

# Digitalizacja zbiorów zielnikowych – krok w dobrą stronę

## Digitisation of herbarium collections – a step in the right direction

DOMINIK TOMASZEWSKI

Instytut Dendrologii PAN, ul. Parkowa 5, PL-62-035 Kórnik  
e-mail: dominito@man.poznan.pl

*Submitted: 13 November 2019; Accepted: 21 November 2019*

**ABSTRACT:** Digitisation involves many types of scientific data, herbarium collections included. We know that herbaria are a source of data whose usage has changed over time but the data do not lose their value for scientists, just the contrary. When taking a decision about digitisation in a herbarium, it is very important that the data acquisition, processing and data access should be organized in a systematic way. Recognised solutions and standards help to choose adequate equipment and software and apply an effective workflow. In the Herbarium of the Institute of Dendrology PAS in Kórnik, we aimed at building a system in which high-quality graphic material is produced at relatively little cost. Here, I present our experience acquired during the first steps of digitisation, which we run as a part of the OZwRCIN project.

### Wstęp

W styczniu tego roku ukazał się najnowszy raport omawiający stan zbiorów zielnikowych na świecie, a dokładniej – tych zbiorów, które są rejestrowane w „Index Herbariorum”. Według tego raportu, sporządzonego przed dr Barbarę M. Thiers, dyrektor New York Botanical Garden Herbarium, zarejestrowanych jest na świecie 3095 zielników, które łącznie przechowują ponad 387,5 mln okazów (Thiers, 2019). Trzy największe: Muséum National d’Histoire Naturelle (Francja), The New York Botanical Garden (Stany Zjednoczone) oraz Royal Botanic Gardens (Wielka Brytania) posiadają odpowiednio: 8, 7,8 oraz 7 mln arkuszy. Czy to dużo? Zdecydowanie tak, jeśli zauważymy, że 32 polskie herbaria zgromadziły ich w sumie około 5,3 mln (l.c.).

Zbiory zasuszonych roślin w postaci arkuszy zielnikowych to jednak nie przybytek czy pozostałość po dawnych i minionych epokach rozwoju botaniki. Od roku 1960 herbaria powiększyły stan posiadania mniej więcej o 160% i trend ten się utrzymuje. Widać z tego wyraźnie, że zielniki

są cenionymi zbiorami naukowymi i bynajmniej nie tracą na znaczeniu. Z ubolewaniem odnotować można jednak przypadki, kiedy są one traktowane nieco po macoszemu i spychane na margines „plant sciences”; prawdopodobnie dzieje się tak z powodu niezrozumienia wagi danych, jakie są w stanie dostarczyć współczesnym badaczom.

Heberling, Prather i Tonsor (2019) w dopiero co opublikowanym opracowaniu przeanalizowali ponad 13 700 artykułów naukowych, w których wykorzystywano dane z herbariów w okresie ostatnich 95 lat. Z tych analiz wypływa szereg interesujących wniosków. Po pierwsze, liczba takich prac bardzo gwałtownie wzrosła w czasie tego stulecia, a przyrost ma charakter wykładniczy. Co ważne, tendencja wzrostowa podąża równoległe do ogólnego trendu w obrębie nauk o roślinach (wspomniane „plant sciences”). Oznacza to, że prace wykorzystujące zielniki jako źródło danych nie tracą na znaczeniu mimo rozwoju nowych technik badawczych i metod analiz, w tym zaawansowanych sposobów badań genetycznych, lecz że znajdują się w zasadniczym nurcie badań nad roślinami.

Co jeszcze ciekawego płynie z lektury tego opracowania? Przeanalizowawszy tysiące artykułów naukowych, badacze (l.c.) przypisali je do 22 obszarów tematycznych i prześledzili, jak ich udział zmieniał się w okresie od 1920 do 2017. W ten sposób otrzymali obraz dynamiki zmian wykorzystania danych z herbariów do celów naukowych. Niektóre z obszarów wyraźnie traciły na znaczeniu (liczonym jako udział procentowy w ogólnej liczbie analizowanych artykułów). Największy spadek odnotowały np. „historia botaników i zbiorów” czy „notki taksonomiczne”. Pewne obszary po okresie intensywnego rozkwitu czy wręcz świetności wyraźnie straciły ostatnio na znaczeniu („prace nad ploidalnością” czy „rozmişczenie gatunków”), podczas gdy inne odnotowują gwałtownie zwiększający się udział („analizy DNA”, „analizy bioróżnorodności”, „fitogeografia i dynamika zasięgów” i „procesy biologiczne a zmiany globalne”).

Udział prac z zakresu taksonomii roślin, jako jednego z obszarów w tych analizach, raczej się nie zmieniał na przestrzeni blisko stu lat. Systematyka to dziedzina tradycyjnie najsilniej związana z herbariami: to na zbiorach zielnikowych opierają się monografie i rewizje systematyczne. Także opisy nowych taksonów nierozzerwalnie wiążą się z wykorzystaniem zbiorów zielnikowych, ale tylko po części wypływa to z obowiązku deponowania okazu, który służył do opisanie nowego taksonu. Doniosłość danych badawczych w postaci takich zbiorów łatwiej zrozumieć, jeśli weźmie się pod uwagę najnowsze szacunki co do ciągle nieopisanego bogactwa gatunkowego. Analiza danych historycznych pozwala na oszacowanie, że wśród roślin okrytonasiennych jeszcze około 70 tysięcy gatunków nie zostało poznanych, przy czym ponad połowa z nich już znajduje się w zbiorach zielników na całym świecie i tam czeka na opisanie (Bebber et al., 2010).

Warto wspomnieć także o tym, że szeroko stosowane metody badań genetycznych rozwinęły się na tyle silnie, że przełamały trudności związane z poprawną interpretacją informacji zakodowanej w DNA pobranego z okazów zielnikowych, co więcej – techniki te mogą być już wykorzystywane do badania materiału z herbariów na skalę masową (Buerki & Baker, 2016). To oznacza, że wkroczyliśmy w etap prawdziwej rewolucji w tej materii, a jej znaczenie trudno przecenić, ponieważ badania takie umożliwią między innymi uwzględnienie danych genetycznych w dodatkowym wymiarze, a mianowicie w aspekcie historycznym. Korzystanie z okazów zielnikowych umożliwia także badanie taksonów rzadkich lub trudno dostępnych w naturze, a nawet takich, które wyginęły w stanie dzikim.

## **Digitalizacja – sposób na upowszechnienie danych zielnikowych**

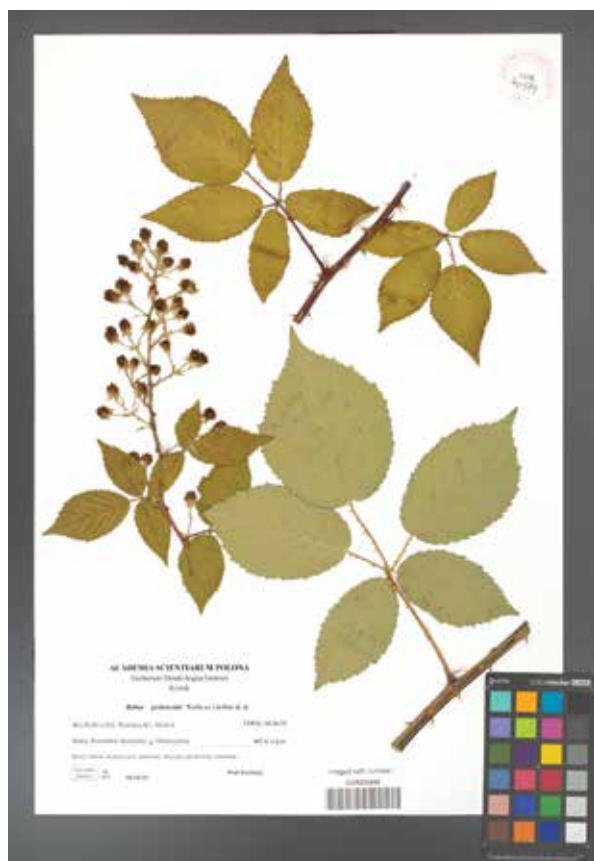
Bogactwo danych badawczych pochodzących ze zbiorów herbariów, czy to realnie wykorzystywanych już dziś, czy o potencjale naukowym, jest ogromne. Naukowcy dostrzegają potrzebę ułatwionego dostępu do tych zasobów i dlatego od kilkudziesięciu lat zbiory zielnikowe są poddawane digitalizacji. W powszechnym rozumieniu proces ten utożsamiany jest z upowszechnianiem zdjęć arkuszy zielnikowych

w Internecie. To jednak tylko część zjawiska. Digitalizację można bowiem rozumieć szerzej – jako wszelkie przetwarzanie informacji o okazach, w tym także informacji z etykiet zielnikowych, do formy cyfrowej. To oznacza, że udostępniane są nie tylko zdjęcia gromadzonych roślin, ale także obrazy etykiet zielnikowych; co niezwykle ważne – zapisane w nich dane i wszelkie informacje o okazie są dodatkowo umieszczone w bazie danych. W ten sposób zgromadzona dokumentacja może następnie służyć do dalszych badań.

Współczesne metody badawcze rozwijają się w kierunku analizy obrazu umożliwiającej pozyskiwanie danych o budowie roślin bezpośrednio ze zdjęć zielników, w tym pozwalającej na przetwarzanie zautomatyzowane (Corney et al., 2012a,b; Carranza-Rojas et al., 2017). Powstają także coraz lepsze narzędzia służące do identyfikacji gatunków (Carranza-Rojas et al., 2017). Jednocześnie dane o występowaniu roślin, jakie pozyskuje się z etykiet zielnikowych mogą służyć nie tylko do precyzyjnego i zakrojonego na szeroką skalę wykreślenia współczesnych zasięgów występowania gatunków. Ich potencjał tkwi także w przewidywaniu zmian w środowisku, jakie dopiero nastaną. Żyjemy w czasach zmian o bezprecedensowym tempie. Dotyczy to zarówno samego klimatu, jak i zmian wywołanych eksploatacją przyrody przez człowieka w szerszym rozumieniu (np. rozwój rolnictwa, zmiany stosunków wodnych, pozyskiwanie kopalin czy rozwój miast i sieci transportowych). To wpływa na przebudowę bioróżnorodności, również w skali globalnej. A zielniki gromadzą od setek lat dowody na istnienie tych zmian w postaci zakonserwowanych roślin, o których wiemy, kiedy i gdzie zostały zebrane. Znając historię i teraźniejszość, możemy trafniej oceniać kierunki zmian w przyszłości. Tym aspektem zajmuje się dział modelowania, który jednak musi być zasilany wiarygodnymi danymi. Te dane znajdują się właśnie w herbariach. Aby je stamtąd pozyskać, trzeba zielniki niejako „uwolnić” poprzez digitalizację.

Jak już wspominałem, myśląc o digitalizacji zbiorów zielnikowych, można wpaść w pułapkę, jeśli wyobrazić sobie ją będziemy jedynie jako wykonanie i udostępnienie cyfrowego obrazu danej rośliny. Nic bardziej błędnego: cyfryzacja taka byłaby nietrafiona i to na różnych poziomach. Po pierwsze, łatwo sobie wyobrazić, jak wyglądałby taki zbiór. Byłaby to kolekcja tysięcy czy nawet milionów zdjęć, w której odnalezienie czegokolwiek mogłoby odbywać się tylko na zasadzie przeszukiwania „jeden po drugim”. Owszem, podgląd byłby możliwy, a z właściwości pliku można by także uzyskać dane na temat daty jego wykonania (zdjęcie, nie zielnika), jego rozmiarów, rozdzielczości, skali barw itp. Są to jednak dane czysto techniczne i nie mówią nic o samym okazie przedstawionym na tym zdjęciu. Właściwa digitalizacja obejmuje połączenie wszelkich dostępnych danych o okazie (metadanych) z plikiem graficznym – dopiero ten tandem stanowi właściwy, oczekiwany owoc digitalizacji i dopiero on stanowi źródło informacji, które można wykorzystać w badaniach naukowych.

Jednak digitalizacja to nie tylko ułatwione pozyskanie danych do badań naukowych. Nie do przecenienia są także wartości czysto praktyczne. Z punktu widzenia zarówno osoby korzystającej ze zbiorów zielnikowych, jak i kuratorów



Ryc. 1. Przykłady zdjęć okazów zielnikowych otrzymanych w ramach digitalizacji herbarium Instytutu Dendrologii PAN

Fig. 1. Images of herbarium specimens which were obtained during digitisation in the Herbarium of the Institute of Dendrology, Polish Academy of Sciences

herbariów, bardzo ważnym aspektem jest możliwość właściwie bezkosztowego udostępniania okazów. Normalną praktyką jest wypożyczanie zbiorów do celów naukowych specjalistom z innych ośrodków. Odpowiednio zabezpieczone okazy są wówczas wysyłane zainteresowanym badaczom (a później odsyłane przez nich do macierzystego zielnika), co oczywiście wiąże się z kosztami. Niektóre herbaria niechętnie udostępniają swe kolekcje w ten sposób, materiały bowiem czasem wracają uszkodzone lub nawet giną. Wypożyczenia to właściwie ostateczność, milej widziane są osobiste wizyty zainteresowanych naukowców w herbariach. I choć fizyczny dostęp do materiałów czasem jest konieczny, to możliwość udostępniania zdjęć okazów jest olbrzymim ułatwieniem dla obu stron: jest tanie, szybkie i nie wiąże się z niebezpieczeństwem utraty czy zniszczenia zbiorów.

## Wybór metod

Przystępując do digitalizacji zbiorów zielnikowych, należy mieć na uwadze dwie podstawowe części tego procesu: wykonanie możliwie jak najlepszych technicznie zdjęć arkuszy zielnikowych oraz opatrzenie ich metadanymi, które będą niezbędne do zarządzania takim zbiorem obrazów, a jednocześnie będą przydatne do prac naukowych (choć i o funkcji edukacyjnej nie można zapominać). Oba zasoby informacji (obraz oraz przyporządkowane mu dane) należy połączyć w taki sposób, by jak najlepiej służyły użytkownikowi, a jednocześnie by sam proces digitalizacji można było prowadzić w rozsądnych ramach czasowych i finansowych.

Kompromis pomiędzy możliwościami finansowymi i ograniczeniami czasowymi a potrzebami użytkownika, także potencjalnymi, stanowi oś podejmowania decyzji co do typu digitalizacji. Użytkownik bez wątplenia chciałby uzyskać najpełniejszą informację o danym okazie (począwszy od nazwy rośliny, jej wieku, stanu rozwojowego, rozmiarów, cechach makro- i mikromorfologii, przez opis populacji i siedliska, w którym rosła, współwystępowaniu innych gatunków, daty i miejsca zbioru, w tym współrzędne geograficzne i wysokość nad poziomem morza, aż po dane osoby, która ją zebrała, autorze oznaczenia, historię identyfikacji na przestrzeni lat itd.). Wiele z tych informacji można odnaleźć bezpośrednio na etykiecie, która stanowi integralną część arkusza zielnikowego. Inne należałoby pozyskać, analizując sam okaz (pomiar, obserwacja w mikroskopie itp.). Jednak ich uzyskanie w takim zakresie, a następnie wprowadzanie do odpowiedniej bazy danych byłoby bardzo pracochłonne i czasochłonne, sprawiając, że wszelkie koszty byłyby nieracjonalne.

To samo dotyczy samego cyfrowego obrazu rośliny: może być on niezwykle wysokiej rozdzielczości, idealnie oświetlony, ukazujący najdrobniejsze szczegóły. Jednak wykonanie takiego perfekcyjnego zdjęcia wymaga najwyższej klasy sprzętu fotograficznego. A potem należałoby jeszcze zabezpieczyć dla niego miejsce na serwerach i szybkie łącza internetowe, by użytkownik mógł z nich wydajnie korzystać, co oczywiście generuje następne pokaźne koszty.

Kompromis między kosztami i jakością udostępnianych danych jest często zatem pojęciem o kluczowym znaczeniu: herbaria chciałyby digitalizować swoje zbiory w możliwie

najpełniejszy sposób, ale jednocześnie nie dysponują nieograniczonymi środkami i czasem na ten cel.

## Zaplecze techniczne

W procesie podejmowania decyzji o trybie digitalizacji uwzględnia się oczywiście aspekty techniczne. Pierwsze pytanie natury technicznej dotyczy wyboru sprzętu do uzyskiwania obrazów cyfrowych. Zasadniczo spotyka się dwa rozwiązania, tj. skanery oraz aparaty fotograficzne. Skanery pozwalają na uzyskiwanie obrazów bardzo wysokiej rozdzielczości otrzymywanych w stałych warunkach oświetlenia. Urządzenia te są coraz szybsze, więc dotychczasowy problem czasochłonnego skanowania obiektu traci stopniowo na znaczeniu. Nie bez znaczenia natomiast pozostają dwie inne cechy takiego rozwiązania. Pierwszy to niebagatelny koszt samego skanera, który wielokrotnie przewyższa wartość stosowania metody opartej na sprzęcie fotograficznym. Drugi – niewielka głębia ostrości, czyli dystans od urządzenia, w którym skanowany obiekt sprawia wrażenie ostrego. Wydawać by się mogło, że w przypadku roślin przechowywanych jako klasyczne arkusze zielnikowe nie powinno to mieć znaczenia, ponieważ zasadniczo są one dobrze sprasowane. Dotyczy to jednak delikatnych i niewielkich roślin zielnych. Co jednak z roślinami o zdrewniałych pędach, silnych cierniach, pokaźnych szyszkach, grubych i twardych owocach? Tak przecież wyglądają kolekcje roślin drzewiastych. W przypadkach zasuszonych roślin o grubych organach głębia ostrości skanera przestaje wystarczać dla wiernego oddania zawartości arkusza zielnikowego. Na świecie większość herbariów wybiera raczej rozwiązanie oparte na mniej czy bardziej klasycznych aparatach fotograficznych. Przy odpowiednim doborze parametrów można uzyskać głębię ostrości wystarczająco dużą dla takich okazów. Dodatkowo sprzęt fotograficzny, choć drogi sam w sobie, nie wymaga aż tak pokaźnych nakładów finansowych w porównaniu ze skanerami. Potrzeba natomiast większej wiedzy specjalistycznej (technicznej, fotograficznej) na etapie kompletowania elementów zestawu do obrazowania oraz przygotowania stanowiska pracy.

Na taki zestaw poza samym aparatem fotograficznym z odpowiednio dobranym obiektywem, składa się zwykle statyw, na którym umieszcza się aparat: powinien być odpowiednio wysoki, by objąć w kadrze cały arkusz i zapewnić właściwą głębię ostrości, a jednocześnie stabilny, by uniknąć efektu poruszenia wywołanego drganiami. Kluczowe znaczenie ma dobór właściwego oświetlenia fotografowanego obiektu. O ile skanery zapewniają równomierne oświetlenie na całej skanowanej powierzchni, o tyle w przypadku zestawów fotograficznych należy zastanowić się, jak dostarczyć wystarczającą ilość światła do arkusza zielnikowego (by móc uchwycić szczegóły morfologii organów) oraz by nie dopuścić do zaburzeń wywołanych światłem z zewnątrz (czy to promieni słonecznych, czy lamp z pomieszczenia); warunki oświetlenia powinny być stabilne. Nie chcę omawiać tu wszelkich szczegółów i możliwości, jakie oferuje współczesna technika, a które znajdują zastosowanie w herbariach; rozwiązań jest wiele i pozwalają na elastyczne podejmowanie decyzji o kształcie procesu.

Dotyczy to także drugiego, nie mniej istotnego aspektu digitalizacji zielników, a mianowicie oprogramowania do wprowadzania, przechowywania i organizacji danych o okazach. Przy wyborze rozwiązań należy brać pod uwagę takie cechy jak: (1) łatwość i szybkość wprowadzania informacji – jeśli to możliwe, warto stosować oprogramowanie OCR do zautomatyzowanego odczytu danych z etykiet; (2) elastyczność systemu, by dane mogły być zbierane w sposób najlepiej dopasowany do specyfiki kolekcji; (3) wykorzystanie przyjętych standardów formatu danych (ułatwia to późniejsze łączenie danych z informacjami z innych zielników); (4) łatwość i szybkość udostępniania danych w Internecie. Nie można zapominać o kosztach zakupu lub stworzenia oprogramowania oraz jego obsługi i wsparcia technicznego. Ważne również, by takie oprogramowanie mogło służyć do zarządzania zbiorami w danej jednostce. Na szczęście powstało na świecie wiele rozwiązań dostosowanych do potrzeb digitalizacji herbariów (np. Brahms czy Specify, by wymienić tylko te najszerszej znane).

### Digitalizacja zielnika Instytutu Dendrologii PAN

W 2018 r. rozpoczęliśmy w Instytucie Dendrologii PAN w Kórniku proces digitalizacji zbiorów herbarium w ramach projektu „Otwarte Zasoby w Repozytorium Cyfrowym Instytutów Naukowych – OZwRCIN” ([rcin.org.pl](http://rcin.org.pl)). Nasze rozwiązania techniczne opierają się na „wariacie

fotograficznym”, a programistyczne – na połączonych siłach własnej bazy danych oraz systemu dLibra (środowisko służące budowie bibliotek cyfrowych) stworzonego przez Poznańskie Centrum Superkomputerowo-Sieciowe.

Zbudowaliśmy dwa stanowiska do obrazowania arkuszy zielnikowych (ryc. 2). W podstawowym wariantcie stanowisko takie składa się z aparatu cyfrowego typu lustrzanka (Canon EOS 5D Mark IV, obiektyw Sigma 70 mm F 2.8 DG) umieszczonego na wysokim statywie reprodukcyjnym ze stabilną podstawą. U podnóża statywu znajduje się specjalnie zaprojektowana komora bezcieniowa, w której umieszczamy okaz zielnikowy. Komora ta wyposażona jest w środku w cztery profesjonalne lampy LED o ustalonej temperaturze barwowej światła. Zapewniają one silne światło padające z kilku stron, co ogranicza powstawanie cieni oraz gwarantuje w miarę równomierne oświetlenie obiektu. Najważniejszą jednak cechą tej konstrukcji jest to, że obiekt zostaje odizolowany od zewnętrznego światła, przez co w pomieszczeniu można pracować w standardowych warunkach oświetlenia biurowego, bez potrzeby jego zaciemnienia.

Aparat jest zamontowany i skonfigurowany tak, by nie było potrzeby jego indywidualnego ustawiania przy kolejnych okazach, co jest wielką zaletą przy masowej digitalizacji, kiedy zaoszczędzone sekundy czy minuty pomnożone przez setki i tysiące okazów dają się przeliczyć na tygodnie pracy. Rozwiązanie to pozwala jednocześnie uzyskiwać obrazy o identycznych parametrach i dużej powtarzalności oraz wierności odwzorowania.



Ryc. 2. Pomieszczenie digitalizacyjne z dwoma stanowiskami fotograficznymi

Fig. 2. The digitisation studio with two imaging stations

Aparat fotograficzny łączy się z komputerem przy pomocy kabla, dzięki czemu obraz jest zapisywany bezpośrednio na dysku twardym, bez potrzeby przenoszenia na dodatkowych nośnikach. Sam komputer jest w miarę standardowym sprzętem biurowym, zadaliśmy jednak, by zaopatrzony był w duży monitor, co ułatwia pracę ze zdjęciami, zwłaszcza że powstają one w dużych ilościach. Podłączony do niego jest także czytnik kodów paskowych (ryc. 3). Sercem systemu jest stworzona na potrzeby obsługi zielnika baza danych, do której wprowadzane są dane o okazie. Dopiero z tej bazy dane są eksportowane i następnie przekazywane do systemu zewnętrznego, który służy do ich upowszechniania w Internecie.

Sam sprzęt to jednak nie wszystko. Niezwykle ważną jest decyzja, jak powinna przebiegać organizacja pracy podczas digitalizacji. Istnieją pewne wypracowane wzorce postępowania i można z nich korzystać, uwzględniając własne zaplecze i potrzeby. Schematy oraz zalecenia dostępne są zarówno w literaturze (Tulig et al., 2012; Nelson et al., 2015; Harris & Marsico, 2017), jak i na stronach internetowych (np. idigbio.org). W Instytucie Dendrologii PAN wybraliśmy ścieżkę, w której poszczególne etapy pracy przebiegają następująco:

1. Etap wstępny. Obejmuje on przygotowanie okazów zielnikowych do digitalizacji. Są one przeglądane pod względem technicznym i merytorycznym. Przegląd techniczny polega na ocenie, czy okaz nadaje się do wykonania zdjęcia, czy roślina jest dobrze przymocowana i właściwie opisana. Jeśli zachodzi potrzeba, wykonywane są drobne naprawy arkusza, poprawiane jest klejenie, części luźne umieszczane są w kopercie. Specjalista sprawdza jednocześnie, czy roślina jest poprawnie oznaczona, a rażące błędy identyfikacji są poprawiane na bieżąco. Równoległe przeglądamy dostępne źródła w poszukiwaniu aktualnej nazwy danego taksonu, ponieważ należy pamiętać, że systematyka jest nauką, która ciągle zmienia nasz sposób patrzenia na klasyfikację roślin, więc identyfikacja okazu przeprowadzona kilkadziesiąt lat temu może nie odpowiadać obecnemu stanowi wiedzy.

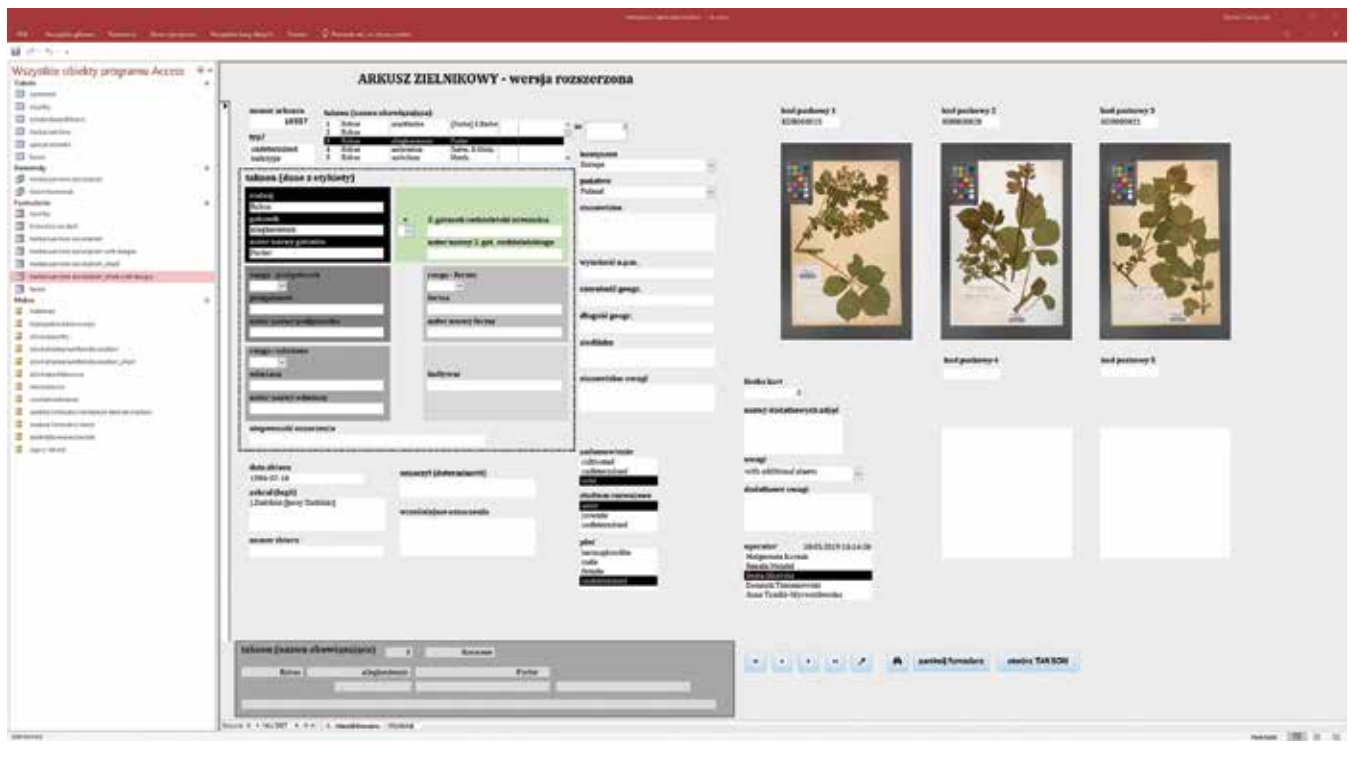
2. Właściwa digitalizacja. Dzielimy ją na dwie fazy. Pierwsza to wprowadzanie danych o okazie do bazy (ryc. 4). Informacje, które uznaliśmy za szczególnie istotne, obejmują np.: numer inwentarzowy, nazwę taksonu z etykiety, nazwę taksonu, którą można uznać za zgodną z aktualnym stanem wiedzy, datę zbioru, nazwisko zbieracza, nazwisko osoby, która oznaczyła okaz, a także miejsce zbioru. Z po-

zyskiwaniem danych na temat miejsca zbioru wiąże się poważna decyzja: czy należy iść szeroko i wprowadzać dane z etykiety najpełniej, jak się da, czy też je ograniczyć, by proces mógł przebiegać w miarę sprawnie. Trzeba brać pod uwagę fakt, że kiedyś etykiety wypisywano ręcznie, często charakterem trudnym do odczytania, a także że robiono to w różnych językach, także w alfabetach obcych w naszej części Europy (choć rozszyfrowywanie etykiety zapisanej po chińsku nie musiałoby być trudniejsze niż identyfikacja nazwy stanowiska zapisanej odręcznie nieczytelną cyrylicą). Uwzględniając te wszystkie czynniki, postanowiliśmy ograniczyć informacje o miejscu występowania zasadniczo do nazwy kraju (według obecnego podziału politycznego). Reszta informacji jest do odszukania na etykietce, którą w razie potrzeby można dowolnie powiększać na ekranie komputera, co ułatwia jej odczytanie.



Fig. 3. Czytnik kodów paskowych ułatwia pracę operatora

Fig. 3. The barcode reader helps an operator in managing the data



Ryc. 4. Dane o okazie gromadzone są w bazie danych stworzonej na potrzeby projektu  
 Fig. 4. Data regarding each specimen are gathered in a database designed for the project

Bardzo istotne od samego początku było dla nas, by gromadzone informacje mogły być w przyszłości łatwo przyswajane przez rosnące w siłę oddziaływania agregatory danych. Chodzi o to, że herbaria digitalizujące swoje kolekcje udostępniają dane systemom, które łączą je w większe aglomeraty, przez co stają się cenniejsze dla użytkowników, którzy na przykład potrzebują dużych zbiorów danych do analiz bioróżnorodności. By takie systemy mogły być bezproblemowo zasilane takimi danymi, te powinny spełniać pewne standardy. Jednym z nich jest ujednolicony sposób zapisu. Dla przykładu, poprawny odczyt daty zbioru okazu może stanowić wyzwanie dla programu, jeśli w różnych ośrodkach będzie stosować się odmienne formaty (np. 12 maja 2001, 12/05/2001, 12/5/2001, 5/12/2001, 12-05-2001, 2001-05-12, 12.05.01 itd., itp.). Na szczęście opracowano pewne standardy, które dobrze sprawdzają się do danych biologicznych, do takich należy Darwin Core czy ISO. W naszych rozwiązaniach informatycznych stosujemy je, licząc, że w przyszłości pozwoli to na łatwe „dogadywanie się” systemów.

Wróćmy jednak do arkusza zielnikowego. Na etapie wprowadzania informacji do bazy danych każda karta zielnikowa zaopatrywana jest w unikatowy kod paskowy. Nie jest on tożsamy z numerem inwentarzowym w ramach całego zielnika. Zarówno kod paskowy, jak i numer arkusza, trafiają do bazy danych. Wartość tekstowa z kodu paskowego staje się nazwą zdjęcia danej karty; w ten sposób zdjęcie i metadane tworzą całość.

W kórnickim herbarium postanowiliśmy prowadzoną od dziesięcioleci numerację arkuszy prowadzić niezależnie od kodów paskowych. W niektórych zielnikach, gdzie wprowadzono system kodów, wybrano rozwiązanie polegające na

tym, że drukowane są kody o numeracji tożsamej z dotychczasową. Ma to swoje zalety – w ten sposób unika się funkcjonowania dodatkowych oznaczeń. My wybraliśmy inne rozwiązanie, ponieważ uznaliśmy, że oba systemy numeracji będziemy mogli skutecznie powiązać w bazie danych. W naszym herbarium spory odsetek okazów reprezentowanych jest przez kilka kart zielnikowych w ramach tego samego numeru inwentarzowego. Dzieje się tak dlatego, że drzewa czy krzewy często wymagają rozdzielania organów z tej samej rośliny na dwie, trzy, nawet więcej pojedynczych kart zielnikowych. Niezwykle często dotyczy to jeżyn (*Rubus*), które wymagają zbioru pędów zarówno wegetatywnych, jak i generatywnych, bo w ten sposób można dysponować kompletem danych morfologicznych potrzebnych do właściwej identyfikacji. Poza tym stosując unikatowe kody niezależne od dotychczasowej numeracji, ułatwiliśmy sobie pracę z ich tworzeniem – drukujemy je szybko w długich seriach. W ten oto sposób nie ma potrzeby drukowania pojedynczych kodów, których wartości należałoby wprowadzać do programu ręcznie podczas digitalizacji konkretnego arkusza – spowalniałoby to pracę, a jednocześnie stanowiłoby źródło błędów, ponieważ wpisywanie zawsze wiąże się z ryzykiem pomyłki.

Po wprowadzeniu danych z określonej partii okazów (zwykle kilkudziesięciu) arkusze te są fotografowane. Dzięki temu, że konfiguracja sprzętu i oprogramowania została już ustalona i działa w oparciu o zadane parametry, ta faza przebiega bardzo szybko (ok. 50–60 s na zdjęcie). Fotografie trafiają najpierw do programu Digital Photo Professional, gdzie nadane zostają im nowe nazwy: operator skanuje czytelnikiem kod paskowy widoczny na monitorze komputera na danym zdjęciu. Dzięki temu zabiegowi każdy plik graficzny

otrzymuje nazwę identyczną z kodem paskowym karty zielnikowej, a odbywa się to bez potrzeby wpisywania go z klawiatury, co wiązałoby się z niebezpieczeństwem pomyłek, o czym już pisałem.

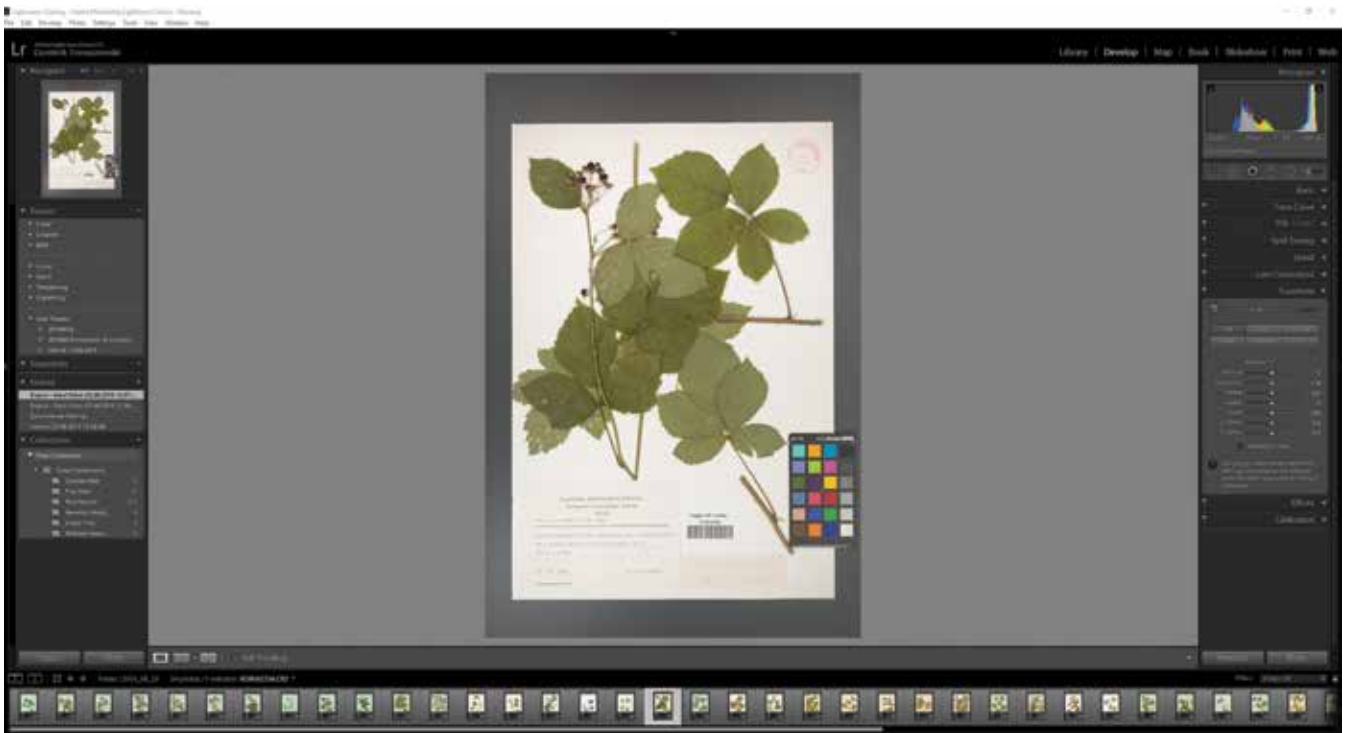
Zdjęcia „surowe” (RAW) o ustalonych w ten sposób nazwach przechodzą następnie proces przetwarzania, który można porównać do wywoływania w ciemni (stąd nazwa programu: Adobe Photoshop Lightroom). Zostaje zastosowany profil kolorów i obiektywu, są lekko wyostrzane, prostowane i przycinane. Co ważne, wszystkie pliki zostają potraktowane dokładnie w ten sam sposób i zbiorczo (ryc. 5). Następnie zostają przekonwertowane do formatów docelowych i zapisane jako pliki w formatach JPG oraz DNG. Pierwszy będzie służył do prezentacji w Internecie (mniejszy rozmiar dzięki kompresji), natomiast drugi do przechowywania długoterminowego.

Arkusze po sfotografowaniu wracają na swoje miejsce w zielniku, na czym kończy się właściwa digitalizacja. To nie koniec jednak pracy z danymi. Informacje zgromadzone w bazie danych są jeszcze sprawdzane i w miarę potrzeb korygowane. To samo dotyczy zdjęć, gdy okazuje się, że mają jakiś defekt (na przykład liść przysłania fragment etykiety zielnikowej, przez co pełna informacja jest niedostępna dla użytkownika; w takim przypadku albo zmieniamy ułożenie liścia, albo przepisujemy zawartość etykiety do bazy danych, by dostarczyć te informacje użytkownikowi w inny sposób).

Digitalizacja nie byłaby pełnowartościowa, gdyby jej wyniki nie stały się dostępne szerokiemu gronu użytkowników, czy to naukowców, czy innym zainteresowanym. W przypadku herbarium Instytutu Dendrologii upublicznienie zasobów odbywać się będzie poprzez portal konsorcjum, któ-

re realizuje projekt digitalizacji zasobów polskich instytutów naukowych, tj. Repozytorium Cyfrowe Instytutów Naukowych (rcin.org.pl). Stosuję tu celowo czas przyszły, ponieważ obecnie weszliśmy w końcową fazę testów technicznych transferu oraz sposobów prezentacji danych (ryc. 6). Wkrótce pierwsze partie danych o drzewach i krzewach z kórnickiego herbarium będą dostępne wszystkim zainteresowanym. W pierwszej kolejności planujemy podzielić się naszymi zbiorami okazów jeżyn (*Rubus*), ponieważ ta kolekcja należy do najcenniejszych w tej części Europy. Warto tu wspomnieć, że herbarium Instytutu Dendrologii PAN (akronim KOR) zgromadziło ponad 54 tysiące okazów zielnikowych roślin drzewiastych. Wśród najważniejszych są te, które reprezentują flory krajów basenu Morza Śródziemnego i obszarów przyległych. Zbiory z terenu Polski to oczywiście trzon kolekcji, przy czym bardzo duży odsetek to rośliny z rodzajów, których systematyka i identyfikacja zawsze stanowiły wyzwanie dla botaników (wierzby, róże i, rzecz jasna, jeżyny). Spora część okazów zainteresuje także osoby zajmujące się odmianami uprawnymi, w tym wiele kultywarów, które rosną lub rosły na terenie kórnickiego arboretum.

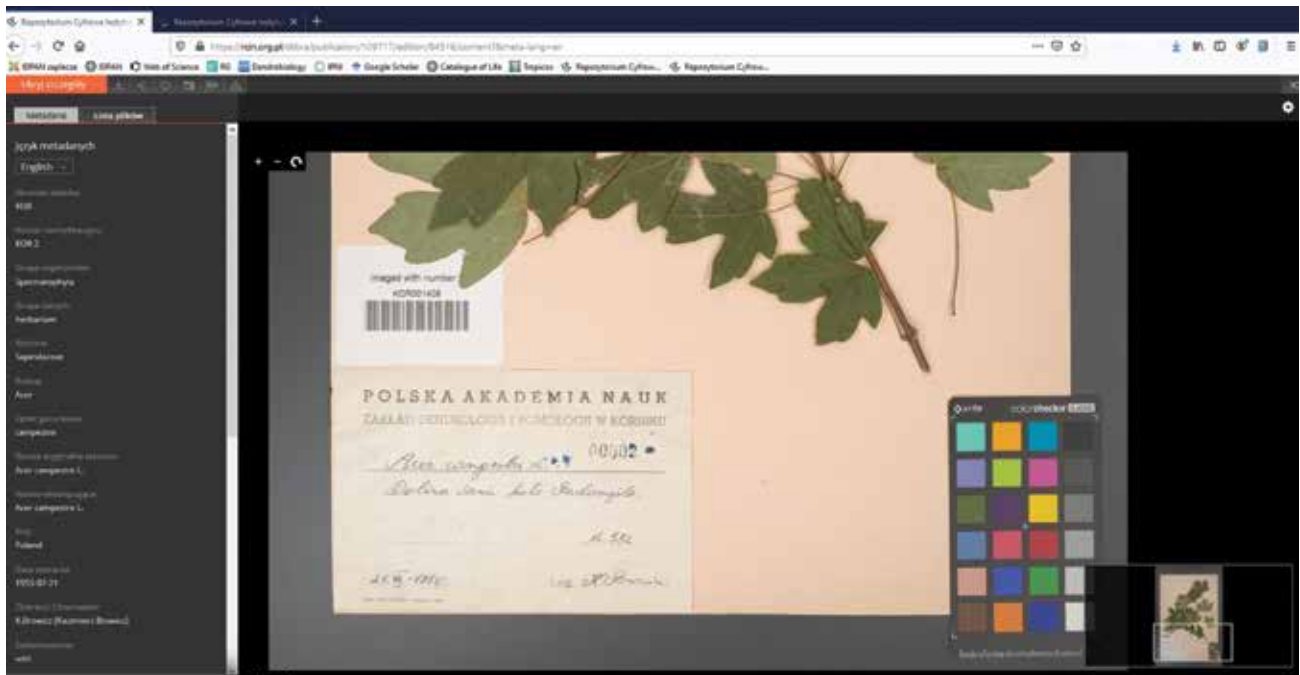
Mamy nadzieję, że ułatwiony dostęp do obrazów roślin pozwoli także nam, dzięki pomocy użytkowników zbiorów, wypełnić część białych plam w naszych zbiorach, np. uzyskamy w ten sposób pomoc w oznaczeniu niektórych roślin, choćby tych zebranych w tropikach. Jesteśmy pewni, że digitalizacja zasobów zielnikowych Instytutu Dendrologii PAN przyczyni się do rozwoju studiów nad roślinami drzewiastymi nie tylko w Polsce, a nasze zbiory przyciągną tych, którzy interesują się drzewami oraz krzewami i poszukują wiarygodnych informacji zarówno o ich wyglądzie, biologii, jak i występowaniu.



Ryc. 5. Zdjęcia z aparatu trafiają bezpośrednio do komputera, gdzie są poddawane zbiorczej edycji

Fig. 5. Images are transferred directly from the camera to the computer, where they are subsequently batch-edited





Ryc. 6. Prezentacja danych o okazie zielnikowym na portalu Repozytorium Cyfrowego Instytutów Naukowych  
 Fig. 6. Presentation of data on a herbarium specimen at the website of the Digital Repository of Scientific Institutes

## Podziękowania

Digitalizacja zielnika Instytut Dendrologii PAN w Kórniku jest częścią projektu „Otwarte Zasoby w Repozytorium Cyfrowym Instytutów Naukowych – OZwRCIN”, finansowanego przez Program Operacyjny Polska Cyfrowa Unii Europejskiej w ramach Działania 2.3 „Cyfrowa dostępność i użyteczność sektora publicznego”, poddziałanie 2.3.1 „Cyfrowe udostępnianie informacji sektora publicznego ze źródeł administracyjnych i zasobów nauki”.

Serdecznie dziękuję osobom zaangażowanym w digitalizację zbiorów zielnikowych, Paniom: Małgorzacie Łuczak, Renacie Mendel, Beacie Sikorskiej i dr Annie Tomlik-Wyremblewskiej, a także Prof. Jerzemu Zielińskiemu, który udziela nam wsparcia merytorycznego oraz dzieli się doświadczeniem i bezcennymi radami.

## Literatura

- Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB, Scotland RW. 2010. Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences* 107:22169–22171 DOI: 10.1073/pnas.101184108.
- Buerki S, Baker WJ. 2016. Collections-based research in the genomic era. *Biological Journal of the Linnean Society* 117:5–10 DOI: 10.1111/bj.12721.
- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017. Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology* 17:181 DOI: 10.1186/s12862-017-1014-z.
- Corney DPA, Clark JY, Tang HL, Wilkin P. 2012a. Automatic extraction of leaf characters from herbarium specimens. *Taxon* 61:231–244.
- Corney DPA, Tang HL, Clark JY, Hu Y, Jin J. 2012b. Automating digital leaf measurement: The tooth, the whole tooth, and nothing but the tooth. *PLoS ONE* 7:e42112 DOI: 10.1371/journal.pone.0042112.
- Harris KM, Marsico TD. 2017. Digitizing specimens in a small herbarium: A viable workflow for collections working with limited resources. *Applications in Plant Sciences* 5:1600125 DOI: 10.3732/apps.1600125.
- Heberling JM, Prather LA, Tonsor SJ. 2019. The changing uses of herbarium data in an era of global change: An overview using automated content analysis. *BioScience* 69:812–822 DOI: 10.1093/biosci/biz094.
- Nelson G, Sweeney P, Wallace LE, Rabeler RK, Allard D, Brown H, Carter JR, Denslow MW, Ellwood ER, Germain-Aubrey CC, Gilbert E, Gillespie E, Goertzen LR, Legler B, Marchant DB, Marsico TD, Morris AB, Murrell Z, Nazaire M, Neefus C, Oberreiter S, Paul D, Ruhfel BR, Sasek T, Shaw J, Soltis PS, Watson K, Weeks A, Mast AR. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applications in Plant Sciences* 3:1500065 DOI: 10.3732/apps.1500065.
- Thiers BM. 2019. The world's herbaria 2018: A summary report based on data from Index Herbariorum. Available at [http://sweetgum.nybg.org/science/docs/The\\_Worlds\\_Herbaria\\_2018.pdf](http://sweetgum.nybg.org/science/docs/The_Worlds_Herbaria_2018.pdf) (dostęp 20.10.2019).
- Tulig M, Tarnowsky N, Bevans M, Kirchgessner A, Thiers B. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209:103–113 DOI: 10.3897/zookeys.209.3125.