

Leśne Big Data

Analiza wielkich zbiorów danych daje tak samo potężne możliwości. SILP to przypuszczalnie największy system informatyczny na świecie, pomagający zarządzać obszarami leśnymi. Czy potrafimy efektywnie z tych zasobów korzystać? Warto się tego nauczyć, ale też przeprowadzić w leśnych danych swego rodzaju inwentaryzację.

System Informatyczny Lasów Państwowych (SILP) – czasami obiekt westchnień, ale częściej przyczyna frustracji zdecydowanej większości nowo zatrudnionych pracowników Lasów Państwowych. Wielogłowa hydra, która swoim znakowym menu zaważnęła większość aktywności znanych w Lasach. Jedyny system na świecie, w którym zatwierdzamy poprzez zaprzeczenie. Oczywiście to nieco przesada, ale jednak trauma paru lat w stanie posiadania i nie tylko pozostawia uraz do końca życia. Nieco poważniej jednak jest to system, który pozwala gromadzić i akumuluje ogromne ilości danych. To najprawdopodobniej największy system informatyczny na świecie spośród tych, które pomagają zarządzać obszarami leśnymi. Oczywiście to także jest pewne uogólnienie, które dotyczy głównie liczby zbieranych danych.

Nic nie robi takiego wrażenia na leśnikach z zagranicy jak nasz SILP (tutaj w bardzo szerokim tego słowa znaczeniu, czyli razem z LMN itp.) i zakres informacji, jaką w nim gromadzimy. Pewnie dlatego, że nie wiemy, ile z nich tak naprawdę wykorzystujemy w praktyce. I dobrze, w poczuciu patriotycznego obowiązku sam tego nigdy nie wyjaśniam.

Big Data i SIP

SILP nie jest wyjątkiem, danych gromadzimy coraz więcej i więcej w różnych bazach, dyskach, telefonach i gdzie się tylko da. Ich pozyskiwanie jest coraz tańsze i sami często nieświadomie się do tego procesu włączamy. W tej powodzi kolejnych terabajtów danych tylko jedna rzecz przychodzi nam coraz trudniej – analiza wszystkich tych informacji, oddzielenie ziarna od plew.

Problemu tak dużych zbiorów danych i ich analizy dotyka (już teraz dziedziny nauki) właśnie Big Data. Termin ten odnosi się do dużych i niejednorodnych zbiorów danych, które trudno przeanalizować, ale jednocześnie, jeśli nam się uda, może to przynieść cenną

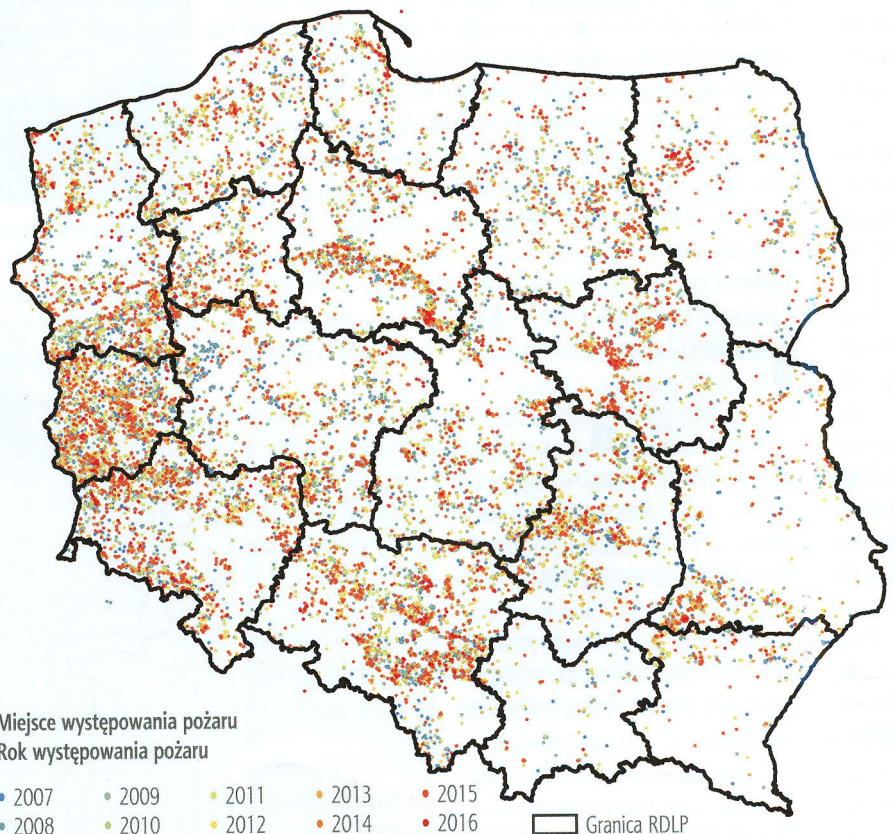
wiedzę. Trudno jest jednoznacznie zdefiniować, jak duże zbiory danych są już zbiorami „big”, ale z reguły przyjęło się przyjmować, że chodzi o takie, których nie da się analizować konwencjonalnymi metodami. W przypadku rejestru drewna z kilku lat dla jednego nadleśnictwa wystarczyłoby pewnie kilkadziesiąt megabajtów znakowych informacji, żeby nas całkowicie pogrążyć, a w przypadku danych z lotniczego skanowania laserowego byłyby to pewnie terabajty.

Nie jest tak, że duże zbiory danych całkowicie w Lasach zaniedbujemy, już jakiś czas funkcjonuje przecież Hurtownia Danych, działa Business Object, a kolejne

aplikacje, również takie połączone z informacją przestrzenną, mają się wkrótce pojawić. Nie znaczy to jednak, że skutecznie i efektywnie z takich zbiorów korzystamy, a jest z czego wybierać.

Pewną specyficzną grupę analiz Big Data stanowią te, które korzystają dodatkowo z dobrodziejstwa, jakie niosą systemy informacji przestrzennej. Dla nas, leśników, są one szczególnie cenne z racji dowiązania większości działań i aktywności do przestrzeni właśnie. Wiemy, gdzie było pozyskane drewno z teoretyczną dokładnością adresu leśnego (być może niebawem będą to rzeczywiste współrzędne), odnotowujemy lokalizacje wykonanych czyn-

Ryc. 1. Pożary obszarów leśnych zarządzanych przez Lasy Państwowe w latach 2009–16



ności i bardzo wielu innych rzeczy, które trafiają do Leśnej Mapy Numerycznej – najczęściej aktualizowanego w skali całego kraju zbioru danych przestrzennych.

Rodzaj analizy

Jako grupa zawodowa korzystamy z map najprawdopodobniej najczęściej w Polsce. Nie tylko nam jednak najłatwiej z reguły interpretować obrazy właśnie.

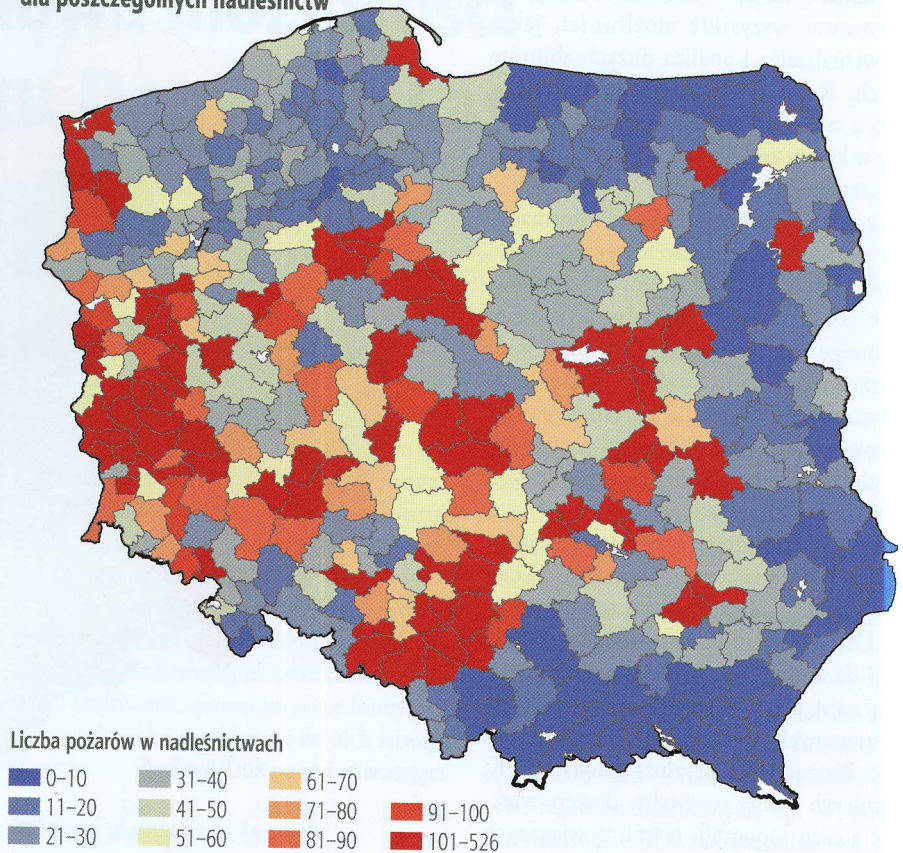
Przykładowo gdybyśmy chcieli zwiualizować pożary lasów, które wybuchały na obszarach leśnych zarządzanych przez Lasy Państwowe w latach 2007 i 2016, to wyglądałoby to dokładnie tak, jak na ryc. 1. Przy bardzo dużej liczbie danych taka generalizacja traci sens, a nasz mózg ma problem z interpretacją tego, co widzą oczy.

Gdybyśmy liczbę tych samych pożarów odnieśli do nadleśnictw, to obraz nieco już zyskuje na czytelności, ale dalej wiele informacji może być zgeneralizowanych błędnie (ryc. 2). Wszystko bowiem będzie zliczane do dużo większych jednostek przestrzennych niż punktowo (w dużym uproszczeniu) występujący pożar.

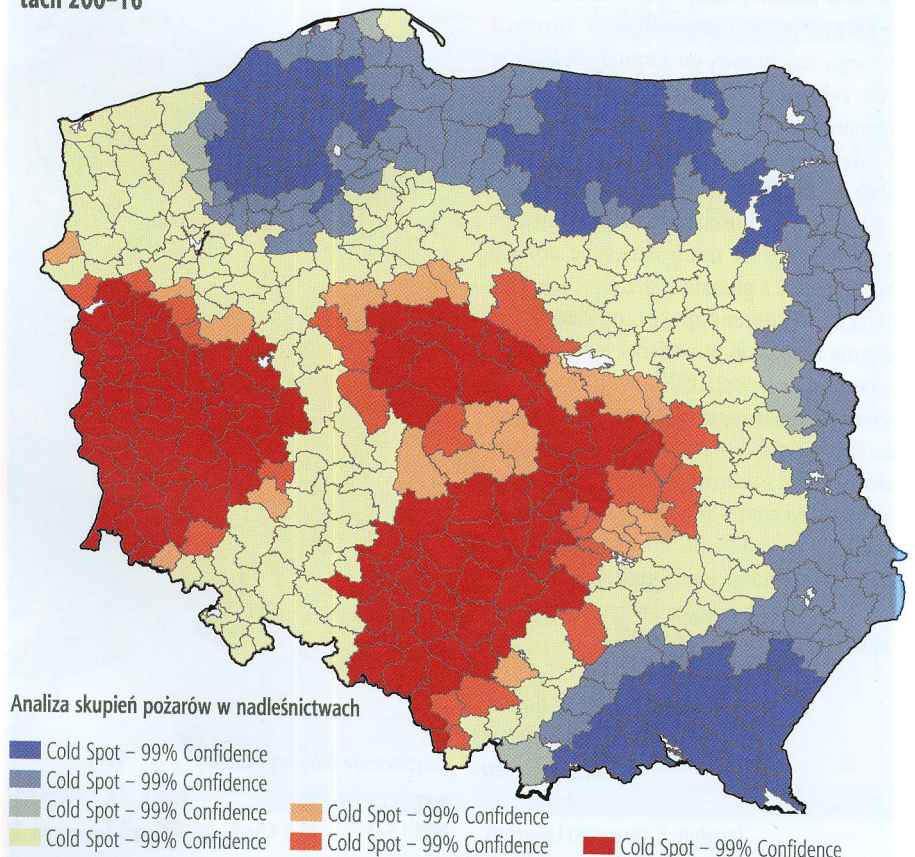
Gdybyśmy jednak skorzystali w tym konkretnym przypadku z jednej z technik GIS i wykonali tzw. analizę skupień dla tych pożarów, dostalibyśmy odpowiedź w postaci wizualizacji niosącej w sobie już znacznie więcej informacji (ryc. 3). W tym przypadku dane przestrzenne zostały pogrupowane, a ich relacje przestrzenne wyrażone intensywnością zastosowanej skali barw. Obszary najbardziej czerwone w latach 2007–16 charakteryzują się największym skupieniem pożarów. Oczywiście każde z analizowanych lat na takiej wizualizacji wygląda inaczej, jednak, co ciekawe, na wszystkich dominuje obszar południowo-zachodniej Polski, a po podsumowaniu także Polska centralna aż po Beskidy. Należy pamiętać, że w tym konkretnym przypadku nie braliśmy pod uwagę powierzchni pożaru, a jedynie fakt jego wystąpienia.

Innym zgoła przykładem jest wizualizacja, która obrazuje liczbę kilometrów dróg leśnych przypadających na kompleksy leśne. Pierwsza z map pokazuje nam liczbę dróg w odniesieniu do poszczególnych dystryktów regionalnych, a druga z dowiązaniem do konkretnych nadleśnictw. Oczywiście analiza ta nie dostarcza prostej odpowiedzi na pytanie, gdzie jest łatwiej zerwać i wywieźć drewno, bo nie brano pod uwagę ani jakości tych dróg, ani dostępności dróg publicznych. Niemniej jednak wizualizacja ta uprawnia do formułowania pewnych ogólnych stwierdzeń. ▶

Ryc. 2. Liczba pożarów lasów na gruntach zarządzanych przez Lasy Państwowe podsumowana dla poszczególnych nadleśnictw



Ryc. 3. Analiza skupień pożarów lasów na gruntach zarządzanych przez Lasy Państwowe w latach 200–16



► Potrzebny „remanent” danych

Nie chodzi o to, by w tak krótkim artykule prezentować wszystkie możliwości, jakie daje wizualizacja i analiza dużych zbiorów danych. Raczej zwrócić uwagę na to, że warto z nich korzystać, ale też opłaca się może w końcu zrobić w tych naszych danych jakiś „remanent”.

Czy naprawdę potrzebujemy informacji, które podczas prac urzędniowych są zwyczajnie kopiowane na bazie innych atrybutów, takich jak np. wilgotność na bazie typu siedliskowego lasu? Każdy rekord w SILP-ie to dodatkowy czas na wpisywanie, drukowanie, zestawianie (bo wciąż nie wszystko robi się automatycznie), a w skali kraju to nie minuty, ale miesiące i lata przepracowanych godzin Służby Leśnej.

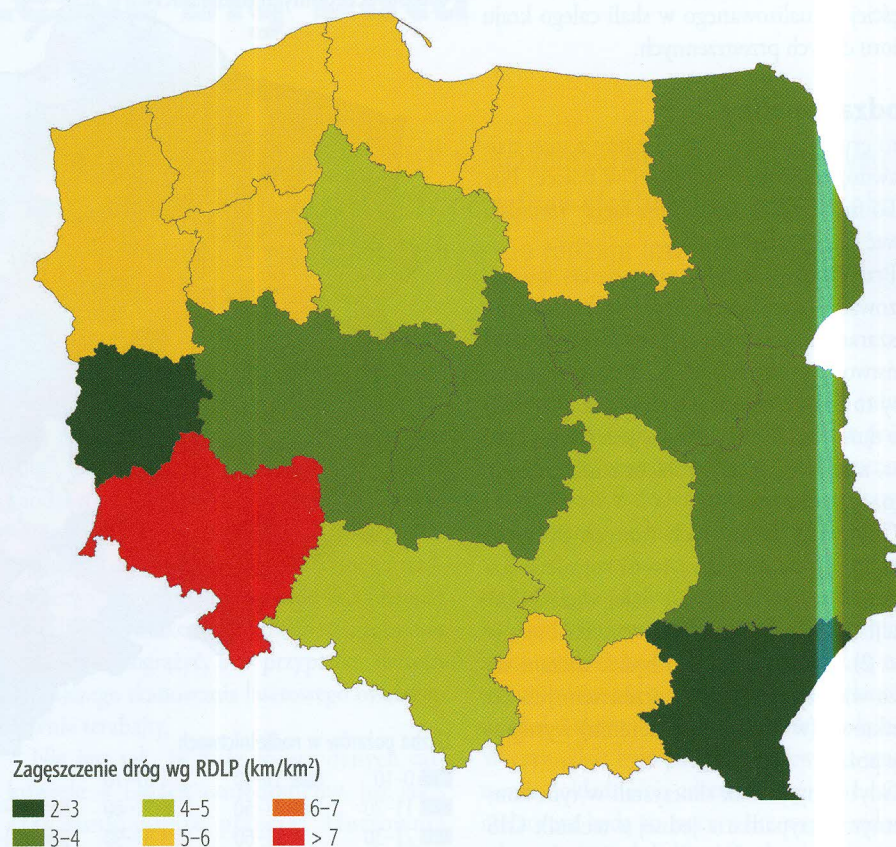
Polska w ostatnim okresie zrobiła bardzo wiele na polu udostępniania danych. Mamy Bank Danych o Lasach, Informatyczny System Ochrony Kraju, Leśne Centrum Informacji dane udostępnia GUS, ale wciąż nie służą nam one do zaawansowanych analiz przestrzennych, o biznesowych nie wspominając. Patrząc przez pryzmat danych ISOK, do których mamy swobodny dostęp, większość z cech topografii terenu dowiązanych do wydzieleni i opisanych w SILP-ie ma się co najmniej bardzo słabo do rzeczywistości. Nie zastanawiamy się, czy dane o bezrobociu w skali lokalnej mają wpływ na kradzież drewna ani gdzie duzi odbiorcy surowca drzewnego najchętniej go kupują. Być może z tych zależności moglibyśmy skorzystać i wygenerować dodatkowy dochód?

W dzisiejszych czasach koordynacja działań nadleśnictw bez przestrzennej analizy istniejących korelacji to zdecydowanie zbyt mało. Parę lat temu Amerykę obiegła wieść, że jedna z globalnie działających firm kurierskich rekomenduje swoim kierowcom wyłącznie skręcanie w prawo i jednocześnie korzystanie ze specjalnie przygotowanych programów, które optymalizują drogę przejazdu i kolejność dostaw. W pierwszym roku reformy wzrosła liczba dostarczanych paczek przy jednoczesnym spadku przejechanych kilometrów o... 33 mln km! Czy nie lepiej szukać oszczędności tam, gdzie ich wygenerowanie można by osiągnąć w naprawdę prosty sposób? ☺

Radomir Bałazy
Mariusz Ciesielski
Tomasz Hycza

Autorzy są pracownikami
Instytutu Badawczego Leśnictwa.

Ryc. 4. Zagęszczenie dróg wg RDLP



Ryc. 5. Zagęszczenie dróg wg nadleśnictw

