета “Індексу географічної назви” – охопити кожний населений пункт (окрім окремих будинків та ферм), який має точну назву та особливості. Це села, селища міського типу, містечка та міста. Кожний район всередині території міста, що має точну назву та особливість, вноситься як окремий запис. Місця, назви яких мають більш ніж одне стандартне написання, включені під кожним своїм записом. Так само місця з валійськими та англійськими, або галійськими та англійськими назвами внесені до списку під обома назвами.

Історичні назви країн Великобританії – це географічні утворення, вік та походження яких змінюються. Більшість із тих, що стосуються Англії, залишилися від часів нормандського захоплення.

Газетир надає деталізовані головні типи адміністративних територій, що пов'язані з кожною назвою місцевості. Містять також території місцевого самоврядування, поліцейські ділянки та стандартні регіони. Ці території вибрано через доволі часте вживання в мові та у ЗМІ. Інші адміністративні території, які на поточний момент не входять до газетира, містять реєстрацію району народження, смерті та шлюбу, території пожежних служб, комісію мирних територій (наприклад, муніципального суду) та комплекс географічних територій Національної служби охорони здоров'я.

В останні роки спостерігалась тенденція використання адміністративних територій як основи загального опису географічного місцеположення. У цьому підході є декілька головних недоліків:

1. Такі адміністративні території були створені для сприяння надання спеціальних публічних послуг, а не для виконання загальної географічної ролі. Використання їх за прямим призначенням – це надання їм не передбачених Парламентом функції і статусу.

2. Назви й території адміністративних одиниць часто змінюються.

3. Існує декілька типів важливих адміністративних територій. Немає очевидної причини вибирати один тип “адміністративної географії” замість іншого як основу географії загального призначення.

4. Багато сучасних адміністративних одиниць мають назви, які запозичені від назв міста або містечка на цій території. Це, своєю чергою, позбавляє їх користі як основу загальної географії.

5. Багато адміністративних територій не відповідають загальним ідеям культурних особливостей [15].

Фрагмент картографічного зображення газетира географічних назв з прив'язкою до карти покажемо на прикладі адміністративних територій Шотландії (рис. 2).



Рис. 2. Газетир географічних назв адміністративних територій Шотландії [18]

Газетири Угорщини

Угорський електронний газетир FNT – звичайна реляційна база даних, в якій можна здійснити комплексні запити, спеціальне або часткове фільтрування назв чи атрибутивних даних. Детальніше розроблене зберігання та краще структурована база даних забезпечують спеціальніші та швидші запити (рис. 3) [10].

Azonosító	Név	Típus	Megye	Település név	EOVY	EOVX	Magasság	Mag	Lakos	Néps	Vízfo
71575	Karancs	14	13	Karancslapujtő	704514	312905	729	161	0	0	0
36717	Karancs	16	13		696000	312000	0	0	0	0	0
36149	Karancsalja	33	13		702856	309839	0	0	1601	1990	0
71576	Karancsapátfalva	2	13	Karancslapujtő	0	0	0	0	0	0	0
36152	Karancsberény	33	13		702094	316020	0	0	1041	1990	0
71516	Karancsberényi-patak	36	13	Karancsberény	706080	316171	0	0	0	0	0
71517	Karancsberényi-patak	36	13	Karancsberény	702053	315238	0	0	0	0	5001
6279	Karancsdülő	20	11	Törökszentmiklő	751500	206500	0	0	61	1980	0

Рис. 3. Електронний газетир FNT [10]

Електронні газетири містять просторові та атрибутивні дані, тому варто створювати газетири і в середовищі ГІС. Інструменти ГІС дають можливість робити запити і за атрибутами, і за місцезнаходженням, що є основними функціями газетирів. У ГІС-газетирі карта є не лише графічним методом візуалізації, а й може брати участь в аналізі просторових явищ у разі використання назв.

Водночас постають проблеми під час використання таких газетирів для створення карт. Кожен газетир містить топоніми як точкові елементи з парою координат. Ці координати визначаються з точністю до мінути дуги. Це означає, що точність даних Угорщини – 1,3–1,8 км, що недостатньо для картографічного використання у певному масштабі [10].

Фрагмент газетирів адміністративно-територіальних одиниць Угорщини зображено на рис. 4.



Рис. 4. Фрагмент газетирів адміністративно-територіальних одиниць Угорщини [19]

Вимоги міжнародного стандарту ISO 19112 – Географічна інформація – Просторова прив'язка за допомогою географічних ідентифікаторів. Вимоги до газетирів

Властивості газетирів. Всі приклади кожного типу місцезнаходження в просторовій системі координат мають описуватись в газетирі. Може існувати декілька газетирів для одного й того самого типу місцезнаходження, з прикладами місцезнаходжень, описаними різними способами. Газетир повинен мати такі атрибути:

- назву;
- територію використання;
- відповідальну установу.

Також мають бути такі записи:

- розміри;
- система координат.

Детально ці елементів показано в табл. 3 [1].

Таблиця 3

Елементи газетира [1]

Назва елемента	UML ідентифікатор	Опис	Макс. Точність один (1) чи багато (N)	Тип даних (UML клас)
Ідентифікатор	Ідентифікатор	Назва газетира	1	RS_Identifier
Сфера	Сфера	Опис типів місцеположення входить у газетир	1	CharacterString
Територія використання	Територія використання	Географічний домен, що покривається в газетирі	1	EX_GeographicExtent
Відповідальна установа	Відповідальна установа	Назва організації, відповідальної за утримання газетира	1	CI_ResponsibleParty
Система координат	Система координат	Назва системи координат, що використовується в газетирі для опису місцеположення	1	SC_CRS
Тип місцеположення	SI_ Тип місцеположення	Назва типу місцеположення, для якого записані приклади в газетирі	N	association

Висновки. Розглянувши електронні географічні газетири різних країн, а також їхні основні характеристики і вимоги до створення, можна зробити висновок, що вимогам міжнародного стандарту ISO 19112 –“Географічна інформація – Просторова прив’язка з допомогою географічних ідентифікаторів” відповідають усі розглянуті вище газетири. Але найбільш повноцінним, точним та операбельним як стосовно інформаційного наповнення, так і з погляду користування та оновлення інформації, є газетир США та Канади, оскільки він містить найбільше параметрів запити, його дані можна використовувати у популярних комерційних програмах за допомогою простих методів побудови запитів. Записи в цьому газетирі оновлюються щомісячно.

В Україні нині немає достатньої бази створення електронних газетирів, що б відповідала вимогам міжнародного стандарту. Річ у тім, що для побудови газетирів використовується інформація, яка зосереджена у різних відомствах. Вона не є уніфікованою. Вирішити цю проблему допоможе інтеграція різноманітних відомчих даних.

1. ISO 19112 Географічна інформація – Просторова прив’язка за допомогою географічних ідентифікаторів.
2. Проект “Цифрова бібліотека Александрії”, Університет Каліфорнії / <http://www.alexandria.ucsb.edu/>.
3. Група експертів ООН з географічних назв / <http://unstats.un.org/unsd/geoinfo/>.
4. Технічний комітет ICO з географічної інформації / <http://www.isotc211.org/>.
5. Адресная сеть улиц города Ханты-Мансийска / www.uriit.ru/otchet2003/2_1_6.pdf.
6. Рада Канади з географічних назв / http://geonames.nrcan.gc.ca/info/gnbc_e.php.
7. Постійний комітет Об’єднаного Королівства з географічних назв для британського службового користування / <http://www.pcgn.org.uk>.
8. Рада США з географічних назв / <http://geonames.usgs.gov/bgn.html>.
9. Карпінський Ю. О., Лященко А. А. Стратегія формування національної інфраструктури георосторових даних в Україні. – К.: НДІГК, 2006. – 108 с. – (Сер. “Геодезія, картографія, кадастр”).
10. Guszlev IA., Lukács L. Hungarian Toponyms in Gazetteers / studiacrescent.com/

HUNGARIAN%20TOPONYMS%20IN%20GAZETTEERS.pdf. 11. Drew D., Royce D.A., A. van Baaren *Address Register Research at Statistics Canada / www.amstat.org/sections/srms/proceedings/papers/1989_042.pdf*. 12. H. Southall *Great Britain Historical Gazetteer / GIS / www.port.ac.uk/.../gbhgis/aboutthebghistoricalgis/.../filetodownload, 22744, en.pdf*. 13. M. McCart Wells, C. Anderson, H. Perkins, E. Wells, S. Yurman *Developing a Comprehensive Standard for US Address Data / xml.coverpages.org/GSDI-10-US-AddressStandard.pdf*. 14. *Електронний газетир поштових та ZIP-кодів США і Канади / http://www.zipinfo.com/products/zsig/zsig.htm*. 15. *Електронний газетир географічних назв Великобританії / http://www.gazetteer.co.uk/section1.htm*. 16. *Електронний газетир поштових та ZIP-кодів Великобританії / www.zorbathegeek.com/.../uk-post-code-geocoding.html*. 17. *Фрагмент картографічного зображення електронного газетира населених пунктів Канади / http://atlas.nrcan.gc.ca/site/english/search/GazetteerMapService*. 18. *Газетир географічних назв адміністративних територій Шотландії / http://www.geo.ed.ac.uk/scotgaz/scotland.html*. 19. *Газетир адміністративно-територіальних одиниць Угорщини / http://www.world-gazetteer.com/wg.php*. 20. *Визначення газетира / http://en.wikipedia.org/wiki/Gazetteer*.