

# DZIEDZICTWO MUZYKI POLSKIEJ – DIGITALIZACJA, OPRACOWANIE I PERSPEKTYWY BADAWCZE

Do niniejszej publikacji tekst został przyjęty w trybie *double blind review process*.

## Wprowadzenie

W badaniach dużych korpusów dzieł muzycznych – niezależnie od tego, czy zdefiniujemy je jako zbiór kompozycji danego kompozytora, twórczość uprawianą w określonym ośrodku, czy w konkretnym okresie – istotna jest możliwie reprezentatywna ilość odpowiednio przygotowanego materiału do analizy. Z tego właśnie powodu od lat podejmowane są próby budowania dużych baz dzieł muzycznych zapisanych w formie cyfrowej. Dotyczy to zresztą nie tylko materiałów nutowych, ale także zawierających je źródła i informacji o nich. Znakomitym przykładem jest tworzona od przeszło 60 lat baza RISM<sup>1</sup>, w której znajdziemy nie tylko dane dotyczące źródeł, jego proveniencji, datowania i in., ale także – co szczególnie ważne – incipity muzyczne. Nie sposób przecenić roli RISM w badaniach źródłoznawczych, a w szczególności przydatności tego narzędzia w ustalaniu atrybucji utworów przekazanych anonimowo. Baza RISM ma jednak także spore ograniczenia, które powodują, że nie pozwala wykorzystać wszystkich możliwości, jakie dają technologie stworzone do prowadzenia badań z wykorzystaniem analizy komputerowej. Konieczne jest powstanie możliwie szerokiej bazy nie tylko cyfrowych odwzorowań źródeł (skanów) połączonych odpowiednio z metadanymi (najlepiej możliwie szeroko powiązanych z innymi danymi na zasadzie *linked open data*), ale także zapisanych w formacie symbolicznym, tj. w postaci transkrypcji nutowych<sup>2</sup>. Baza taka do celowo powinna pełnić funkcję tzw. *big data* dla zasobu źródeł muzycznych znajdujących się na terenie Polski, co pozwoli nie tylko w pełni wykorzystać potencjał istniejącego już oprogramowania do skomputeryzowanej analizy muzycznej, ale w przyszłości umożliwi wykorzystanie narzędzi sztucznej inteligencji w badaniach źródłoznawczych. Otwiera to zupełnie nowe perspektywy, do tej pory zupełnie niedostępne muzykologom. Warto w tym kontekście zwrócić uwagę nie tylko na podnoszoną wielokrotnie kwestię atrybucji, ale także na możliwość wychwycenia konkordancji między źródłami, tworzenia swoistych modeli dla analizy stylokrytycznej czy też mapowania źródeł, by zobrazować przepływy repertuaru pomiędzy ośrodkami. Osiągnięcie powyższych celów możliwe będzie jedynie dzięki realizacji określonych działań, po pierwsze możliwie wysokiej jakości digitalizacji, następnie zebraniu najwyższej jakości metadanych<sup>3</sup>, powiązaniu danych i metadanych oraz ich publikacji, a wreszcie

1 Zob. <https://opac.rism.info/index.php?id=4> oraz <https://rism.online/> [dostęp: 18 X 2023].

2 Na temat formatów stosowanych do cyfrowego zapisu notacji muzycznej zob. Marcin Konik, *Repertuar muzyki polskiej wieku XIX – wydania cyfrowe i analiza skomputeryzowana*, „Studia Chopinowskie” 2018, nr 1, s. 24–36.

3 Wysoka jakość metadanych niekoniecznie musi oznaczać szczegółowe naukowe opracowanie konkretnego źródła – to zadanie w zasadzie pojawia się na końcu procesu. W przypadku metadanych RISM kluczowym czynnikiem decydującym o jakości jest możliwość szerokiego wykorzystanie słownictwa kontrolowanego, zastosowanie identyfikatorów pozwalających łączyć dane i osadzić je w szerszym kontekście, a także stworzenie takiego opisu źródła, który zawiera możliwie pełną informację o nim – m.in. autorstwo (o ile jest znane), datowanie czy proveniencje.

sporządzeniu cyfrowej transkrypcji. Ten ostatni element z uwagi na wygodę użytkownika powinien uwzględniać także odniesienie transkrypcji cyfrowej do źródła z wykorzystaniem technologii IIF<sup>4</sup> (o tym aspekcie mowa będzie w dalszej części artykułu).

Publikacja źródeł w sposób, o którym mowa powyżej (z uwzględnieniem różnych narzędzi informatycznych i analitycznych), pozwala zbudować nowy rodzaj warsztatu krytyki źródła dysponujący zupełnie nowymi możliwościami. Istotą zmiany jest nie tylko publikacja edycji cyfrowej, która pozwala na zastosowanie rozwiązań z oczywistych względów niemożliwych do uzyskania w tradycyjnej publikacji papierowej lub statycznie wyrenderowanym pliku np. w formacie PDF, ale nade wszystko algorytmizacja procesu edycji. Algorytmizacja ta umożliwi nie tylko jego przyspieszenie, ale także uniknięcie ryzyka wielu błędów, jakie mogą wystąpić w efekcie stosowania metod tradycyjnych. Jednak bodaj największą korzyścią jest jednocześnie tworzenie bazy materiału analitycznego, ponieważ wszystkie opisane poniżej typy źródłowych edycji cyfrowych<sup>5</sup> stanowią mogą podstawę do badań i analizy.

W artykule niniejszym zostaną omówione kolejno proces tworzenia kompleksowej bazy źródeł wraz z ich cyfrowymi transkrypcjami oraz wynikające stąd nowe możliwości i perspektywy badawcze. Wszystkie wskazane przykłady, a także charakterystyka stosowanej metodologii prac oparte są na doświadczeniach Narodowego Instytutu Fryderyka Chopina w realizacji dwu projektów finansowanych ze środków Unii Europejskiej: „Dziedzictwo Chopinowskie w otwartym dostępie” oraz „Dziedzictwo muzyki polskiej w otwartym dostępie”<sup>6</sup>. Projekty te realizowane były odpowiednio w latach 2017–2021 i 2018–2022. W efekcie powstały dwa portale<sup>7</sup> oraz odrębne strony, na których dostępne są transkrypcje cyfrowe<sup>8</sup>. Co istotne, wszystkie dane wytworzone w obu projektach udostępniane są w domenie publicznej i mogą być wykorzystywane bez dodatkowych ograniczeń – dotyczy to także kodów źródłowych stworzonego oprogramowania.

## Kolekcje

Podczas realizacji wspomnianych projektów opracowane zostały częściowo lub w całości kolekcje muzykaliów (oraz materiałów okłomuzycznych) z różnych archiwów i bibliotek polskich. W pierwszej kolejności w ramach projektu POPC-1 przygotowano zbiory Instytutu Chopina (kolekcja Muzeum Fryderyka Chopina oraz zbiory biblioteczne). Udostępniono niemal 40 tys. w większej części niepublikowanych wcześniej online materiałów, w tym unikalną kolekcję XIX-wiecznych czasopism, zbiory ikonograficzne i fotograficzne, archiwalne nagrania oraz kolekcję wczesnych wydań dzieł Chopinowskich. Szczególnie istotne w kontekście badań źródłoznawczych było wprowadzenie do bazy RISM kolekcji

4  
Zob. <https://iif.io/> [dostęp: 18 X 2023].

5  
W projektach cyfrowych Narodowego Instytutu Fryderyka Chopina przyjęto za instrukcją wydawniczą trzy podstawowe typy edycji cyfrowej: 1) dyplomatyczną, 2) wykonawczą, 3) krytyczną. Podstawowe założenia edycji wskazanych typów zostaną omówione w dalszej części artykułu.

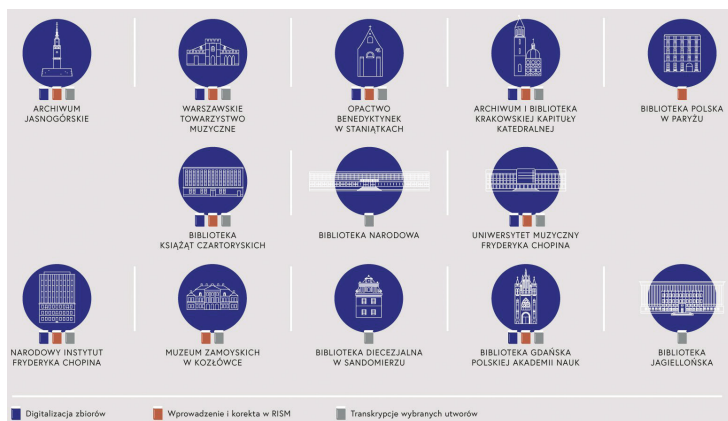
6  
Na potrzeby niniejszego tekstu projekty oznaczamy skrótami – odpowiednio „Dziedzictwo Chopinowskie w otwartym dostępie” jako POPC-1 oraz „Dziedzictwo muzyki polskiej w otwartym dostępie” jako POPC-2. Obydwa finansowane były w ramach Programu Operacyjnego Polska Cyfrowa.

7  
Zob. <https://chopin.musicsources.pl/pl> oraz <https://polish.musicources.pl/pl> [dostęp: 18 X 2023]. Obydwa portale docelowo mają zostać całkowicie zintegrowane (prace nad tym trwają).

8  
Zob. <https://chopinscores.org/> oraz <https://polishscores.org/> [dostęp: 18 X 2023].

zdigitalizowanych pierwodruków, które stały się podstawą do stworzenia transkrypcji cyfrowych.

Projekt POPC-2<sup>9</sup> objął w sumie kilkanaście kolekcji, z czego dziewięć w maksymalnym zakresie przyjętym w założeniach. Ustalono trzy możliwe poziomy opracowania polegające na: 1) digitalizacji zbiorów, 2) tworzeniu lub korekcie istniejących metadanych RISM oraz 3) przygotowaniu transkrypcji cyfrowych. Na przykładzie 1 przedstawiono uwzględnione kolekcje wraz ze wskazaniem rodzaju wykonanych prac.



Przykład 1. Instytucje objęte projektem POPC-2

W kilku przypadkach (np. Biblioteki Narodowej czy Biblioteki Diecezjalnej w Sandomierzu) prace ograniczone były do sporządzenia transkrypcji na bazie już uprzednio udostępnionych skanów lub wyłącznie do stworzenia metadanych RISM (np. Biblioteka Polska w Paryżu). Odnosząc się w dalszej części do konkretnych kolekcji, dla ułatwienia używać będziemy przypisanych im sigłów RISM<sup>10</sup>. Warto odnotować, że większość kolekcji objętych projektem nie była wcześniej nawet w najmniejszym zakresie udostępniona w sieci, a w przypadku niektórych brakowało nawet kompletnych katalogów dających pojęcie o zawartości zbioru.

## Digitalizacja

W projekcie POPC-2 udało się zdigitalizować ponad 25 tys. źródeł muzycznych, niejednokrotnie trudno dostępnych fizycznie. W zasadzie jedynym ograniczeniem kompletności prac był stan zachowania, czyli przypadki, w których z uwagi na zalecenia konserwatorskie digitalizacja nie mogła zostać przeprowadzona. W projektach NIFC przyjęto za wytycznymi obowiązującymi w Instytucie najwyższe standardy digitalizacji. W szczególności oznacza to, że stosowana rozdzielczość skanów wynosiła minimalnie 600 dpi przy 24-bitowej

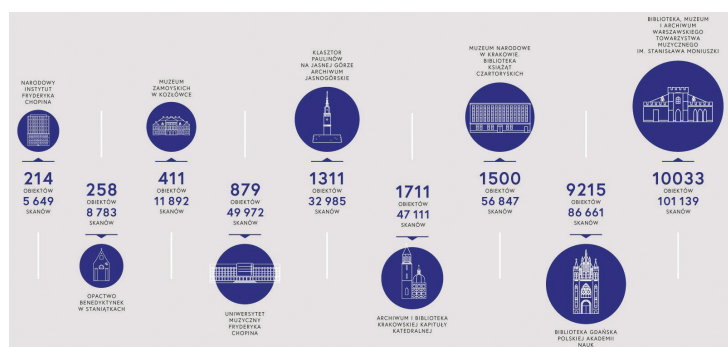
9

Pod poniższymi linkami można pobrać broszury informacyjne w języku polskim [https://polishscores.org/info/booklet/popc-2022\\_pl.pdf](https://polishscores.org/info/booklet/popc-2022_pl.pdf) oraz angielskim [https://polishscores.org/info/booklet/popc-2022\\_en.pdf](https://polishscores.org/info/booklet/popc-2022_en.pdf) [dostęp: 18 X 2023].

10

Poniżej zestawiono wszystkie sigła używane w projekcie POPC-2: **PL-CZ** – Archiwum oo. Paulinów na Jasnej Górze, **PL-GD** – Biblioteka Gdańska Polskiej Akademii Nauk, **PL-Kc** – Muzeum Narodowe w Krakowie – Biblioteka Książąt Czartoryskich, **PL-Kk** – Archiwum i Biblioteka Krakowskiej Kapituły Katedralnej, **PL-KOZmzk** – Muzeum Zamoyskich w Kozłowiec, **PL-STAb** – Biblioteka Opactwa św. Wojciecha ss. Benedyktynek w Staniątkach, **PL-Wnifc** – Narodowy Instytut Fryderyka Chopina, **PL-Wtm** – Biblioteka Warszawskiego Towarzystwa Muzycznego im. Stanisława Moniuszki, **PL-Wumfc** – Biblioteka Główna Uniwersytetu Muzycznego Fryderyka Chopina. Obok powyższych kolekcji znalazły się też udostępnione online zbiory, na podstawie których przygotowano transkrypcje cyfrowe – są to: **PL-Wn** – Biblioteka Narodowa, **PL-SA** – Biblioteka Diecezjalna w Sandomierzu.

głębi kolorów, a w uzasadnionych przypadkach nawet z 48-bitową głębią. W efekcie większość wytworzonych w procesie digitalizacji plików (zapisanych w niekompresowanym formacie TIFF) ma znaczne rozmiary (od 200 MB do nawet 700 MB). Całość wytworzonego materiału cyfrowego (skany wraz z tzw. derywatami<sup>11</sup>) to w sumie ponad 200 TB danych. Oczywiście w celu umożliwienia wygodnego korzystania z portalu domyślnie wyświetlane są niższej jakości skompresowane pliki w formacie JPEG, jednak dla każdego obiektu możliwe jest pobranie plików w trzech formatach: pojedynczego skanu (strony) w formacie skompresowanym JPEG, całego obiektu w formacie PDF oraz pliku źródłowego w formacie TIFF. Statystyka efektów prac digitalizacyjnych została ujęta na przykładzie 2<sup>12</sup>.



Przykład 2. Statystyka prac digitalizacyjnych w projekcie POPC-2

Pliki archiwizacyjne wraz z derywatami przechowywane są w wewnętrznym repozytorium NIFC, opartym na systemie Islandora<sup>13</sup>, który bazuje na platformie Fedora wraz z wbudowanym zaawansowanym silnikiem indeksowania i przeszukiwania Solr<sup>14</sup>. Repozytorium stanowi centrum, z którego pobierane są do portalu pliki prezentacyjne i źródłowe, a także identyfikatory, które umożliwiają pobranie i automatyczną aktualizację metadanych poprzez API RISM. Przykład 3 ukazuje ogólny schemat przepływu danych w projekcie.

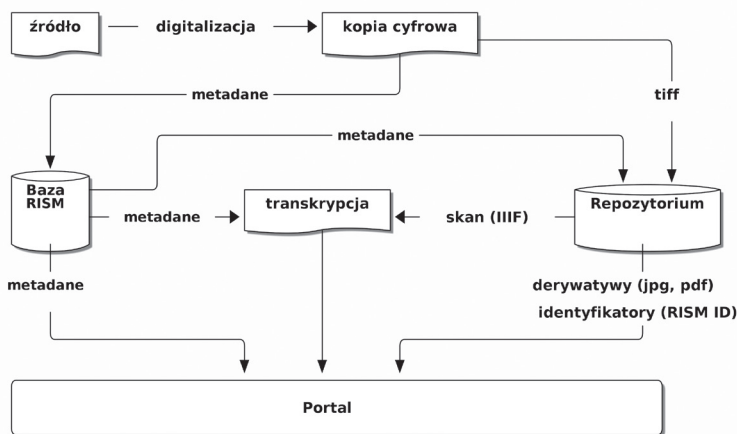
Przyjęta metodologia pozwala osiągnąć maksymalną ekonomię pracy oraz utrzymania całości systemu. Zwróćmy uwagę, że elementem, który może podlegać najczęstszym zmianom, są metadane RISM. Wynika to z faktu, że dzięki rozrastającej się bazie możliwe są np. nowe odkrycia dotyczące autorstwa, datowania źródeł etc. Ponieważ (jak widać na schemacie 3) do każdego z komponentów końcowych (jak repozytorium, portal czy plik zawierający transkrypcję) metadane pobierane są bezpośrednio z bazy RISM, wystarczy korekta w jednym miejscu, aby aktualne dane pojawiły się w każdym ze wspomnianych komponentów. Rozwiązanie powyższe ma jeszcze jedną wartość dodaną, wynikającą z tego, że baza RISM

11 Derywaty to pliki pochodne tworzone automatycznie w procesie archiwizacji w systemie repozytoryjnym. Podczas umieszczania plików źródłowych (nieskompresowanych plików TIFF) skrypty automatycznie tworzą pochodne pliki w formacie JPEG, JPEG2000 czy PDF.

12 Stan na dzień zakończenia projektu tj. 31 VIII 2022.

13 Zob. <https://wiki.lyrasis.org/display/ISLANDORA714/Introduction> [dostęp: 18 X 2023].

14 Zob. <https://solr.apache.org> [dostęp: 18 X 2023].



Przykład 3. Ogólny schemat przepływu danych

i przechowywane w niej metadane są aktualizowane i korygowane przez bardzo dużą, międzynarodową społeczność badaczy. W efekcie, jeśli osoba specjalizująca się w XIX-wiecznym drukarstwie muzycznym na podstawie analizy egzemplarza druku znajdującego się w zbiorach np. Staatsbibliothek w Berlinie dokona korekty metadanych, zostaną one automatycznie i bez konieczności jakiegokolwiek ingerencji ze strony NIFC zaktualizowane w portalu (oczywiście dotyczy to informacji odnoszących się do publikacji jako takiej, a nie konkretnego egzemplarza – może to być datowanie czy osoby powiązane, ale już nie sygnatura czy cechy fizyczne danego źródła).

Z uwagi na zastosowane standardy i wytyczne możliwe jest dalsze wykorzystanie wytworzonych w projektach plików w profesjonalnych publikacjach, wszystkie obiekty skanowane były bowiem z użyciem próbnika koloru i należycie skalibrowanym sprzętem. Przyjęto zasadę, że w przypadku materiałów rękopiśmiennych próbnik znajduje się na każdym skanie, w przypadku druków umieszczany jest zaś na ostatnim skanie w celu umożliwienia skalibrowania koloru wydruku dla całości obiektu. Pliki źródłowe sporządzone i udostępnione są w taki sposób, że umożliwiają nie tylko wydruk czy inną formę publikacji, ale także tworzenie na podstawie ich analizy opisu fizycznego źródła. Gwarantuje to właśnie obecność próbnika koloru, który dzięki umieszczonej na nim miarce pozwala odczytać informację o wymiarach obiektu. Skany mają jednak tę właściwość, że dzięki zastosowaniu odpowiedniego oprogramowania (np. darmowego programu Gimp<sup>15</sup>) można dokonywać bardzo precyzyjnych pomiarów na obiekcie cyfrowym, mając gwarancję, że wyniki odzwierciedlają faktyczne cechy źródła fizycznego. O ile ma to mniejsze znaczenie w przypadku ogólnej charakterystyki źródła (np. w odniesieniu do jego wymiarów), o tyle w przypadku bardziej precyzyjnej analizy może przynieść badaczowi istotną nową wiedzę. Jest tak m.in. w przypadku pomiarów rozmiaru czcionek czy tzw. światła

15  
Zob. <https://www.gimp.org/> [dostęp: 18 X 2023].

między systemami partytury, co może mieć istotne znaczenie dla datowania źródeł XIX-wiecznych (dane takie pomagają określić, z którym kolejno nakładem mamy do czynienia). Analiza obrazów cyfrowych, stanowiących odwzorowanie obiektów fizycznych, to prężnie rozwijająca się, zwłaszcza Stanach Zjednoczonych, nowa dziedzina badań<sup>16</sup>. Szczególnie obiecujące wydają się w tym aspekcie innowacyjne projekty mające na celu stworzenie z użyciem narzędzi SI oprogramowania pozwalającego już nie tylko na automatyczne rozpoznawanie tekstu na podstawie skanowanych druków, ale także zapisów rękopiśmiennych, w tym nutowych.

## Metadane RISM

Metadane dla wszystkich typów źródeł, jakie opracowywane są w bazie RISM, tworzono właśnie w tej bazie, wychodząc z założenia, że ma to kluczowe znaczenie dla międzynarodowej społeczności badaczy z uwagi na najszerszy możliwie dostęp do informacji o źródłach. Warto podkreślić, że przed rozpoczęciem prac przez NIFC w bazie RISM dostępne były zaledwie 154 rekordy odnoszące się do Fryderyka Chopina (z czego tylko 24 ze zbiorów polskich)<sup>17</sup>, obecnie zaś liczba ta wynosi 6 791 (z czego 6 391 ze zbiorów polskich)<sup>18</sup>. Stawia to Chopina na ósmym miejscu (nie licząc kompozycji anonimowych), jeśli chodzi o liczbę rekordów źródeł w bazie RISM, zaraz za Mozartem, Haydnem czy Händlem, a przed Beethovenem czy niezwykle przecież popularnym w swoim czasie Johannem Adolffem Hassem (stąd wielka liczba zachowanych źródeł kompozycji tego autora w europejskich archiwach).

	do 2015	od 2016 do 2022
Chopin, Fryderyk	154	6 791
Moniuszko, Stanisław	10	2 271
Noskowski, Zygmunt	2	573
Stefani Józef	8	457
Hertz, Michał	0	306
Elsner, Józef	154	294
Paderewski, Ignacy J	7	261
Kurpiński, Karol	41	206
Wieniawski, Adam	0	184
Niewiadomski, Stanisław	4	146
Kozłowski, Józef	48	130
Dobrzyński, Ignacy F	3	119
Nowakowski, Józef	7	115

Przykład 4. Metadane RISM dla wybranych kompozytorów

Na przykładzie 4 uwidoczniło przyrost rekordów dla wybranych kompozytorów. Warto zauważyć, że dla części tych autorów brakowało jakichkolwiek informacji związanych z ich zachowanym

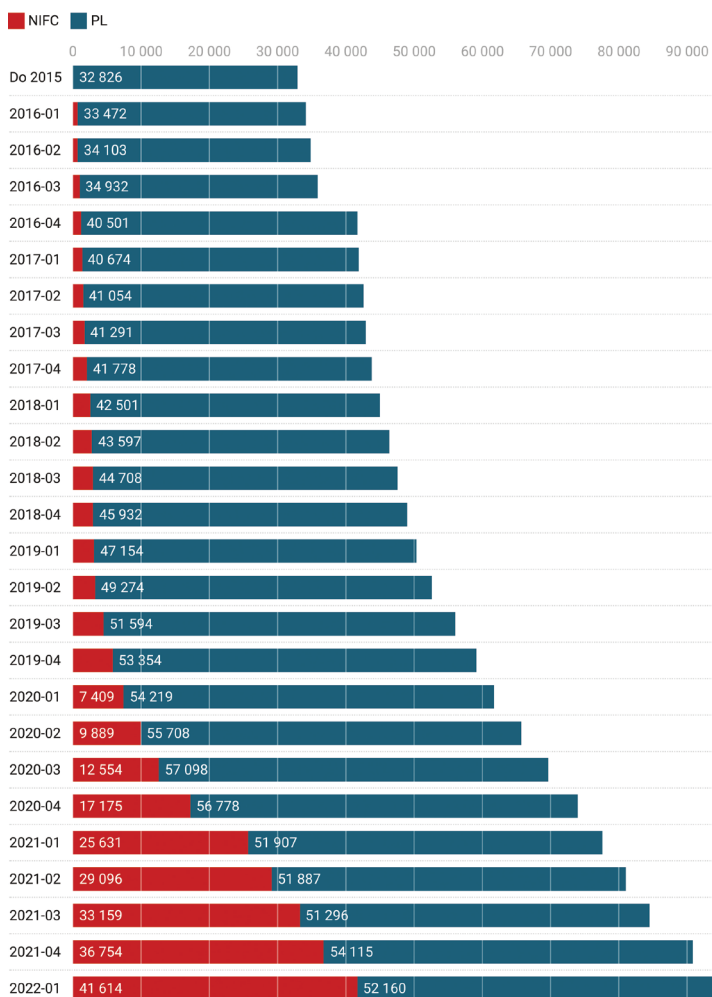
<sup>16</sup> Zob. Ching-chih Chen, James Z. Wang, Howard D. Wactlar, Kevin Kiernan, *Digital Imagery for Significant Cultural and Historical Materials*, „International Journal on Digital Libraries” 2005, 5 (4), s. 275–286.

<sup>17</sup> Stan na dzień 1 I 2016.

<sup>18</sup> Stan na dzień 12 VII 2023.

materialnym dziedzictwem. Co istotne, przy tworzeniu rekordów źródłowych w RISM powstają także liczne rekordy słownikowe (tzw. *authorities*) dla osób i instytucji. Dzięki temu do obiegu naukowego wchodzi informacje nie tylko o samych kompozytorach, ale także o osobach powiązanych (jak np. adresaci dedykacji) czy instytucjach (jak oficyny wydawnicze czy ośrodki gromadzące muzykalia). Jest to podstawa dla budowy w przyszłości opartej na tych danych sieci semantycznej, co pozwoli znacznie skuteczniej prowadzić m.in. badania proveniencyjne.

Obecnie zespół RISM w NIFC koordynuje opracowanie najważniejszych polskich kolekcji muzycznych. O skali tych prac świadczy fakt, że od momentu rozpoczęcia działalności do końca lutego 2022 roku rekordy RISM wytworzone przez NIFC stanowiły już ponad 44% wszystkich istniejących rekordów źródeł polskich. Zjawisko to dobrze ilustruje przykład 5. Uwidoczniono tu proporcję rekordów

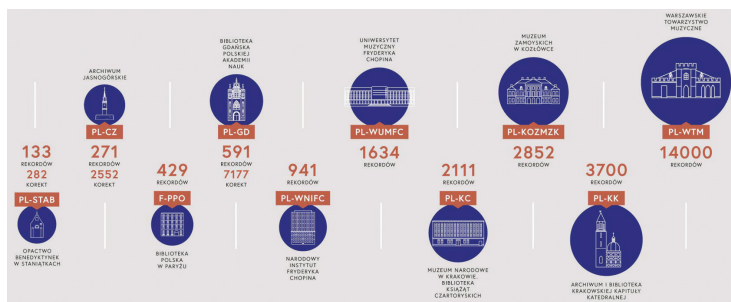


Przykład 5. Polskie źródła w RISM opracowane przez NIFC



opracowanych przez zespół NIFC w podziale na kwartały kolejnych lat działalności.

Z zestawienia jednoznacznie wynika, jak bardzo dynamiczny był przyrost danych (a co za tym idzie – wiedzy) dotyczących źródeł do muzyki polskiej. Widać także, że wszystkie pozostałe polskie grupy RISM (jest ich kilkanaście) w tym samym okresie stworzyły w sumie razem niecałe 20 tys. rekordów, podczas gdy sama tylko grupa RISM w NIFC wypracowała niemal 42 tys. rekordów. Było to możliwe dzięki ścisłej koordynacji kwerend, prac digitalizacyjnych i katalogowania oraz wypracowaniu efektywnej metodologii działań, a także skompletowaniu odpowiedniego zespołu. Warto zwrócić uwagę na skokowy przyrost danych (należy pamiętać, że dotyczy to zarówno metadanych RISM, jak i powiązanych z nimi zawsze cyfrowych kopii źródeł) w dwu aspektach: a) dostępności wiedzy o zbiorach oraz b) dostępności wiedzy na temat twórczości konkretnych kompozytorów. Przykład 6 obrazuje, jak istotna zmiana nastąpiła w tym zakresie.



Przykład 6. Dane RISM wytworzone dla wybranych kolekcji

Zaprezentowane dane wskazują jednoznacznie na kluczowy wkład NIFC w zakresie udostępniania źródeł muzycznych oraz uzupełniania informacji na temat konkretnych kolekcji. W tym kontekście warto odnotować fakt, że na portalu prezentującym źródła i metadane uruchomiono mechanizm, który pozwala na wprowadzanie korekt i informowanie o ewentualnych błędach. Dzięki temu możliwość poprawiania metadanych w bazie RISM pośrednio uzyskały także osoby, które nie mają uprawnień do pracy w programie Muscat<sup>19</sup>. Ma to duże znaczenie dla budowania społeczności osób zaangażowanych w rozbudowę i poprawę jakości tworzonej bazy.

## Transkrypcje cyfrowe

Na gruncie klasycznego edytorstwa muzycznego wskazuje się kilka typów edycji, w tym m.in. edycję krytyczną, wydanie urtekstowe, edycję wykonawczą (praktyczną), transkrypcję dyplomatyczną, wydanie faksymilowe i in. W projektach NIFC przyjęto,

19

Muscat to oprogramowanie webowe, które umożliwia katalogowanie źródeł w bazie RISM. Zob. <https://rism.info/community/muscat.html> [dostęp: 18 X 2023].

że w pierwszej kolejności tworzone będą transkrypcje dyplomatyczne, które staną się podstawą do sporządzenia w dalszej kolejności edycji zmodernizowanych (wykonawczych) oraz – docelowo – cyfrowych edycji krytycznych (w ramach odrębnego projektu)<sup>20</sup>.

W trakcie trzech lat prac w projekcie POPC-1 udało się sporządzić niemal 400 transkrypcji, które obejmują prawie wszystkie dzieła Chopina. Większość edycji cyfrowych ma co najmniej dwie lub trzy wersje źródłowe oparte na różnych pierwodrukach (np. na wydaniu francuskim, niemieckim, polskim i in.). W efekcie stworzono system, który pozwala na wyrenderowanie partytur z zapisu symbolicznego, a także umożliwia ich przeszukiwanie i porównywanie. W projekcie POPC-2 powstało niemal 7,5 tys. transkrypcji (plików), na które składają się zarówno pojedyncze dzieła, jak i duże utwory cykliczne. Gdyby wyrenderować te dane do formatu PDF w standardowym układzie partytury, uzyskalibyśmy około 60 tys. stron materiału nutowego. Tak wysoka efektywność możliwa była dzięki zastosowaniu specjalnie stworzonego oprogramowania do kontroli przepływu pracy, zarządzania i kontroli wersji plików. Poniżej zostaną scharakteryzowane zastosowana metodologia, wykorzystane narzędzia informatyczne, a także założenia edycyjne.

## Metodologia prac

Podstawowe cele, jakie przyjęto w projekcie w zakresie przygotowanych transkrypcji, to:

- 1) przygotowanie najwyższej jakości składów,
- 2) stworzenie bazy repertuaru muzyki polskiej,
- 3) możliwie wierne zakodowanie informacji zapisanej w źródłach,
- 4) wytworzenie materiału, który będzie mógł być wykorzystany do analizy skomputeryzowanej i stworzenia wyszukiwarki muzycznej.

Realizacja tych założeń wymagała postępowania z należytą starannością w trakcie tworzenia transkrypcji oraz ich dalszego opracowania. Na potrzeby bazy transkrypcji przyjęto, że tworzone będą trzy typy składów:

- 1) tzw. transkrypcja dyplomatyczna (możliwie wierne oddająca zapis źródłowy),
- 2) edycja modernizowana (do celów wykonawczych),
- 3) cyfrowa edycja krytyczna (uwzględniająca ingerencje i poprawki redaktora wraz z ich wykazem w formie komentarza rewizyjnego).

Proces tworzenia edycji cyfrowych polega na: 1) pozyskaniu danych poprzez ręczne wykonanie transkrypcji lub OCR<sup>21</sup> (w przypadku nadających się do tego druków muzycznych), 2) zapisaniu danych do formatu XML, 3) konwersji do formatu KRN, 4) korekty

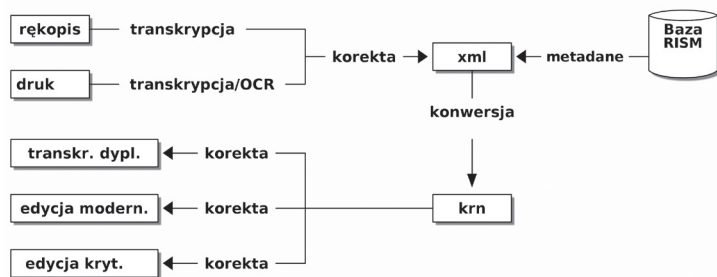
20

Przygotowując roboczą instrukcję wydawniczą na potrzeby NIFC, oparto się m.in. na następujących publikacjach: Willi Apel, *The Notation of Polyphonic Music, 900–1600*, Cambridge 1953; John Caldwell, *Editing Early Music*, Oksford 2001; Theodor Dumitrescu, Karl Kügle, Marnix van Berchum (red.), *Early Music Editing: Principles, Historiography, Future Directions*, Turnhout 2013; James Grier, *The Critical Editing of Music. History, Method, and Practice*, Cambridge 2004; Don Harrán, *New Light on the Question of Text Underlay Prior to Zarlino*, „Acta Musicologica” 1973, nr 45 (1), s. 24–56; idem, *Vicentino and His Rules of Text Underlay*, „The Musical Quarterly” 1973, nr 59 (4), s. 620–632; idem, *How to „Lay” the „Lay”: New Thoughts on Text Underlay*, „Musica Disciplina” 1997, nr 51, s. 231–262; Cristle Collins Judd, *Reading Renaissance Music Theory: Hearing with the Eyes*, Cambridge 2000; Gary Towne, *A Systematic Formulation of Sixteenth-Century Text Underlay Rules (Part I of II)*, „Musica Disciplina” 1990, nr 44, s. 255–287; idem, *A Systematic Formulation of Sixteenth-Century Text Underlay Rules (Part II)*, „Musica Disciplina” 1991, nr 45, s. 143–168.

21

*Optical Character Recognition*, czyli optyczne rozpoznawanie znaków. Dla zapisu muzycznego to właściwie OMR (*Optical Music Recognition*). Jest to proces, w którym za pomocą odpowiedniego oprogramowania zapis graficzny zostaje automatycznie odczytany i zapisany w formie tekstowej.

zapisu do wersji możliwie najbardziej zbliżonej do źródła, 5) stworzeniu na tej podstawie edycji modernizowanej i krytycznej (zob. przykł. 7).



Przykład 7. Przepływ danych w procesie tworzenia transkrypcji

Główne cele, dla których przyjęto wskazaną metodologię pracy, to zagwarantowanie maksymalnie efektywnego przepływu pracy i danych, a także zapewnienie możliwie najlepszej kontroli jakości danych. Dzięki założonej metodzie każda transkrypcja dyplomatyczna jest poddana co najmniej dwukrotnej rewizji przez dwie różne osoby, edycja krytyczna zaś dwukrotnej lub trzykrotnej rewizji przez różne osoby. Ma to nie tylko zagwarantować najwyższą jakość danych z uwagi na konieczność spełnienia wysokich standardów edycji krytycznych, ale także zapewnić, że dane poddawane późniejszej analizie (także analizie statystycznej obejmującej całe korpusy dzieł) nie zawierają błędów mogących zniekształcać wyniki badań.

Z powyższych założeń uwidocznionych na diagramie wynika metodologia pracy oraz specyficzne zasady edycji stosowane przez redaktorów. Podział zadań, a także konieczność dysponowania odrębnymi kompetencjami wymagały dokonania podziału redaktorów na dwie grupy:

- 1) redaktorzy plików XML,
- 2) redaktorzy plików KRN.

Aby zapewnić spójność z pozostałymi elementami portalu (skany źródeł i informacje o nich) metadane do plików XML pobierane są z bazy RISM. Kiedy redaktor rozpoczyna pracę nad transkrypcją w pierwszym etapie (kodowanie zapisu do formatu MusicXML), otrzymuje pusty plik, w którym uzupełnione są metadane, w tym identyfikatory RISM ID pozwalające później połączyć automatycznie transkrypcję z cyfrowym odwzorowaniem źródła. Po zakończeniu prac nad pierwszą wersją pliku (XML) jest on konwertowany do formatu KRN (format zapisu plików zgodnie z syntaksą Humdrum)<sup>22</sup>. Redaktor pliku KRN dokonuje korekty oraz uzupełnia transkrypcję o elementy niemożliwe do zapisania w formacie XML. W przypadku

22

Zob. David Huron, *Humdrum Toolkit: Reference Manual*, Menlo Park 1994.

utworów cyklicznych (jak np. msze, symfonie etc.) prace nad całym dziełem prowadzone są jednocześnie przez kilku redaktorów. Pozwala to rozbudować kompetencje zespołu oraz stanowi dodatkowy element wsparcia i korekty, ponieważ korzystając z systemu kontroli pracy i wersjonowania plików, redaktorzy komunikują się ze sobą oraz z koordynatorami prac w wewnętrznym forum dyskusyjnym. Całość takiej wymiany uwag wraz z wszystkimi wersjami każdego pliku (jego pełną historią z komentarzami) zapisywana jest w systemie. Pozwala to z jednej strony na późniejsze odtworzenie od samego początku procesu powstawania transkrypcji (łącznie z możliwością dotarcia do uzasadnień decyzji podejmowanych przez redaktorów), z drugiej strony stanowi podstawę do stworzenia na dalszym etapie komentarza rewizyjnego (w przypadku cyfrowej edycji krytycznej).

Wydania krytyczne muszą spełniać wysokie standardy pracy naukowej, zostać opatrzone stosownym wstępem (z możliwie pełnym opisem źródła, biogramu kompozytora) oraz – obowiązkowo – tzw. komentarzem rewizyjnym. Ten ostatni stanowi opis zmian – niezbędnych niemal we wszystkich rodzajach dawniejszych źródeł – w stosunku do źródła wprowadzonych w edycji przez redaktora. Mogą one dotyczyć np. pewnych niezrozumiałych już obecnie konwencji notacyjnych; stanowią niejednokrotnie uzupełnienia zapisów niekonsekwentnych, jak np. częste w rękopisach pomijanie kreski taktowej bądź akcydencji (znaków chromatycznych). W tym sensie praca redaktora sprowadza się do „krytyki” źródła. Komentarz rewizyjny jako lista wszelkich odstępstw od oryginału powinien pozwolić w całości zrekonstruować oryginalną jego treść. Ma to na celu umożliwienie badaczom oraz muzykom podjęcie ewentualnie innych niż proponowane przez redaktora decyzji, dotyczących wątpliwych momentów zapisu muzycznego. Warto w tym momencie przypomnieć, że w edytorstwie muzycznym, a także – szerzej – wszelkiej krytyce tekstu (także literackiego) jedną z głównych zasad jest zasada *difficilior lectio potior* (dosł. „trudniejsza lekcja ma większą moc”), która oznacza, że w przypadku sprzecznych przekazów mniej oczywiste, wydające się w pierwszej chwili błędnym odczytaniem tekstu należy uznać za bliższe oryginalnemu zamysłowi autora. Zasada ta wynika z przyjęcia tezy, że skrypcy, kopiści mają tendencję do automatycznego poprawiania np. nietypowej pisowni, w efekcie czego powstają ujednolicenia, które nie były zamysłem twórcy tekstu i w istocie stanowią zagubienie czynnika „oryginalności” oraz bliskości pierwowzoru. Wypisanie różnic pomiędzy źródłem a jego edycją niejednokrotnie okazuje się najbardziej czasochłonnym etapem procesu edycyjnego. W projektach NIFC z założenia jest to proces przynajmniej w części zautomatyzowany i wykonywany przez odpowiednie narzędzia programistyczne.

W dobie coraz powszechniejszych i łatwo dostępnych baz i repozytoriów cyfrowych, dzięki którym można zapoznać się z oryginałem, tworzenie transkrypcji dyplomatycznych wydaje się coraz mniej celowe, jak wskazuje m.in. John Caldwell<sup>23</sup>. W projektach

23

Zob. Caldwell, op. cit., s. 1.

NIFC jest to niemal oczywiste, ponieważ transkrypcje będą tworzone wyłącznie na bazie źródeł bądź zdigitalizowanych przez NIFC, bądź dostępnych w domenie publicznej. Jednak z uwagi na metodologię pracy oraz sposób działania narzędzi komputerowych tworzenie takich transkrypcji jest konieczne. Stanowią one ogniwo pośrednie pomiędzy źródłem a jego krytyczną edycją. Jak wskazano, pliki KRN, czyli główny typ formatu, w jakim zapisywane są partytury w projekcie, to pliki tekstowe. Plik KRN w dużym uproszczeniu przypomina arkusz kalkulacyjny, w którym w kolumnach wypisane są głosy kompozycji<sup>24</sup>, poszczególne linie zaś to kolejne „wydarzenia” w utworze. Przyjmijmy, że mamy jednotaktowy, dwugłosowy utwór w metrum 4/4, a w każdym z głosów występują trzy ćwierćnuty i jedna pauza ćwierćnutowa. Plik KRN będzie zatem zawierał dwie kolumny, sam zapis nutowy zaś – cztery linijki (nie licząc innych elementów, jak np. klucze, metrum, kreski taktowe etc.).

Wspomniana automatyzacja procesu tworzenia komentarza rewizyjnego (lub porównania wersji tego samego utworu na podstawie różnych źródeł) polega na wykonaniu komputerowego porównania plików tekstowych: 1) transkrypcji dyplomatycznej i 2) edycji krytycznej (lub dwu transkrypcji dyplomatycznych w przypadku wielu źródeł do tej samej kompozycji). Narzędzia komputerowe bardzo dobrze radzą sobie z tego typu zadaniami. Na najprostszym przykładzie wygląda to w sposób następujący. Załóżmy, że w źródle mamy zapisaną ćwierćnutę *c'*. Redaktor wie jednak, że w tym miejscu powinien znajdować się krzyżyk podwyższający *c'* na *cis'*. W pliku transkrypcji dyplomatycznej znajdzie się zatem zapis: 4c, natomiast w pliku edycji krytycznej – 4c#. Narzędzia komputerowe porównają obydwie pliki i wskażą różnice pomiędzy nimi<sup>25</sup>:

$$4c\# - 4c = \#$$

Analogicznie do powyższej sytuacji możemy wyłowić rozbieżności w wysokości dźwięków, wartościach rytmicznych etc. Każda, najdrobniejsza nawet różnica pomiędzy plikami zostanie odnaleziona przez algorytm i wskazana. Lista owych odmienności stanowi istotę komentarza rewizyjnego. Oczywiście przytoczone przykłady są niezwykle proste, jednak Humdrum to nie tylko rodzaj zapisu, ale także bardzo złożony system oraz zestaw kilkudziesięciu programów, za pomocą których można wykonywać bardzo skomplikowane operacje. Przykłady te mają unaocznić, jak niezwykle ważne jest, aby transkrypcja dyplomatyczna możliwie najwierniej odwzorowała zapis źródłowy. W przeciwnym razie efekt porównania da fałszywe wyniki.

Tradycyjne metody pracy redaktora przygotowującego edycję krytyczną można sprowadzić do kilku najważniejszych etapów. W pierwszej kolejności tworzona jest transkrypcja – już w tej fazie niejednokrotnie wprowadzane są korekty oczywistych pomyłek oraz uzupełniane braki i pominięcia<sup>26</sup>. Następnie wykonuje się głębszą

24  
Każdy z głosów to osobna kolumna, w ten sposób zapisywany jest też tekst słowny oraz wiele innych elementów dzieła.

25  
W pewnym sensie przypomina to np. tzw. śledzenie zmian w dokumentach tekstowych.

26  
Np. w sytuacji, kiedy głosy kompozycji „nie schodzą się” wskutek pomyłki skryptora, który mógł przeoczyć takt utworu lub całą linię.

korektę – poprawia współbrzmienia, ujednolica oznaczenia dynamiczne i artykulacyjne, dodaje brakujące akcydencje etc. Praca ta wymaga częstego sięgania do źródła, np. w tym celu, aby porównać analogiczne do siebie fragmenty kompozycji i sprawdzić, czy to porównanie pomaga rozstrzygnąć wątpliwości. W efekcie powstaje edycja poprawiona i uzupełniona. Stopień ingerencji edytora zależny jest najczęściej od przyjętych założeń wydawniczych lub celu publikacji. W edycjach krytycznych przyjmuje się możliwie najmniejszą liczbę ingerencji redaktora. Kolejny etap pracy to ponowne porównanie tak sporządzonej edycji ze źródłem i wypisanie wszystkich ingerencji. W ten sposób powstaje komentarz rewizyjny. Ten etap pracy bywa najbardziej żmudny, a jednocześnie jest niejednokrotnie źródłem pomylek.

W projektach NIFC dzięki zastosowaniu narzędzi – wypracowanych przez twórcę Humdrumu, Davida Hurona z Ohio State University<sup>27</sup> i od wielu lat rozwijanych na Uniwersytecie Stanforda w Center for Computer Assisted Research in the Humanities<sup>28</sup> przez współpracującego z NIFC Craiga Sappa – można ten proces uprościć. Co ważne, tworzenie komentarza rewizyjnego zostaje w dużym stopniu zautomatyzowane, co redukuje do minimum ryzyko błędu i czyni sam proces nieporównanie szybszym. Umożliwia też redaktorowi skupienie się wyłącznie na merytorycznej pracy, czyli analizie źródła i podejmowaniu na jej podstawie konkretnych decyzji.

Przy takiej metodologii pracy dochodzi do przesunięcia punktu ciężkości: niewralgicznym momentem nie jest tworzenie komentarza, lecz weryfikacja transkrypcji dyplomatycznej. Przyjęty podział pracy powoduje też, że transkrypcja ta poddawana jest znacznie dokładniejszej rewizji, niż ma to miejsce zazwyczaj w modelu tradycyjnym. W efekcie powstaje edycja o dużym stopniu wiarygodności, a uzyskane dane mają o wiele wyższą jakość. Co więcej, jeśli dysponujemy co najmniej dwoma wersjami edycji utworu (dyplomatyczną i krytyczną bądź wykonawczą), możemy nie tylko wygenerować automatycznie listę korektur, ale także pokazać różnice pomiędzy źródłem a jego edycją, wyświetlając je w partyturze cyfrowej. Najbardziej czytelną metodą będzie np. podświetlenie lub pokolorowanie tych nut w edycji, które różnią się od zapisu oryginalnego. Możliwe jest też m.in. stworzenie narzędzi pozwalających na wyświetlanie wariantów i różnych decyzji redaktorów lub wręcz rozwiązań alternatywnych<sup>29</sup>.

Mając na uwadze powyższe, a także model przepływu danych i pracy przedstawiony wcześniej, możemy wywnioskować, że dla każdego utworu opracowywanego w przyjętym *workflow* powinny docelowo powstać następujące pliki:

- 1) plik XML – możliwie wiernie odzwierciedlający przekaz źródłowy,
- 2) plik KRN – transkrypcja dyplomatyczna,
- 3) plik KRN – transkrypcja zmodernizowana (wykonawcza),
- 4) plik KRN – edycja krytyczna.

27  
Zob. <https://www.musiccognition.osu.edu/people/david-huron/> [dostęp: 18 X 2023].

28  
Zob. <http://www.ccarh.org/> [dostęp: 18 X 2023]

29  
Spektakularne tego przykłady można zobaczyć na stronie programu Verovio, który NIFC wykorzystuje do renderowania partytur: <https://book.verovio.org/> [dostęp: 18 X 2023].

Aktualnie w bazie transkrypcji NIFC udostępnionej na stronie internetowej<sup>30</sup> znajduje się 7 448 plików autorstwa 370 kompozytorów z okresu od XVI do początków XX wieku. Dla 1 018 transkrypcji sporządzono także wersje zmodernizowane (wykonawcze). Ponadto 1 066 plików jest połączonych z obrazami cyfrowymi za pomocą interfejsu IIIF (połączenie to realizuje się na poziomie systemu transkrypcji do systemu partytury w źródle). Niezależnie od tego oczywiście wszystkie transkrypcje zawierają odnośniki do kompletnych źródeł muzycznych, których skany znajdują się w portalu. Bardziej szczegółowo funkcjonalności systemu renderowania partytur cyfrowych opisujemy poniżej.

## System renderowania i wyszukiwania

Wypracowany przez NIFC system renderowania online wraz z wyszukiwarką muzyczną wzbogacony o elementy skomputeryzowanej analizy muzycznej jest dostępny pod adresem <https://polish-scores.org/>. Technologia systemu opiera się na dwu podstawowych komponentach: 1) narzędziach Humdrumu, które umożliwiają zapis tekstu muzycznego oraz jego efektywne przeszukiwanie i analizę, a także 2) Verovio, czyli programie renderującym dynamiczne partytury online. Predefiniowane filtry pozwalają na wskazanie kryteriów, jakie ma spełnić wyszukiwanie, a łączenie owych filtrów z elementami przeszukiwania tekstowego (w tym muzycznego) pozwala bardzo precyzyjnie określić oczekiwane wyniki.



Przykład 8. Widok strony głównej systemu renderowania partytur cyfrowych

Na stronie głównej wyświetlana jest lista wszystkich dostępnych partytur. W pasku głównym tego widoku znajduje się kilka przycisków, które pozwalają wstępnie filtrować i sortować listę: 1) **M** – ogranicza listę do partytur z wersjami zmodernizowanymi (wykonawczymi), 2) **I** – pozwala wyświetlić wyłącznie te partytury, które nie tylko mają link do skanów w portalu, ale także są zintegrowane z obrazem za pomocą technologii IIIF (o szczegółach w dalszej części), 3) **T** – wyświetla nuty, które mają także warstwę tekstu

<sup>30</sup>  
Zob. <https://polishscores.org/> [dostęp: 18 X 2023].

słownego lub są jej pozbawione (przycisk można kliknąć kilkakrotnie, aby osiągnąć oczekiwany rezultat), 4) **D** – sortuje listę według daty dodania pliku (w kolejności od najnowszych). W tym samym pasku znajduje się także przycisk (dwie krzyżujące się strzałki), który pozwala wybrać losowo dowolną partyturę z bazy, oraz symbol tabeli, który kopiuje listę wyników (wraz z linkami i najważniejszymi metadanymi) do pliku w formacie CSV. Listę taką można wkleić za pomocą komendy CTRL-V np. do arkusza kalkulacyjnego, aby zapisać spis wyników wyszukiwania. Na poziomie strony głównej działają też skróty klawiaturowe, które uruchamiają te (oraz inne filtry): litera **m** wyświetla partytury wykonawcze, litera **d** sortuje według daty dodania, litery **c** lub **k** sortują listę według nazwiska kompozytora, litera **s** – według sygnatury, litera **t** zaś – według tytułu dzieła. Poniżej omówionego paska znajdują się rozwijalne listy wyboru, dzięki którym możemy ograniczyć liczbę wyników według: 1) nazwiska kompozytora, 2) okresu powstania źródła<sup>31</sup>, 3) miejsca przechowywania źródła (oznaczone za pomocą *siglum* RISM), 4) gatunku muzycznego<sup>32</sup>, 5) instrumentu (obecnie można wybrać tylko jeden – docelowo system pozwoli dobierać zestawy instrumentarium), 6) tonacji<sup>33</sup>, 7) skali (czyli trybu dur-moll lub skali kościelnej)<sup>34</sup>.

Oprócz wskazanych filtrów na stronie głównej umieszczono także trzy pola pozwalające na wyszukiwanie tekstowe. Pierwsze z nich wyszukuje tekst w polu tytułu utworu. Po wpisaniu w nim fragmentu tytułu – np. „regi” – otrzymamy listę utworów, w których tytule znajduje się m.in. słowo „regina”, ale także „confregit” czy „regius”. W podobny sposób działa wyszukiwarka dla tekstu słownego, przy czym przeszukiwane są całe partytury (wraz ze wszystkimi głosami zawierającymi tekst słowny), a na liście wyników wskazana jest także liczba wystąpień wyszukiwanej frazy w utworze. Oczywiście wpisanie dokładnie całego słowa lub frazy pozwala uzyskać bardziej precyzyjny wynik. Wyszukiwarka w takim przypadku stosuje domyślnie operator logiczny AND, czyli połączy koniunkcją wpisane frazy, a wśród wyników znajdują się tylko utwory zawierające łącznie wszystkie wskazane sformułowania<sup>35</sup>.

Osobnego omówienia wymaga trzecie pole, które umożliwia wyszukiwanie według zapisu muzycznego. Narzędzia Humdrumu pozwalają na bardziej precyzyjne lub uogólnione przeglądanie indeksów pod tym kątem. Można zatem szukać melodii wyłącznie według konturu (np. wpisując jej kierunek za pomocą + lub -), według interwałów wraz z określeniem ich jakości i kierunku lub bez nich (np. +P5 oznacza skok kwinty czystej w górę, -3 skok o dowolną tercję w dół etc.) czy według konkretnych dźwięków (zgodnie z nomenklaturą amerykańską, czyli np. melodię chorałową *Ave maris stella* znajdziemy, wpisując sekwencję znaków „d a b g a”). Aktualna wersja systemu umożliwia wyszukiwanie wyłącznie według ostatniego ze wskazanych sposobów. Poszerzenie funkcjonalności o wspomniane powyżej planowane jest w nieodległej przyszłości<sup>36</sup>. Po wpisaniu

31 Uwaga, data na liście odnosi się do datowania źródła, a nie okresu działalności kompozytora czy powstania dzieła, dlatego np. po wybraniu opcji „XIX w.” możemy odnaleźć na liście wyników dzieła starsze, ale publikowane właśnie w XIX wieku.

32 Ten filtr nie działa jeszcze precyzyjnie i trwają prace nad jego usprawnieniem.

33 Tonacja znajdującego się w bazie utworu określana jest za pomocą algorytmu i w szczególnych, nielicznych przypadkach może być określona nieprecyzyjnie.

34 Algorytm badający skale kościelne jest dopiero tworzony na podstawie zgromadzonego materiału, a zatem filtr wskazuje na skale wyłącznie tam, gdzie zostały one określone ręcznie. Po zakończeniu prac nad udoskonaleniem algorytmu badającego tonalność zostanie on w pełni wdrożony.

35 Na przykład po wpisaniu do wyszukiwarki tekstów słownych kolejno słów oddzielonych spacją: „regina caeli ave maris” znajdziemy wyłącznie jeden utwór, w którym system wskaże następujące wystąpienia fraz: „3 ave 1 caeli 1 maris 1 regina 1 reginae 2 suave”.

36 O pełnych możliwościach wyszukiwarki opartej na narzędziach Humdrumu można przeczytać w dokumentacji na stronie <https://extras.humdrum.org/man/themax/> [dostęp: 18 X 2023].



sekwencji dźwięków do wyszukiwarki lista wyników zostanie zaktualizowana wyłącznie do tych utworów, które zawierają dany motyw. Co więcej, po wybraniu konkretnego utworu, na wyrenderowanym składzie wyszukiwany motyw będzie oznaczony kolorem czerwonym, a u góry strony znajdzie się lista taktów, w których motyw występuje. Jest to zatem niezwykle przydatne narzędzie nie tylko pozwalające odnajdywać np. melodie chorałowe czy melodie pieśni stosowane w utworach wielogłosowych, ale także pomocne w poszukiwaniu konkordancji pomiędzy źródłami.



Przykład 9. Widok transkrypcji systemu renderowania partytur cyfrowych

Poniżej omówione zostaną pokrótce wybrane funkcjonalności dostępne z poziomu widoku konkretnej partytury (zob. przykład 9). Kluczowym elementem wspomnianego poziomu jest wyrenderowana za pomocą programu Verovio<sup>37</sup> partytura cyfrowa. Verovio jest programem stworzonym i rozwijanym przez Laurenta Pugina z RISM Digital Center<sup>38</sup>, umożliwiającym wyrenderowanie w przeglądarce internetowej (bez konieczności instalowania jakiegokolwiek dodatkowego oprogramowania) składu nutowego zapisanego w formacie MEI<sup>39</sup> (w przypadku transkrypcji tworzonych przez NIFC jest to format KRN automatycznie konwertowany do MEI i renderowany). Algorytmy, które odpowiadają za automatyczny układ partytury oraz rozmieszczenie poszczególnych jej elementów, zostały pierwotnie stworzone z myślą o muzyce dawnej, dlatego najlepsze efekty dają właśnie na gruncie tego typu repertuaru. W ostatnich latach rozwijanych jest jednak wiele projektów, które wykorzystują Verovio do tworzenia cyfrowych edycji, dzięki czemu także edycje muzyki późniejszej (a zwłaszcza fortepianowej) zyskują w coraz większym stopniu spełniać wymogi profesjonalnych wydań. Ważnym impulsem do rozwoju Verovio był m.in. projekt POPC-I realizowany przez NIFC, ze względu na konieczność dostosowania oprogramowania do dzieł Chopina, niezwykle trudnych w automatycznym renderowaniu za pomocą algorytmów.

Na stronie prezentującej cyfrową transkrypcję powyżej składu nutowego umieszczony jest pasek, na którym znajdują się informacje odnoszące się do autora kompozycji i źródła. *Signum* kolekcji

37  
Zob. <https://www.verovio.org/index.xhtml> [dostęp: 18 X 2023].

38  
Zob. <https://rism.info/digital-center.html> [dostęp: 18 X 2023].

39  
Zob. <https://music-encoding.org/> [dostęp: 18 X 2023].

oraz sygnatura są odnośnikami, które po kliknięciu przenoszą do strony RISM<sup>40</sup>, gdzie możemy zapoznać się z rekordem słownikowym instytucji lub rekordem bibliograficznym źródła. Poniżej sygnatury obiektu umieszczona jest ikonka prowadząca do skanu źródła w portalu lub – o ile dana transkrypcja została połączona ze skanami za pomocą serwera IIF – także do konkretnych jego fragmentów (w takim przypadku pojawia się też ikona IIF). Podobne odnośniki umieszczono obok nazwiska kompozytora – w tym przypadku przenoszą one do serwisów takich jak Wikipedia, stron z rekordami kontroli autorytatywnej (jak np. VIAF) oraz do Portalu Muzyki Polskiej rozwijanego przez NIFC<sup>41</sup>. Portal ten stanowi bazę wiedzy na temat twórców, dzieł i szeroko pojętej sztuki (w kontekście dziedzictwa muzycznego). Powyżej paska z informacjami nt. źródła i autora znajduje się pasek narzędziowy. Umożliwia on m.in. nawigowanie między kolejnymi transkrypcjami z listy, odtworzenie utworu w formacie midi, wyszukanie melodii w danym utworze czy też dostosowanie graficzne składu (ikonka kółka zębatego). Ikona wykresu kołowego daje dostęp do funkcji analitycznych, w tym tych pozwalających na wyświetlenie struktury rytmicznej utworu (narzędzie przydatne w poszukiwaniu rytmów tanecznych) czy wyodrębnienie tekstu słownego. Symbol klucza wyświetla model tonalny kompozycji zwizualizowany za pomocą trójkątnego grafu<sup>42</sup>. Ikona pobierania umożliwia otwarcie kodu źródłowego, otwarcie w edytorze Verovio Humdrum Viewer<sup>43</sup> lub pobranie pliku w formatach KRN, MEI, MIDI, MusicXML, PDF lub SVG. Wybór ten daje ogromne możliwości dalszego wykorzystania materiału, są to bowiem w zasadzie wszystkie najpopularniejsze formaty zapisu tekstu muzycznego lub wyrenderowanego składu – jak np. format grafiki wektorowej SVG, którego można z powodzeniem użyć w publikacji czy na stronie internetowej. Warto dodać, że całość materiału stworzonego w projektach NIFC jest przechowywana i może zostać kompletnie pobrana z repozytorium GitHub<sup>44</sup>.

## IIF

Szczególnie przydatną oraz ułatwiającą korzystanie z systemu transkrypcji cyfrowych funkcją jest zastosowanie serwera IIF w repozytorium NIFC. International Image Interoperability Framework to zestaw narzędzi programistycznych ułatwiających i ujednolicejących sposób dostępu do zasobów cyfrowych, a w szczególności obrazów powstałych w wyniku digitalizacji obiektów oryginalnych. Dzięki zastosowaniu tzw. pliku manifestu użytkownik nie tylko pobiera komplet informacji o obrazie, ale jest też w stanie wyświetlić jego fragment lub całość. Na stronie transkrypcji NIFC – jak już wskazano – dla części partytur wykonano pracę polegającą na połączeniu cyfrowej edycji z odnośnym fragmentem skanu źródła (prace nad połączeniem wszystkich transkrypcji trwają). Przyjęto,

---

40

Zob. <https://rism.online/> [dostęp: 18 X 2023].

---

41

Zob. <https://portalmuzyki-polskiej.pl/pl> [dostęp: 18 X 2023].

---

42

Modele te bliżej scharakteryzowane zostały m.in. w artykule poświęconym analizie komputerowej repertuaru XIX-wiecznego – zob. Konik, *Repertuar...*, op. cit., przyp. 13, a także w pracy nt. repertuaru staniąteckiego: idem, *Monografia zespołu muzycznego oraz katalog muzykantiów klasztoru ss. benedyktynek w Staniątkach*, Musica Iagellonica, Kraków 2016, s. 169–172.

---

43

Edytor online przeznaczony dla formatu KRN.

---

44

Zob. <https://github.com/pl-wnifc> [dostęp: 18 X 2023].

że połączenie to nie będzie realizowane na poziomie nuta do nuty, tylko system muzyczny do systemu. Rozwiązanie takie jest nie tylko bardziej korzystne z metodologicznego punktu widzenia, ale pozwala wyświetlić dany fragment źródła wraz z jego kontekstem. W praktyce interfejs IIIF na stronie transkrypcji NIFC działa w taki sposób, że po dwukrotnym kliknięciu danej nuty zostaje w osobnym oknie wyświetlony fragment skanu źródła, na podstawie którego wpisano właśnie ten znak. Jest to niezwykle użyteczne narzędzie dla badaczy, pozwala bowiem w bardzo wygodny sposób zweryfikować jakość transkrypcji (gdy mamy wątpliwości co do konkretnej decyzji edytora), a ponadto przyspiesza i czyni wygodniejszym tworzenie komentarza rewizyjnego w przypadku samodzielnego opracowywania danej kompozycji. Także dla wykonawców zastosowanie serwera IIIF jest nie bez znaczenia, ponieważ ułatwia przyjęcie konkretnych rozwiązań poprzez sięgnięcie do źródła.

## Portal

Portal, na którym prezentowane są zdigitalizowane źródła, zaopatrzone jest w typowe dla podobnych systemów funkcjonalności. W związku z tym skupimy się jedynie na kilku specyficznych rozwiązaniach, jakie zastosowano z uwagi na rodzaj publikowanego materiału. Ponieważ centrum, wokół którego został zbudowany cały system, stanowią źródła muzyczne, przyjętym standardem dla metadanych jest Marc21 w formacie wypracowanym na potrzeby RISM. Komplet metadanych dla każdego źródła znajduje się w bazie RISM, natomiast na portalu wyświetlane są jedynie wybrane pola, pozwalające na skuteczne i precyzyjne wyszukiwanie. Metadane aktualizowane są poprzez API RISM – dzieje się tak w momencie dodania nowych obiektów lub zgłoszenia konieczności korekty. Należy mieć na uwadze, że wewnętrzne procedury RISM Zentralredaktion (jednostki odpowiadającej za jakość metadanych) powodują, że zgłoszone nowe rekordy lub korekty istniejących publikowane są z pewnym opóźnieniem, a co za tym idzie, także w portalu pojawiają się dopiero po pewnym czasie (zwykle trwa to około tygodnia). Dzięki wdrożeniu standardu RISM do bazy NIFC wszystkie metadane charakteryzuje bardzo wysoki poziom powiązań z perspektywy *linked open data*. Z tego powodu w widoku konkretnego obiektu zawsze znajdują się bezpośrednie odnośniki do bazy RISM, tak aby każdy korzystający z zasobów mógł w prosty sposób przejść do powiązanych systemów. W panelu metadanych dostępna jest też lista transkrypcji cyfrowych wykonanych na podstawie danego źródła (o ile zostały sporządzone, ponieważ nie wszystkie prezentowane obiekty były poddane transkrypcji).

Portal źródeł jest ciągle rozbudowywany o nowe funkcje oraz uzupełniany o kolejne zasoby. W ostatnim czasie opublikowano archiwa Biblioteki Uniwersytetu Łódzkiego, w trakcie publikacji są materiały

Biblioteki im. Zielińskich Towarzystwa Naukowego Płockiego, w dalszej kolejności – Muzeum Teatralnego w Warszawie (przy Teatrze Wielkim – Operze Narodowej), obecnie zaś digitalizowane są i będą opracowane, a następnie opublikowane zbiory Zakładu Narodowego im. Ossolińskich. W przygotowaniu do digitalizacji jest też kolekcja Archiwum Archidiecezjalnego i Biblioteki Kapitulnej we Wrocławiu oraz tamtejszej Akademii Muzycznej im. Karola Lipińskiego. Aktualnie niemal w całym kraju prowadzone są kwerendy, które pozwolą przygotować digitalizację kolejnych zbiorów. Dotyczy to zwłaszcza mniejszych ośrodków, w których przypadku dostęp do źródeł jest niejednokrotnie utrudniony czy wręcz niemożliwy. Co istotne, w przyszłych działaniach NIFC z zakresu digitalizacji i opracowania źródeł muzycznych zostanie poszerzony nie tylko zakres czasowy (także o źródła wcześniejsze niż XVI-wieczne), ale też merytoryczny – włączone do opracowania zostaną źródła liturgiczne (chorałowe) oraz materiały kontekstualne (jak archiwalia dotyczące życia muzycznego, czyli np. księgi rachunkowe i inne dokumenty).

W zakresie nowych funkcjonalności portal zostanie w przyszłości całkowicie zintegrowany z Portalem Chopinowskim (stworzonym w projekcie POPC-i), którego wszystkie dane zostaną włączone do docelowego systemu. Nastąpi też pełniejsza integracja z Portalem Muzyki Polskiej. Ponadto planowane jest wdrożenie wyszukiwarki muzycznej z poziomu portalu, pozwalającej na wyświetlenie po wpisaniu melodii listy wyników tych źródeł, w których dany motyw się pojawia (dotyczy to oczywiście utworów transkrybowanych cyfrowo bądź mających incipit muzyczny w bazie RISM). W kolejnych krokach planowane jest stworzenie zmapowanego – na podstawie danych proveniencyjnych – modelu źródeł, co pozwoli nie tylko wyszukiwać je według lokalizacji przechowywania, ale także śledzić przepływ repertuaru. W momencie, w którym baza będzie wystarczająco rozbudowana, zostanie także stworzony semantyczny model relacji pomiędzy źródłami. Warto podkreślić, że już na obecnym etapie rozwoju omawiana baza jest największym w Polsce (i jednym z największych w Europie) repozytorium źródeł muzycznych, natomiast w zakresie transkrypcji cyfrowych – największą tego typu bazą w świecie.

## Perspektywy badawcze

Zaprezentowane rozwiązania stanowią podstawę do tworzenia zupełnie nowych narzędzi warsztatu krytyki źródła oraz – szerzej – otwierają nowe perspektywy badawcze. Wydaje się, że szczególnie obiecująca może być analiza dużych korpusów dzieł, pozwalająca na tworzenie modeli do analizy stylokrytycznej. Z dużym powodzeniem tego typu badania prowadzone są w Stanach Zjednoczonych m.in. przez Jesego Rodina z Uniwersytetu Stanforda. Zastosowanie w przyszłości elementów SI w doniesieniu do tego typu zbiorów

danych otwiera kolejne, nieznane obszary w źródłoznawstwie. Jest to jednak ciągle perspektywa przyszłości, choć być może już niezbyt odległej. W tym momencie skupimy się na kilku wybranych aspektach, ważnych dla badań repertuaru muzyki polskiej (i w Polsce tworzonej oraz wykonywanej).

## Cyfrowa edycja krytyczna

Spśród wspomnianych wcześniej typów transkrypcji cyfrowych w zasadzie wyłącznie dyplomatyczna oraz krytyczna (w nieco mniejszym stopniu) mają wartość dla badacza. W przypadku transkrypcji modernizowanej (wykonawczej) jej walory sprowadzają się głównie do tego, że stanowi gotowy materiał dla muzyka chcącego wykonać dany utwór. Cyfrowa edycja krytyczna z perspektywy przetwarzania maszynowego ma mniej zalet niż dyplomatyczna, zwłaszcza gdy korzystamy z możliwości interfejsu IIF i na bieżąco możemy konfrontować edycję z zapisem źródłowym. Tak bezpośredni dostęp do źródła w pewnym sensie stanowi ekwiwalent komentarza rewizyjnego, a przynajmniej wykazu korektur. Zakładając, że dany system renderowania partytur daje też dostęp do informacji bibliograficznych na temat źródła oraz autora kompozycji (jak jest w przypadku portali NIFC), możemy mówić o niemal wszystkich elementach nieodzownych dla tradycyjnie pojmowanej edycji krytycznej. Dotyczy to także muzykologicznej charakterystyki utworu, która może znaleźć się na Portalu Muzyki Polskiej. Warto zastanowić się, które z elementów procesu przygotowywania tak pojętej cyfrowej edycji krytycznej mogą podlegać algorytmizacji, a przez to usprawnieniu. Z pewnością wspomniany już proces sporządzania wykazu korektur dzięki uprzedniemu stworzeniu transkrypcji dyplomatycznej i porównaniu jej z wersją finalną (edycją krytyczną) podlega automatyzacji. A jest to jeden z bardziej żmudnych etapów w tradycyjnym procesie tworzenia edycji krytycznej. W praktyce na prostym przykładzie zastosowanie algorytmu w tym celu może wyglądać, jak poniżej<sup>45</sup>:



Przykład 10. Zastosowanie algorytmu do porównania zapisu

45

Oczywiście jest to jedynie uproszczony przykład, ponieważ wszystkie dające się zanotować elementy i aspekty zapisu muzycznego oraz cech źródła poddają się takiej analizie.

Załóżmy, że przedstawione na przykładzie 10 zapisy, to: A – zapis źródłowy, B – poprawiona edycja (krytyczna). W formacie KRN ich kod źródłowy wygląda odpowiednio (B – kolumna lewa, A – kolumna prawa):

```

**kern      **kern
*c1efG2     *c1efG2
*M4/4       *M4/4
=1          =1
4d          4d
4a          4a
4b          4b-
4g          4g
=2          =2
4a          4a
4b          4b
4dd         4dd
4cc         4cc
=3          =3
2b          4b
.           4a
2a          2r
==          ==
*_          *_

```

Aby uzyskać listę różnic między tymi wersjami w formie tekstowej, wystarczy uruchomić znane ze środowiska unixowych systemów operacyjnych narzędzie `diff`. Jego zastosowanie da następujący efekt w postaci listy:

```

7c7
< 4b
---
> 4b-
15,16c15,17
< 2b
< 2a
---
> 4b
> 4a
> 2r

```

Jest to zestaw danych w pełni pozwalający sporządzić listy korektur dla tego fragmentu zapisu. Wystarczy jedynie napisać program (prace nad nim trwają w NIFC), który na tej podstawie wygeneruje listę korektur w tradycyjnej i bardziej czytelnej dla odbiorcy formie. Co najważniejsze, metoda ta nie tylko uwalnia redaktora od konieczności wielokrotnego powracania do tych samych miejsc bądź zapamiętania, w którym momencie dokonano zmiany względem zapisu źródłowego, ale niemal zupełnie eliminuje ryzyko błędu. W tym przypadku najistotniejsze jest, by zadbać o najwyższą jakość transkrypcji dyplomatycznej. To właśnie te dane stanowią punkt odniesienia dla algorytmu w zakresie odnotowywania zmian w stosunku do źródła.

## Badania atrybucji i konkordancje

Spośród ponad 1 mln 600 tys. rekordów w bazie RISM 300 tys. to rekordy anonimowe. Dla danych z polskim *siglum* proporcja ta to 19 292 rekordy anonimowe na 102 605 wszystkich rekordów<sup>46</sup>. Oznacza to, że niemal 20% metadanych dla źródeł pozbawionych jest kluczowej informacji na temat autorstwa. Oczywiście problem narasta, gdy mowa o źródłach starszych. Ustalenie atrybucji bądź wskazanie konkordancji między źródłami umożliwia baza RISM głównie dzięki zapisanym w niej incipitom muzycznym w formacie Plaine and Easie Code<sup>47</sup>. Niestety, jeśli mamy do czynienia ze źródłem niekompletnie zachowanym lub gdy nie zakodowano w bazie incipitów (sytuacja taka nie jest rzadka), nie sposób sparować rekordów, a zatem nie da się rozstrzygnąć kwestii autorstwa czy ewentualnych konkordancji.

Inaczej jest w przypadku bazy transkrypcji NIFC, gdzie w otwartym formacie tekstowym zapisywano całe utwory, a nawet fragmentarycznie zachowane źródła (gdy brak np. części głosów lub obiektu). W tym przypadku możliwe jest zastosowanie algorytmu, który porównuje wszystkie zapisane dane i wskazuje procentowe podobieństwo pomiędzy utworami, głosami czy frazami<sup>48</sup>. Aby zaprezentować na prostym przykładzie działanie i potencjał algorytmu, porównamy automatycznie repertuar zachowany w dwu źródłach – będzie to kancjonał E ze Staniątek oraz kancjonał L 1642 z Sandomierza. Oba źródła pochodzą z kręgu benedyktynek, powstały w podobnym okresie (pocz. XVIII wieku) i zawierają podobny repertuar (są pomiędzy nimi liczne konkordancje). W pierwszej kolejności za pomocą programu *tindex* tworzymy indeks kompletu utworów. Indeks taki zapisuje wszystkie głosy każdego utworu za pomocą różnych możliwych do przeszukania sposobów (np. rejestrując kierunek melodii, poszczególne interwały, wysokości dźwięków etc.). W obydwu omawianych kancjonałach zapisano w sumie niemal 140 kompozycji. Plik indeksu stworzony na podstawie plików KRN zawiera niemal 500 linii kodu, co oznacza, że algorytm z całego repertuaru wyodrębnił właśnie tyle niezależnych rekordów do analizy. Następnie uruchamiamy program *XSIM*, który pobiera kolejno z każdego utworu głos, dzieli go na zdefiniowane liczby nut frazy (długość frazy jest określana przez użytkownika) i porównuje te fragmenty z każdym głosem każdego utworu, podzielonego na analogiczne do wskazanych fragmenty. Po zakończeniu pracy – która może być czasochłonna z uwagi na dużą liczbę przetwarzanych danych – algorytm zapisuje wynik do pliku wymiany danych w formacie JSON. Dla omawianego repertuaru wygenerowany plik z danymi analitycznymi zawiera ponad 28 tys. linii kodu, co daje pojęcie o ogromnej ilości przetworzonych informacji nawet dla tak niewielkiego wyjściowego zbioru danych. Przykładowy pojedynczy rekord zapisany w pliku JSON wygenerowanym przez algorytm wygląda następująco:

46  
Stan na dzień 15 VII 2023.

47  
Zob. <https://www.iaml.info/plaine-easie-code> [dostęp: 18 X 2023].

48  
Funkcje takie ma program *xsim* napisany przez Craiga Sappa z Uniwersytetu Stanforda.

```

{
  "filename": "L_1642a_001_Omni_die.krn",
  "voice": "1",
  "matches": [
    {"percent":100, "filevoice":"L_1642a_001_Omni_die.krn"},
    {"percent":100, "filevoice":"L_1642a_001_Omni_die.krn::1"},
    {"percent":8.7, "filevoice":"L_1642b_020_Chrystus_zmartwychwstal.krn"},
    {"percent":6.5, "filevoice":"St_E_002_Vesp_Magnificat.krn"},
    {"percent":6.5, "filevoice":"St_E_029_Angelus_Domini.krn"},
    {"percent":6.5, "filevoice":"St_E_029_Angelus_Domini.krn::1"}
  ]
}

```

Interpretacja wyników należy do badacza. Z powyższego rekordu dowiadujemy się, że algorytm wyodrębnił pierwszy głos (czyli najniższy) z pliku o nazwie *L\_1642a\_001\_Omni\_die.krn*, zawierającego utwór *Omni die dic Mariae* z kancjonału **L 1642**. Testowe porównanie (zawsze pierwsze w kolejności) głosu ze sobą samym dało wynik 100%, co potwierdza poprawność procedury. Następnie wśród całego analizowanego zbioru danych algorytm znalazł podobieństwa do wybranych fragmentów innych utworów na poziomie 8,7% oraz 6,5%, co oznacza, że zbieżności te są zupełnie przypadkowe (we wskazanych utworach muszą pojawiać się podobne krótkie motywy). Z doświadczeń pracy z programem XSIM wynika, że z badawczego punktu widzenia istotne są podobieństwa wskazane przez algorytm w przedziale od 60% do około 99%. Analogie występujące rzadziej zwykle są nieistotne i przypadkowe, natomiast wynik 100% daje niemal wyłącznie tautologiczne porównanie testowe. Nieczęsto zdarza się, że nawet w przypadku tych samych utworów zapisanych w różnych źródłach podobieństwo wynosić będzie właśnie 100%, najczęściej zapisy te różnią się bowiem drobnymi szczegółami, jak np. zamiana wartości rytmicznych, inne wysokości pojedynczych dźwięków etc. *Nota bene* właśnie tego typu różnice stanowią największą trudność w bazie RISM.

Sprawdźmy jednak, czy w wyniku analizy pojawiły się jakieś potencjalnie interesujące wyniki. W omawianym pliku json znajduje się 11 rekordów (co można wyszukać automatycznie bez konieczności przeglądania wszystkich niemal 30 tys. linii) z podobieństwem wskazanym na poziomie < 90%. Jest tak np. w przypadku głosu najwyższego w pliku *L\_1642a\_013\_Krolewnie\_wieczney.krn* oraz pliku *St\_E\_021\_Krolewnie\_wieczny.krn*. Już same nazwy tych plików wskazują, że mamy do czynienia z tym samym utworem zapisanym w dwu różnych kancjonałach. Porównajmy ich głosy w zapisie nutowym na przykładzie 12.

Widzimy wyraźnie, że są to niemal identyczne zapisy, różniące się drobnymi szczegółami. Omawiany przykład ma na celu jedynie wykazanie przydatności narzędzia, sam w sobie bowiem jest dość oczywisty. Jednak prowadzona aktualnie z wykorzystaniem metod



Canto I

Kró - lew - nie wiecz - nej nie - ba wy - so - kie - go I po - śre - dni - cy świa - ta u - pa - dle - go

12

Śpie - waj - my chwa - łę od - da - jąc sa - me - mu Bo - gu wie - czne - mu.

Canto I

Kró - lew - nie wie - cznej nie - ba wy - so - kie - go I po - śre - dni - czce świa - ta u - pa - dle - go

13

Śpie - waj - my chwa - łę od - da - jąc sa - me - mu Bo - gu wie - czne - mu.

Przykład 12. Podobieństwa zapisów po wyszukiwaniu za pomocą programu XSIM

skomputeryzowanych analiza repertuaru kolekcji PL-STab, PL-Kk, PL-SA oraz PL-GD daje niezwykle obiecujące rezultaty w zakresie odnajdywania konkordancji pomiędzy źródłami, a także badania atrybucji dzieł anonimowych.

## Wnioski

Dwa niezwykle dobrze rokujące w badaniach źródłoznawczych nowe obszary, które pojawiły się dzięki standaryzacji zapisu metadanych (jak w RISM) oraz tworzeniu spójnych pod względem jakości i formatu danych baz transkrypcji cyfrowych, to mapowanie źródeł i rozwijanie kolejnych narzędzi analizy stylokrytycznej. Są to dziedziny badań z dużym powodzeniem praktykowane na gruncie *digital humanities* w różnych dyscyplinach naukowych. Nie miejsce tu, aby szczegółowo omawiać możliwe aplikacje tych metod na gruncie muzykologii. Zwłaszcza w zakresie badań źródłowych dotyczących zasobów polskich archiwów brak jeszcze wystarczającej ilości odpowiednio przygotowanego i opracowanego materiału, aby prowadzić na większą skalę takie działania. Zakładając jednak, że tempo prac prowadzonych przez NIFC w tym zakresie zostanie utrzymane, można przyjąć, że już w ciągu najbliższych trzech lub czterech lat będziemy dysponowali odpowiednią bazą, spełniającą definicję muzycznego *big data*. Zwłaszcza rozbudowa bazy transkrypcji o kolejne zapisy źródłowe, w tym uzupełnienie o: 1) źródła liturgiczne (chorałowe), 2) repertuar pieśni kancjonałowych z XVI i XVII wieku, 3) zbiór pieśni ludowych (w tym Kolbergowski), da nowe możliwości, choćby w wyszukiwaniu melodii chorałowych czy ludowych w muzyce polifonicznej, rytmów tanecznych czy śledzenia kolejnych konkordancji. Narzędzia do prowadzenia takich badań już istnieją i – jak staraliśmy się wykazać na kilku przykładach – spełniają bardzo dobrze swoje funkcje. Obecnie kluczowe wydaje się poszerzenie bazy materiałów do analizy.

---

**ABSTRACT**

*Heritage of Polish Music – digitalization, development, and research perspectives*

In the study of large corpora of musical works, regardless of whether we define them as a collection of compositions by a given composer, works cultivated in a specific center or in a specific period, it is important to have a possibly representative number of appropriately prepared material for analysis. This article discusses the process of creating a comprehensive database of sources along with their digital transcriptions and the resulting new research opportunities and perspectives. All the indicated examples as well as the characteristics of the work methodology used are based on the experience of the Chopin Institute in the implementation of two projects financed by the European Union: (1) Chopin Heritage in Open Access and Heritage of Polish Music in Open Access. These projects were implemented in 2017–2021 and 2018–2022, respectively. As a result, two portals and separate pages were created where digital transcriptions are available.

---

**KEYWORDS**

digital musicology, computer assisted analysis, digital editions, rism

---

**ABSTRAKT**

W badaniach dużych korpusów dzieł muzycznych, niezależnie od tego czy zdefiniujemy je jako zbiór kompozycji danego kompozytora, twórczość uprawianą w określonym ośrodku czy w konkretnym okresie, istotną jest możliwie reprezentatywna liczba odpowiednio przygotowanego materiału do analizy. W artykule niniejszym omówiono kolejno: proces tworzenia kompleksowej bazy źródeł wraz z ich cyfrowymi transkrypcjami oraz wynikające stąd nowe możliwości i perspektywy badawcze. Wszystkie wskazane przykłady, a także charakterystyka stosowanej metodologii prac oparta jest na doświadczeniach Instytutu Chopina w realizacji dwu projektów finansowanych ze środków Unii Europejskiej: (1) „Dziedzictwo Chopinowskie w otwartym dostępie” oraz „Dziedzictwo muzyki polskiej w otwartym dostępie”. Projekty te realizowane było odpowiednio w latach 2017–2021 oraz 2018–2022. W ich efekcie powstały dwa portale oraz odrębne strony, na których dostępne są transkrypcje cyfrowe.

---

**SŁOWA KLUCZOWE**

muzykologia cyfrowa, analiza skomputeryzowana, edycje cyfrowe, rism

---

**DR MARCIN KONIK**

absolwent filozofii i muzykologii na Uniwersytecie Jagiellońskim w Krakowie, kierownik Działu Informacji Naukowej Narodowego Instytutu Fryderyka Chopina w Warszawie. Doktorat uzyskał na podstawie rozprawy poświęconej wątkom kosmologicznym (ze szczególnym uwzględnieniem koncepcji *musica mundana* Boecjusza) w średniowiecznych traktatach z zakresu teorii muzyki. Odbił staże naukowe w prestiżowych ośrodkach: Jacob's School of Music na Indiana University w Bloomington oraz na Wydziale Muzyki Uniwersytetu Stanforda. Współpracuje m.in. z Centre for Computer Assisted Research in the Humanities Uniwersytetu Stanforda. W swoich badaniach skupia się na problematyce skomputeryzowanej analizy muzycznej, muzycznych bazach danych, projektach digitalizacyjnych. Specjalizuje się w badaniach źródłowych muzyki polskiej okresu XVII–XIX wieku.