БЕЗПЕКОВИЙ СЕКТОР ДЕРЖАВИ

УДК 351.74:004.056(477)
DOI https://doi.org/10.32850/LB2414-4207.2024.34.26

OSINT В КОНТЕКСТІ КІБЕРБЕЗПЕКИ

Демедюк Сергій Васильвович,
кандидат юридичних наук,
заступник Секретаря Ради
(Рада національної безпеки і оборони
України, м. Київ, Україна)

У статті акцентовано на важливості використання розвідки з відкритих джерел в сучасних умовах протидії кіберзлочинності. Зазначено, що спільна інфраструктура Інтернету створює потенціал для взаємопов'язаних вразливостей для всіх користувачів. Водночас, кіберпростір має унікальні особливості, що дозволяють злочинцям вчиняти злочини: глобалізація, яка надає злочинцям нові можливості для виходу за межі звичайних кордонів; розподілені мережі, які створюють нові можливості для віктимізації; синоптизм і паноптизм, які дають можливість стежити за жертвами віддалено; та сліди даних, які можуть надати злочинцям нові можливості для вчинення крадіжки особистих даних.

Акцентовано увагу на терміні «OSINT», його змісті та походженні. Зазначено, що OSINT охоплює різні публічні джерела, такі як академічні публікації (наукові роботи, публікації конференцій тощо), джерела ЗМІ (газети, радіоканали, телебачення тощо), веб-контент (веб-сайти, соціальні мережі тощо) та публічні дані (відкриті урядові документи, оголошення публічних компаній) тощо. Наводиться коротке змістовне порівняння з подібними термінами LITINT, WEBINT, SOCMINT.

Зазначено, що правоохоронні органи і служби безпеки звертаються до OSINT за додатковим розширенням і поглибленням інформації, щоб підсилити і допомогти перевірити контекстні знання. Збільшення кількості і типів викликів для сучасних фахівців з національної безпеки та правоохоронних органів прискорило використання відкритих джерел в Інтернеті з метою забезпечення розуміння більш цілісної картини щодо осіб або діяльності.

Акцентовано на тому, що для аналітика OSINT є ключовим завдяки можливостям браузерів, пошукових систем, веб-сайтів, баз даних, індексації, пошуку та аналітичних додатків. Однак існують ключові проблеми, які можуть відволікти від правильного напрямку OSINT, такі як збір даних з великих відкритих записів в Інтернеті та інтеграція даних для розширення можливостей. Тому фахівцями в усьому світі використовується величезна різноманітність інструментів та методів OSINT, серед яких виділяються найпоширеніші інструменти для збору, зберігання та класифікації даних з відкритих джерел: збір даних (Data acquisition), походження даних (Data provenance), зберігання даних (Data storage), управління даними (Data curation) та візуалізація (та сумісність) даних (Data visualization (and interaction)).

Ключові слова: Open Source Intelligence, OSINT, intelligence, кібербезпека, кіберзлочин.

OSINT IN THE CONTEXT OF CYBERSECURITY

Demediuk Serhii Vasylvovych,
Candidate of Juridical Sciences,
Deputy Secretary
(National Security Council of Ukraine,
Kyiv, Ukraine)

The article emphasizes the importance of using open-source intelligence in the current context of countering cybercrime. It is noted that the shared Internet infrastructure creates the potential for interconnected vulnerabilities for all users. At the same time, cyberspace has unique features that allow criminals to commit crimes: globalization, which provides criminals with new opportunities to go beyond conventional borders; distributed networks, which create new opportunities for victimization; synoptism and panoptism, which allow victims to be monitored remotely; and data traces, which can provide criminals with new opportunities to commit identity theft.

Attention is focused on the term "OSINT", its content and origin. It is noted that OSINT covers various public sources, such as academic publications (research papers, conference publications, etc.), media sources (newspapers, radio channels, television, etc.), web content (websites, social networks, etc.) and public data (open government documents, announcements of public companies), etc. A brief meaningful comparison with similar terms such as LITINT, WEBINT, SOCMINT is provided.

It is noted that law enforcement and security agencies are turning to OSINT for additional expansion and deepening of information to enhance and help verify contextual knowledge. The increase in the number and types of challenges for modern national security and law enforcement professionals has accelerated the use of open sources on the Internet to provide an understanding of a more holistic picture of individuals or activities.

It is emphasized that OSINT is key for analysts due to the capabilities of browsers, search engines, websites, databases, indexing, searching, and analytical applications. However, there are key challenges that can distract from the right direction of OSINT, such as collecting data from large public records on the Internet and integrating data to expand capabilities. Therefore, a huge variety of OSINT tools and techniques are used by professionals around the world, including the most common tools for collecting, storing, and classifying data from open sources: Data acquisition, Data provenance, Data storage, Data curation, and Data visualization (and interaction).

Key words: Open Source Intelligence, OSINT, intelligence, cybersecurity, cybercrime.

Постановка проблеми. Вплив кіберзлочинності змусив розвідувальні та правоохоронні органи по всьому світу боротися з кіберзагрозами. Перед усіма секторами зараз стоять схожі дилеми: як найкраще захиститися від кіберзлочинності і як ефективно сприяти безпеці людей і організацій. Отримання унікальних і цінних розвідувальних даних шляхом збору публічних записів для створення всебічного профілю певних цілей швидко стає важливим засобом для розвідувального співтовариства. Оскільки кількість доступних відкритих джерел стрімко зростає, протидія кіберзлочинності все більше залежить від передових програмних засобів і методів збору та обробки інформації в ефективний і результативний спосіб.

Метою статті є розкриття ролі OSINT в контексті кібербезпеки. Розглядаються інструменти і методи збору та аналізу OSINT, а також суміжні роботи, які є основними складовими його внеску.

Виклад основного матеріалу. У XXI сторіччі цифровий світ став «палицею з двома кінцями» [1; 2]. Завдяки революції загальнодоступних джерел (тобто відкритих

джерел) цифровий світ надав сучасному суспільству величезні переваги, але в той же час проблеми інформаційної незахищеності висвітлили вразливості та слабкості [3; 2]. Спільна інфраструктура Інтернету створює потенціал для взаємопов'язаних вразливостей для всіх користувачів [4]: «Віруси, хакери, витік безпечної і приватної інформації, системні збої і переривання послуг» з'явилися в бездонному потоці [2].

Wall [5; 6; 7] та Nykodym та ін. [8, 9] обговорювали, що кіберпростір має чотири унікальні особливості, які називаються «трансформаційними ключами», що дозволяють злочинцям вчиняти злочини:

– глобалізація, яка надає злочинцям нові можливості для виходу за межі звичайних кордонів;

– розподілені мережі, які створюють нові можливості для віктимізації;

– синоптизм і паноптизм, які дають можливість стежити за жертвами віддалено;

– сліди даних, які можуть надати злочинцям нові можливості для вчинення крадіжки особистих даних.

На додаток до вищезазначеного, Хоббс та ін. [3] стверджують, що однією з основних тенденцій розвитку Інтернету останніх років є те, що «підключення до Інтернету може бути дуже ризикованою справою».

Так само, як і епідемічне використання та розвиток технологій мобільного зв'язку, використання відкритих джерел поширюється на сфери розвідки, політики та бізнесу [3]. У той час як традиційні джерела та інформаційні канали (ЗМІ, бази даних, енциклопедії тощо) були змушені адаптуватися до нового віртуального простору, щоб зберегти свою присутність, багато «нових» медіа-джерел (особливо з соціальних мереж) поширюють велику кількість користувацького контенту, який згодом змінив інформаційний ландшафт. Прикладами масштабу користувацької інформації є 500 мільйонів твітів на день у Twitter та 98 мільйонів щоденних записів у блогах на Tumblr [3], а також мільйони індивідуальних персональних сторінок у Facebook. З розвитком інформаційного ландшафту для запобігання та виявлення терористичної діяльності правоохоронним органам необхідно збирати відповідний контент за допомогою розслідувань і регульованого спостереження [10].

Open Source Intelligence (OSINT) походить від служб національної безпеки та правоохоронних органів [11] і в межах нашого аналізу, посилаючись на джерела, будемо цей термін визначати як «сканування, пошук, збір, вилучення, використання, перевірка, аналіз і обмін розвідувальною інформацією зі споживачами відкритих джерел і загальнодоступних даних з несекретних, нетаємних джерел» [12; 10]. OSINT охоплює різні публічні джерела, такі як академічні публікації (наукові роботи, публікації конференцій тощо), джерела ЗМІ (газети, радіоканали, телебачення тощо), веб-контент (веб-сайти, соціальні мережі тощо) та публічні дані (відкриті урядові документи, оголошення публічних компаній тощо) [13; 14].

OSINT традиційно описується як пошук у загальнодоступних опублікованих джерелах [15], таких як книги, журнали, газети, брошури, звіти тощо. Це часто називають літературною розвідкою або LITINT [16].

Однак, швидке зростання цифрових медіа-джерел в Інтернеті та суспільному мовленні розширило сферу діяльності з відкритим вихідним кодом [17]. Оскільки існують різноманітні загальнодоступні онлайнові джерела, з яких ми можемо збирати розвідувальну інформацію, багато авторів описують цей тип OSINT як WEBINT. Дійсно, терміни WEBINT і OSINT часто використовуються як взаємозамінні [13; 14]. Соціальні медіа, такі як соціальні мережі, спільноти для обміну медіа та спільні проекти – це сфери, де створюється більшість користувацького контенту. Тому існує також і розвідка у соціальних медіа (Social Media Intelligence (SOCMINT)) – це «розвіддані, які

збираються з сайтів соціальних медіа». Деяка інформація з них може бути у відкритому доступі без будь-якої автентифікації, необхідної для проведення розслідування [18, с. 36; 13; 14].

Багато правоохоронних органів і служб безпеки звертаються до OSINT за додатковим розширенням і поглибленням інформації, щоб підсилити і допомогти перевірити контекстні знання. На відміну від типових ІТ-систем, які можуть приймати лише обмежений діапазон вхідних даних, джерела даних OSINT настільки ж різноманітні, як і сам Інтернет, і будуть продовжувати розвиватися в міру розширення технологічних стандартів [11]: «OSINT може забезпечити фон, заповнити епістемічні прогалини і створити зв'язки між, здавалося б, непов'язаними джерелами, в результаті чого розвідувальна картина стає більш повною» [3, с. 2].

OSINT дедалі більше залежить від асиміляції збору та аналізу даних з усіх джерел. Така розвідка є невід'ємною частиною «національної безпеки, конкурентної розвідки, бенчмаркінгу і навіть інтелектуального аналізу даних на підприємстві» [4]. OSINT вже давно використовується урядом, військовими і в корпоративному світі для того, щоб стежити за конкурентами і мати конкурентну перевагу [13; 14]. Крім того, велика кількість інтернет-користувачів займається легальною діяльністю «від комунікацій і комерції до ігор, знайомств і ведення блогів» [4, с. 6], і OSINT відіграє в цьому контексті важливу роль.

Збільшення кількості і типів викликів для сучасних фахівців з національної безпеки, розвідки, правоохоронних органів і служб безпеки прискорило використання відкритих джерел в Інтернеті, щоб допомогти скласти більш цілісну картину про людей, організації та діяльність [4]. Американське опитування PWC (PricewaterhouseCoopers) [19] під назвою «Ключові висновки з дослідження стану кіберзлочинності в США за 2015 рік», в якому взяли участь понад 500 керівників американських компаній, правоохоронних органів і державних установ, стверджує, що «кіберзлочинність продовжує з'являтися в заголовках газет і викликати головний біль у керівників підприємств». 76 % керівників, відповідальних за кібербезпеку, заявили, що цього року вони більше занепокоєні кіберзагрозами: «Кількість інцидентів у сфері кібербезпеки не лише зростає, вони також стають дедалі більш руйнівними і націленими на все ширший спектр інформації та векторів атак» [19].

У звіті Міністерства внутрішньої безпеки США зазначено, що до критично важливих сфер діяльності, де застосування OSINT є життєво необхідним, належать загальна розвідка, завчасне попередження, внутрішня боротьба з тероризмом, захист критичної інфраструктури (в тому числі кіберпростору), захист від катастрофічного тероризму, а також готовність до надзвичайних ситуацій і реагування на них [20]. Тому розвідка, служби безпеки і громадська безпека збирають великі обсяги даних з різних джерел, включаючи кримінальні справи про терористичні інциденти і загрози кібербезпеці [20].

Глассман і Канг [21] розглядають OSINT як результат змін у відносинах між людиною та інформацією, що відбуваються в результаті появи і зростаючого домінування Всесвітньої мережі в повсякденному житті. Соціально неприйнятну поведінку було виявлено на веб-сайтах, блогах і онлайн-спільнотах усіх видів «від експлуатації дітей до шахрайства, екстремізму, радикалізації, переслідувань, крадіжки особистих даних і витоку приватної інформації». Крадіжка особистих даних і розповсюдження незаконно «скопійованих фільмів, телепередач, музики, програмного та апаратного забезпечення є гарними прикладами того, як Інтернет збільшив вплив злочинності» [3].

Глобалізація, швидкість поширення, анонімність, транскордонний характер інтернету, а також відсутність відповідного законодавства чи міжнародних угод зробили

деякі з них дуже поширеними і дуже складними для судового розгляду [22]. Існують різні типи темних сторін інтернету, а також програми для висвітлення темних сторін, що включають як технологічно орієнтовані, так і нетехнологічно орієнтовані.

До технологічно орієнтованих темних сторін належать спам, шкідливе програмне забезпечення, хакерство, DoS атаки, фішинг, порушення прав на цифрову власність. Темні сторони, не пов'язані з технологіями, включають онлайн-scams і шахрайство, фізичну шкоду, кібербулінг, поширення неправдивої інформації та нелегальні азартні ігри в Інтернеті. При цьому, нетехнологічні заходи реагування передбачають розробку відповідного законодавства, активність правоохоронних органів, посидення судових процесів, розвиток міжнародного співробітництва, активна позиція громадянського суспільства, розвиток освіти, посилення обізнаності і обережності людей [22].

Комп'ютерні злочини і цифрові докази збільшуються такими темпами, які поки що не піддаються вимірюванню, за винятком поодиноких досліджень [3]. Для аналітика OSINT є ключовим завдяки можливостям браузерів, пошукових систем, веб-сайтів, баз даних, індексації, пошуку та аналітичних додатків [4]. Однак існують ключові проблеми, які можуть відволікти від правильного напрямку OSINT-проектів, такі як збір даних з великих відкритих записів в Інтернеті та інтеграція даних для розширення можливостей параметрів OSINT-проектів [11].

Сучасні інформаційні фахівці використовують різноманітні методи для організації відкритих джерел, включаючи, але не обмежуючись ними, аналіз веб-посилань, метрики, методи сканування, картографування джерел, видобуток тексту, створення онтології, аналіз блогів і методи розпізнавання образів. Алгоритми розробляються з використанням обчислювальної топології, гіпер-графів, аналізу соціальних мереж (SNA), виявлення знань та інтелектуального аналізу даних (KDD), агентного моделювання, аналізу динамічних інформаційних систем тощо [23].

Фахівцями в усьому світі використовується величезна різноманітність інструментів та методів OSINT, серед яких можемо виділити найпоширеніші інструменти для збору, зберігання та класифікації даних з відкритих джерел.

Кодування даних (*Data encoding*) – термін «кодування» означає процес перетворення послідовності символів у спеціальний формат для передачі або зберігання. У веб-середовищі відповідні набори даних відновлюються з сервісів даних, доступних локально або глобально в Інтернеті. Залежно від сервісу і типу інформації, дані можуть бути представлені в різних форматах. Платформи моделювання необхідні для взаємодії з різними форматами даних, включаючи звичайний текст, мови розмітки та бінарні файли [24; 25]. *Приклади: Система «Геоінформатика для геохімії» (веб-сервіси баз даних, що приймають звичайний текстовий формат), base 64online Encoder, XML encoder.*

Збір даних (*Data acquisition*) – Автоматичний збір даних з різних джерел (наприклад, датчиків і зчитувачів на заводі, в лабораторії, медичному або науковому середовищі). Збір даних зазвичай проводився через точки доступу до даних і веб-посилання, такі як http або ftp-сторінки, але вимагав періодичного оновлення. Використання каталогу дозволяє перевірити доступні джерела даних перед їх збором [24; 26]. *Приклади: Каталоги метаданих.*

Походження даних (*Data provenance*) – Цей термін використовується для позначення процесу відстеження та реєстрації походження даних та їх переміщення між базами даних. За концепцією походження стоїть динамічна природа даних. Замість того, щоб створювати різні копії одного і того ж набору даних, важливо відстежувати зміни і зберігати запис процесу, який призвів до поточного стану. Таким чином, походження даних може гарантувати надійність даних і відтворюваність результатів. Наразі походження даних стає дедалі важливішим питанням у наукових базах даних, де воно

відіграє центральну роль у перевірці якості, зручності та надійності даних (зокрема, у семантичних веб-сервісах) [24; 27; 28; 29]. *Приклади: Розподілені системи контролю версій, такі як Git, Mercurial. Розподілені системи управління версіями були розроблені для полегшення відстеження змін у документах, кодах, наборах текстових даних, а віднедавна і в геопросторових даних.*

Зберігання даних (*Data storage*) – Цей термін стосується практики зберігання електронних даних за допомогою сторонніх сервісів, доступ до яких здійснюється через Інтернет. Це альтернатива традиційним локальним сховищам (наприклад, дисковим або стрічковим накопичувачам) і портативним сховищам (наприклад, оптичним носіям або флеш-накопичувачам). Його також можна назвати «хостинговим сховищем», «інтернет-сховищем» або «хмарним сховищем». Реляційні бази даних (БД) наразі є найкращим вибором для зберігання та обміну даними [24; 25]. *Приклади: Postgre SQL, MySQL, Oracle, NoSQL.*

Управління даними (*Data curation*) – спрямоване на виявлення та пошук даних, забезпечення якості даних, додавання цінності, повторне використання та збереження з плином часу. Включає відбір та оцінку даних творцями та архівістами; еволюційне забезпечення інтелектуального доступу; надлишкове зберігання; трансформацію даних. Управління даними має вирішальне значення для оцифрування, обміну, інтеграції та використання наукових даних [25; 30]. *Приклади: Сховища даних, вітрини даних, інструменти Плану управління даними (DMPTool). Інструменти DMP створюють готові до використання плани управління даними для конкретних фінансових установ, щоб задовольнити вимоги донорів до планів управління даними, отримати покрокові інструкції та рекомендації щодо даних, а також дізнатися про ресурси та послуги, доступні в установі.*

Візуалізація (та сумісність) даних (*Data visualization (and interaction)*) – Цей термін означає представлення даних у візуальному або графічному форматі (наприклад, створення таблиць, зображень, діаграм та інших інтуїтивно зрозумілих способів розуміння даних). Інтерактивна візуалізація даних йде ще далі: вона виходить за межі відображення статичної графіки та електронних таблиць і дозволяє використовувати комп'ютери та мобільні пристрої для більш детального вивчення діаграм і графіків, а також інтерактивно (і миттєво) змінювати дані, які ви бачите, і способи їх обробки [24; 25]. *Приклади: Poly Maps, NodeBox, FF Chartwell, SAS visual Analytics, Google Map*

Аналітичні інструменти OSINT формують системи методів інтелектуального аналізу даних, візуалізації закономірностей і пропонують аналітичні моделі для розпізнавання та реагування на виявлені закономірності. Ці інструменти повинні поєднувати незамінні функції і містити інтегровані алгоритми і методи, що підтримують типові методи інтелектуального аналізу даних, включаючи (але не обмежуючись ними) класифікацію, регресію, асоціативний аналіз і аналіз наборів елементів, схожість і кореляцію, а також нейронні мережі [31]. Такі аналітичні інструменти – це програмні продукти, які надають прогнозні та рекомендаційні аналітичні програми, деякі з них працюють на великих обчислювальних платформах з відкритим вихідним кодом, як правило, паралельних обчислювальних системах, заснованих на кластерах товарних серверів, масштабованих розподілених сховищах і технологіях, таких як бази даних Hadoop і NoSQL. Загалом інструменти призначені для того, щоб надати користувачам можливість швидко аналізувати великі обсяги даних [32].

Висновки. Таким чином, розвиток інформаційних технологій, кіберпростору та медіа ресурсів, з однієї сторони, формують технологічні умови поширення нових видів кіберзлочинів, з іншого, спонукають до поглиблення безпекової діяльності на основі тих же IT технологій, використання сучасних методів та інструментів збору, обробки та аналізу даних.

Список використаних джерел:

1. Gregory M., Glance, D. Cyber-crime, cyber security and cyber warfare. Security and networked society. Springer, 2013. p. 51–95.

2. Yuan T., Chen P. Data mining applications in E-Government information security. 2012 international workshop on information and electronics engineering (IWIEE). Proc. Eng., 2012, vol. 29, p. 235–240.

3. Hobbs Ch. Morgan M., Salisbury D. Open source intelligence in the twenty-first century. Palgrave, 2014, pp. 1–6.

4. Appe E. J. Behavior and technology, Internet Searches for Vetting, Investigations, and Open-Source Intelligence. Taylor and Fransic Group, 2011, pp. 3–17.

5. Wall D. S. The internet as a conduit for criminal activity. In: Pattavina, A. (ed.) Information technology and the criminal justice system. Sage Publications, USA. 2005.

6. Wall D. S. Hunting shooting, and phishing: new cybercrime challenges for cybercanadians in the 21st century. The ECCLES centre for american studies. 2007.

7. Wall D. S. Hunting shooting, and phishing: new cybercrime challenges for cyber canadians in the 21st Century. The Eccles Centre for American Studies. www.bl.uk/ecclescentre. The British Library Publication. 2008.

8. Nykodym N., Taylor R., Vilela J. Criminal profiling and insider cyber crime. Digital Investigation, 2005.vol. 2, p. 261–267.

9. Babak Akhgar, P. Saskia Bayerl, Fraser Sampson. Open Source Intelligence Investigation. From Strategy to Implementation. Advanced Sciences and Technologies for Security Applications. Springer International Publishing AG. 2016.

10. Koops B. J., Hoepma J. H., Leenes R. Open-source intelligence and privacy by design. Computer Law and Security Review, 2013, vol. 2(9), p. 676–688.

11. Kapow Software. 2013. URL: http://www.kofax.com: http://www.kofax.com/go/kapow/wp-building-yourosint-capability

12. Fleisher C. OSINT: its implications for business/competitive intelligence analysis and analysts. Inteligencia y Seguridad, 2008. vol. 4, p. 115–141.

13. Chauhan S., Panda K. Open source intelligence and advanced social media search. Hacking web intelligence open source intelligence and web reconnaissance concepts and techniques. Elsevier, 2015. pp. 15–32.

14. Chauhan S., Panda K. Understanding browsers and beyond. Hacking web intelligence open source intelligence and web reconnaissance concepts and techniques. Elsevier, 2015, pp. 33–52.

15. Burwell H. P. Online competitive intelligence: increase your profits using cyber-intelligence. Facts on Demand Press, Tempe, AZ. 2004.

16. Clark R. M. Intelligence analysis: a target-centric approach. CQ Press, Washington, DC. 2004.

17. Boncella R. J. Competitive intelligence and the web. Commun AIS, 2003, vol. 12, p. 327–340.

18. Omand D., Miller C., Bartlett J. Towards the discipline of social media intelligence (2014). In: Hobbs, Morgan, Salisbury (eds.) Open source intelligence in the twenty-first century. Palgrave, 2014, p. 24–44.

19. PWC cyber security. 2015. URL: https://www.pwc.com/us/en/increasing-it-effectiveness/publications/assets/2015-us-cybercrime-survey.pdf. Retrieved from http://www.pwc.com/cybersecurity

20. Chen H., Chiang R. H. L., Storey V.C. Business intelligence and analytics: from big data to big impact. Business Intelligence Research, 2012, vol. 36(4), p. 1–24.

21. Kang M. J. Intelligence in the internet age: the emergence and evolution of Open Source Intelligence (OSINT). Computers in Human Behavior, 2012, vol. 2, p. 673–682.

22. Kim W., Jeong O. R., Kim Ch., So J. The dark side of the Internet: attacks, costs and responses. Information Systems, 2011, vol. 36, p. 675–705.

23. Brantingham P. L. Computational Criminology. 2011 European intelligence and security informatic conference. IEEE Computer Society. 2011. DOI:10.1109/EISIC.2011.79

24. Vitolo C., Elkhatib Y., Reusser D., Macleod C. J. A., Buytaert W. Web technologies for environmental Big Data. Environmental Modelling and Software, 2015, vol. 63, p. 185–198.

25. Webopedia.com. (n.d.). URL: Webopedia.com.

26. Ames D. P., Horsburgh J. S., Cao Y., Kadlec J., Whiteaker T., Valentine D. Hydro desktop: web services-based software for hydrologic data discovery, download, visualization, and analysis. Environmental Modelling and Software, 2012, vol. 37, p. 146–156.

27. Buneman P., Khanna S., Chiew Tan W. Data provenance: some basic issues. University of pennsylvania scholarly commons. 2000. URL: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1210&context=cis_papers

28. Szomszor M., Moreau L. Recording and reasoning over data provenance in web and grid services. On the move to meaningful internet systems, 2003, pp. 603–620.

29. Tilmes C., Yesha Ye., Halem M. Distinguishing provenance equivalence of earth science data. International Conference on Computational Science (ICCS). 2010. p. 1–9.

30. Dou L., Cao G., Morris P. J., Morris R. A., Ludäscher B., Macklin J. A., Hanken J. Kurator: a Kepler package for data curation workflows. International Conference on Computational Science, ICCS 2012, Procedia Computer Science, 2012, vol 9, pp. 1614–1619. DOI:10.1016/j.procs.2012.04.177

31. Harvey C. 50 top open source tools for big data. 2012. URL: http://www. datamation.com/data-center/50-top-open-source-tools-for-big-data-1(2,3).html

32. Loshin D. How big data analytics tools can help your organization. 2015. URL: http://searchbusinessanalytics.techtarget.com/feature/How-big-data-analytics-tools-can-help-yourorganization.