

ARTYKUŁY I ROZ PRA WY

Anna Skolimowska, Magdalena Turska

Uniwersytet Warszawski, kontakt: as@aluw.edu.pl,
ORCID ID: 0000-0002-2346-9149;
e-editiones / TEI Publisher, kontakt: tuurma@gmail.com,
ORCID ID: 0009-0003-9458-2119

Długie trwanie edycji cyfrowej

*Przypadek Korpusu Tekstów
i Korespondencji Jana Dantyszka*

Wstęp

Jan Dantyszek (1485–1548), nobilitowany mieszczanin gdański, sekretarz królewski w służbie Zygmunta I i Bony Sforzy, ich długoletni ambasador na dworach europejskich, biskup chełmiński i warmiński, utalentowany poeta nowolaciński, przyjaciel humanistów, uczonych, polityków i kupców z całej Europy i Nowego Świata, wyróżnia się spośród innych wybitnych obywateli Królestwa Polskiego pierwszej połowy XVI wieku przede wszystkim niezwykle ważką spuścizną w postaci korespondencji.

Korespondencja ta, rozproszona w zbiorach rękopiśmiennych dziesiątków archiwów i bibliotek na terenie kilkunastu krajów europejskich, od wieków budzi zainteresowanie badaczy dawnych dziejów, dawnych tekstów i zjawisk kulturowych. Wyjątkowo obszerny korpus ponad 6000 zachowanych listów (głównie w języku łacińskim i niemieckim) dokumentuje sieć powiązań Dantyszka (655 korespondentów) i stanowi cenne, wielowątkowe świadectwo czasów i okoliczności, w jakich przyszło mu działać.

Szczególne znaczenie listów Dantyszka zauważył już współczesny mu Stanisław Górski, komponując rękopiśmienną kolekcję, tzw. Acta Tomicianiana – zbiór korespondencji i dokumentów królewskich, pomyślany jako materiał edukacyjny dla młodego króla Zygmunta II Augusta¹. W pierwszej redakcji Tomicianów listy-sprawozdania dyplomatyczne Dantyszka zostały potraktowane wyjątkowo – wydzielono je w osobny tom, podczas gdy inne materiały zestawiono chronologicznie.

Korespondencja Dantyszka jest również obficie reprezentowana w odpowiadających jej chronologicznie kilkudziesięciu spośród tzw. Tek Naruszewicza². Ponad 200 tomów kopiariuszy stanowiło plon wielkiej akcji kopiowania źródeł do historii Polski, zorganizowanej w XVIII wieku przez Adama Naruszewicza na zlecenie króla Stanisława Augusta.

Wyjątkowo ciekawe są losy kopiowania tzw. kodeksów upsalskich – dwóch poszytów oryginalnych listów do Dantyszka, wybranych spośród setek innych jego listów (złupionych z archiwum biskupów warmińskich w początkach XVIII wieku) przez bibliotekarza Uniwersytetu w Uppsali Erica Benzeliusa do osobnego opracowania ze względu na szczególne znaczenie dla dziejów reformacji. (Notabene poszyty te nie zostały zwrócone wraz z innymi łupami szwedzkiimi podczas akcji rewindykacyjnej w XIX wieku, ponieważ akademicy szwedzcy uznali, że zawierają zbyt cenny dla nich materiał źródłowy). W jakiś czas po opracowaniu tych poszytów, w 1717 roku Benzeliusz odpisał z nich sześćdziesiąt dziewięć listów. Odpis ten stał się podstawą kolejnych kopii. Zachowało się siedem zależnych od siebie kopiariuszy tej kolekcji listów, wiadomo też o czterech lub pięciu innych – dziś zaginionych³.

Na początku XX wieku, zapewne dzięki ukazaniu się pierwszych drukowanych tomów serii Acta Tomicianiana, znaczenia korespondencji Dantyszka świadomi byli historycy i literaturoznawcy z Polskiej Akademii Umiejętności w Krakowie⁴. Postulaty Kazimierza Morawskiego podjął wówczas Antoni Marian Kurpiel, który drogą systematycznych kwerend archiwalnych gromadził wszelkie informacje o Dantyszku i jego listach z myślą o stworzeniu monografii humanisty⁵. Wprawdzie zamiar ten udaremniła przedwczesna śmierć Kurpiela, lecz pozostało po nim trzynaście tek zawierających ułożone chronologicznie regesty, excerpty, a niekiedy także pełne kopie, obejmujące łącznie nieco ponad połowę spośród znanych obecnie 6117 listów z korespondencji Dantyszka. Z niezwykle starannie uporządkowanej spuścizny Kurpiela chętnie korzystali związani z Polską Akademią Umiejętności wydawcy następnych tomów Tomicianów, w szczególności zaś wybitny znawca polskiego odrodzenia Władysław Pocięcha. Inny członek PAU, filolog klasyczny Stanisław Skimina, w połowie ubiegłego stulecia wydał i opracował utwory poetyckie Dantyszka⁶.

Spuścizna epistolograficzna Dantyszka stanowiła również przedmiot zainteresowań wielu innych, zarówno polskich, jak i zagranicznych badaczy i wydawców źródeł – zajmujących się m.in. przeszłością Polski, Hiszpanii, Prus, Erazmem z Rotterdamu, Mikołajem Kopernikiem. Trudno

je wszystkie wymienić w krótkim artykule. Od początku XIX wieku do momentu uruchomienia w latach osiemnastych XX wieku na Uniwersytecie Warszawskim przedstawianego tu projektu *Rejestracja i Publikacja Korespondencji Jana Dantyszka (1485–1548)* (zwanego też w skrócie *Projektem Dantyszek*) powstało nie mniej niż 146 częściowych drukowanych publikacji źródłowych zawierających fragmenty tej korespondencji. W ostatnich latach liczba ta wzrosła do ok. 200 pozycji. Są to publikacje różnego typu, o zróżnicowanym poziomie opracowania edytorskiego, służące różnym celom. Niektóre zawierają ledwie jeden bądź kilka listów ilustrujących jakieś zjawisko, inne są tomami rozbudowanych, opatrzonych streszczeniami listów, inwentarzy archiwalnych kolekcji rękopiśmiennych. Wszystkie one są rejestrowane w bazie danych stanowiącej jądro systemu informatycznego służącego do realizacji naszego projektu. W bazie tej gromadzone są podstawowe metadane dotyczące korespondencji Dantyszka i tekstów jego autorstwa, z uwzględnieniem wszystkich znanych przekazów źródłowych (rękopisów i starych druków), oraz, jak wyżej wspomniano, wszystkich publikacji. Teksty Dantyszka niebędące listami – mowy, memoriały i instrukcje poselskie, utwory poetyckie czy wreszcie wystawiane przez niego dokumenty – objętościowo wprawdzie stanowią margines w stosunku do liczby zachowanych listów, lecz jest to materiał ważny dla edycji listów, ponieważ niejednokrotnie pozwala umieścić opracowywaną korespondencję w szerszym kontekście źródłowym⁷.

Projekt Dantyszek jest realizowany przez interdyscyplinarny zespół badawczy pod kierunkiem Anny Skolimowskiej i monitorowany przez Międzynarodową Radę Naukową Projektu, której przewodniczącym jest Jerzy Axer. W skład Rady wchodzi badacze mający doświadczenie w zakresie wydawania korespondencji wczesnonowożytnej.

Projekt przewiduje publikację internetową całości materiału w wersji pełnotekstowej w postaci *Corpus of Ioannes Dantiscus' Texts & Correspondence* / *Korpus Tekstów i Korespondencji Jana Dantyszka* (CIDTC) oraz publikację pełnego inwentarza korespondencji i wyboru listów w ukazującej się od 2004 roku serii drukowanej „Corpus Epistularum Ioannis Dantisci” (CEID).

Do dzisiaj, po z górą trzydziestu latach prac edytorskich, został opublikowany pełen inwentarz korespondencji (zarówno w wersji papierowej – trzy tomy, jak i internetowej), ok. 60% tekstów udostępniono *in extenso* w internecie (w tym wszystkie znane teksty autorstwa Dantyszka) i opublikowano drukiem cztery tomy wybranych listów. Horyzonty

czasowe projektu obejmują przynajmniej dziesięć lat dalszych intensywnych prac edytorskich.

Publikowane listy opatrzone są komentarzem naukowym w języku angielskim. Komentarz ten obejmuje aparat krytyczny oraz podstawowe objaśnienia rzeczowe (streszczenie listu, identyfikacja osób i miejsc, wybrane similia, a także niezbędne objaśnienia kontekstu). Przyjęto założenie, że tekst listu jest udostępniany w internecie od razu po opracowaniu

aparatu krytycznego, komentarz rzeczowy może być natomiast uzupełniany później, podobnie jak indeksy osób i miejsc, które są systemowo powiązane z przypisami rzeczowymi generowanymi automatycznie w rezultacie znakowania tekstu. Z tego samego systemu generowane są wstępne wersje tomów publikowanych drukiem. Materiał ten po wygenerowaniu automatycznym wymaga zazwyczaj stosunkowo niewielkich prac redakcyjnych.

Tabela 1.

	Realizatorzy	Technologia	Cel	Publikacja
XVI wiek	Stanisław Górski i zespół kopistów	archiwizacja materiałów źródłowych kancelarii królewskiej; kopiowanie odręczne	dostępność źródeł dla elit władzy	kolekcja rękopisów „Teki Górskiego”; rękopiśmienne Acta Tomiciana
XVIII wiek	Adam Naruszewicz i zespół badaczy źródeł oraz kopistów	kwerendy archiwalne; kopiowanie odręczne	zebranie podstawy źródłowej do wielotomowej <i>Historii narodu polskiego</i>	kolekcja rękopisów „Teki Naruszewicza”; publikacja <i>Historii narodu polskiego</i>
XIX wiek	Tytus Działyński, Zygmunt Celichowski		opracowanie rękopiśmiennych Acta Tomiciana do publikacji drukiem	seria drukowana Acta Tomiciana, t. 1–13
XX wiek	badacze związani z PAU, w szczególności Antoni Marian Kurpiel, Władysław Pociecha i inni	kwerendy archiwalne; system fiszek; kopiowanie odręczne; mikrofilmowanie	inwentarz rękopiśmienny; opracowanie listów do publikacji drukiem	„Teki Kurpiela”; seria drukowana Acta Tomiciana, t. 14–18 (oparte na rozszerzonej w stosunku do rękopiśmiennych AT podstawie źródłowej)
	Zakład Edytorstwa Greckiego i Łacińskiego IFK UW: Jerzy Axer, Jerzy Mańkowski, Katarzyna Kotońska, Joanna Ziabicka i zespół	kwerendy archiwalne; system fiszek; dBase +	inwentarz listów; opracowanie listów do publikacji drukiem	<i>Espanoles y Polacos en la Corte de Carlos V. Cartas del embajador Juan Dantisco</i> , Madrid 1994
	Zakład Edytorstwa Greckiego i Łacińskiego IFK UW: Jerzy Axer, Jerzy Mańkowski, Anna Skolimowska i zespół	kwerendy archiwalne; dBase + Microsoft Fox Pro	inwentarz listów; opracowanie listów do publikacji drukiem	stworzenie podstawy dalszych prac edytorskich
XXI wiek	Pracownia Edytorstwa Źródeł OBTA UW: Anna Skolimowska i zespół	kwerendy archiwalne; MS Access; MS Word	inwentarz listów, źródeł rękopiśmiennych i publikacji drukowanych; transkrypcje wybranych listów	seria drukowana CEID, 2004–
	Pracownia Edytorstwa Źródeł i Humanistyki Cyfrowej WAL UW: Anna Skolimowska, Magdalena Turska i zespół	baza danych MySQL oraz system publikacji oparty na językach programowania PHP i XSLT (LAMP + XSLT) TEI XML	ewidencja listów i źródeł; transkrypcje całego korpusu listów i innych dokumentów	seria drukowana CEID, 2004–; publikacja internetowa CIDTC, 2010–

Źródło: opracowanie własne.

Powyższa tabela podsumowuje historię opracowań spuścizny Dantyszka od kolekcji Stanisława Górskiego aż po chwilę obecną.

Wyzwania migracji

Po raz pierwszy komputery do potrzeb projektu zastosowano w latach osiemdziesiątych. Od tego momentu regularnie zmieniał się otaczający nas krajobraz technologiczny, a co za tym idzie, stosowane oprogramowanie. Początkowo zespół badawczy wykorzystywał standardowe systemy baz danych i edycji tekstów. Z czasem został stworzony specjalny, krojony na miarę system informatyczny, oczywiście opierający się na ogólnie dostępnych technologiach. Analizując sytuację na przestrzeni ostatnich trzydziestu lat, możemy łatwo zauważyć, że mniej więcej co dekadę niezbędne okazywało się całkowite przeprojektowanie systemu przechowywania i edycji danych. Po części potrzeba ta była podyktowana czynnikami zewnętrznymi: naturalnym upadkiem przestarzałych technologii i pojawianiem się nowych. Zwłaszcza ten drugi aspekt, nowe możliwości techniczne, skutkowały rozbudowanymi aspiracjami zarówno co do celów projektu (idea stale aktualizowanej publikacji internetowej), jak i praktycznej organizacji pracy nad nim.

Charakterystyczną cechą ewolucji *Projektu Dantyszek* był wzrost – z jednej strony zakres i wieloaspektowość opracowywanego materiału, z drugiej nowe cele projektu (inwentarz *vs* transkrypcje *in extenso* z komentarzem, druk *vs* publikacja online) wymusiły znaczne rozbudowanie modelu danych i towarzyszącej infrastruktury informatycznej, a co za tym idzie, drastyczny wzrost ich złożoności.

O ile w kwestii zakresu materiału możemy mówić o postępie liniowym, dzięki opracowaniu następnych podzbiorów korespondencji, o tyle decyzje dotyczące nowych form pracy i publikacji wymagały każdorazowo skokowej zmiany i konwersji danych między różnymi systemami komputerowymi. Pomiędzy okresami transformacji, wymagającymi skoncentrowanego i skoordynowanego wysiłku, techniczny aspekt utrzymania projektu ograniczał się do okazjonalnych aktualizacji oprogramowania i niezbędnych adaptacji do jego nowszych wersji. W miarę wzrostu złożoności projektu w jego kolejnych inkarnacjach migracje oraz bieżące utrzymanie wymagały znacząco rosnących nakładów.

Dla ilustracji skali zmian: baza danych w systemie Fox Pro była pojedynczą tabelą relacyjną; w systemie Access (1998) tabel było już sześć; po przejściu na MySQL

(2007–2009) możemy już mówić o ok. siedemdziesięciu tabelach⁸. Złożoność systemu wzrosła zatem skokowo, co miało daleko idące konsekwencje dla systemu edycji i publikacji. Każda tak poważna zmiana modelu danych pociąga za sobą potrzebę stworzenia nowego oprogramowania do obsługi nowego modelu, jako że z konieczności (warunkowanej nowością koncepcji edycji cyfrowych oraz złożonością problemów metodologicznych edytorstwa dawnych tekstów) zawsze były to rozwiązania szyte na miarę. O ile przejście od jednej tabeli do sześciu było jeszcze stosunkowo łatwe i zajęło kilka tygodni⁹, o tyle praca nad kolejną wersją trwała ponad dwa lata.

Z jednej strony nie będzie nas to dziwić, jeśli przypomnimy sobie, jak wyglądały komputery i metody korzystania z nich w 1990, 2005 roku i obecnie. Z drugiej chcielibyśmy uniknąć nakładów i trudności związanych z ciągłym przebudowywaniem systemów informatycznych, zwłaszcza w uniwersyteckich realiach budżetowych i przy ograniczonej dostępności wykwalifikowanych akademickich konsultantów IT.

Z początkiem nowego milenium brak konsensusu i powszechnie akceptowanych standardów stał na przeszkodzie optymalizacji i racjonalizacji wykorzystania zasobów przez adaptowanie istniejących rozwiązań. Ta rewolucja miała dopiero nadejść, a pierwszą oznaką nowego była rosnąca popularność XML i narzędzi do jego przetwarzania oraz otwarcie się badaczy, również w Polsce, na metody i zyskujące uznanie standardy humanistyki cyfrowej.

Zwiastunem zmiany paradygmatu w *Projekcie Dantyszek* była decyzja o konwersji transkrypcji do standardu TEI/XML. Wymagała ona sporych nakładów od strony technicznej, a także przeszkolenia oraz dostosowania metod pracy zespołu badawczego. Można jednak śmiało stwierdzić, że ten wysiłek nie poszedł na marne, wręcz przeciwnie, był warunkiem koniecznym rozwoju projektu. Początki były dość trudne z tej racji, że stosunkowo hermetyczna dokumentacja TEI Guidelines była dla nas jedynym źródłem informacji, podpartym bardzo nielicznymi dostępnymi wówczas przykładami wykorzystania TEI do edycji korespondencji¹⁰.

Krótkoterminową konsekwencją zmiany była możliwość równoczesnego opracowywania materiału na potrzeby publikacji drukiem oraz online. Dzięki oparciu edycji na formacie TEI/XML mogłyśmy automatycznie wygenerować z tego samego zbioru danych zarówno wersję internetową w formacie HTML, jak i wersję do składu dla wydawcy formy książkowej. Zapewniło to zgodność obu publikacji oraz zagwarantowało spójność formy i treści przypisów w całym projekcie. Ponadto nasz materiał źródłowy nareszcie był, przynajmniej

na podstawowym poziomie, kompatybilny z innymi tego typu projektami edytorskimi na świecie.

Począwszy od migracji z prostej bazy danych w MS Fox Pro (*de facto* w ówczesnej formie porównywalnej z arkuszem kalkulacyjnym) do systemu Access czy późniejszej konwersji do MySQL, stawałyśmy zawsze przed problemem opracowania mapowania między istniejącym modelem danych a nowym: obszerniejszym i w odmiennej technologii. Niestety zawsze był to proces wymagający stworzenia dedykowanego rozwiązania do konwersji, ponieważ były to modele opracowane wyłącznie na potrzeby *Projektu Dantyszek*. Niemniej dopóki poruszaliśmy się w obszarze technologii relacyjnych, było to zajęcie czasami żmudne, lecz niewymagające zbyt ścisłego łączenia kompetencji technicznych z wiedzą o edytorstwie.

W momencie uruchomienia publikacji cyfrowej obszerny i szczegółowy zbiór metadanych, nad którym zespół badawczy pracował wtedy już od wielu lat, miał, jak można się spodziewać, formę relacyjnej bazy danych – w zgodzie z niepodzielnie królującymi od lat osiemdziesiątych rozwiązaniami opartymi na modelu relacyjnym i języku zapytań SQL. Transkrypcje w TEI/XML wydawały się nam wtedy stosunkowo luźno połączone z systemem opracowywania metadanych tak ze względu na stosowane narzędzia informatyczne, jak metodykę pracy, co zaowocowało hybrydowym systemem opartym na kombinacji bazy danych MySQL i języka PHP (LAMP)¹¹ z procesorem XSLT.

Z perspektywy czasu możemy stwierdzić, że była to konsekwencja praktycznej hegemonii modelu relacyjnego, co skutkowało tym, że fraza „baza danych” domyślnie była rozumiana jako „relacyjna baza danych” i nawet uniwersyteckie kursy baz danych jedynie wspominały o innych modelach, głównie jako ciekawostkach historycznych. Niemniej od początku funkcjonowania komputerowych baz danych równolegle egzystowały inne paradygmaty, w ostatnich zaś dwóch dekadach możemy obserwować ciągły wzrost dostępności i popularności systemów nierelacyjnych. Ze względu na naturalne osadzenie edycji tekstowych w XML logiczną konsekwencją tego wyboru okazało się bowiem zastosowanie takiego formatu nie tylko do transkrypcji, ale również do metadanych. Na horyzoncie zarysowało się odejście od baz relacyjnych w stronę „baz dokumentów” obsługujących format XML. Ta zmiana pociąga za sobą perspektywę kolejnego przeprojektowania systemu publikacji.

Najgwałtowniejszy rozwój technologii noSQL, w tym XML-owych systemów baz danych, a także języków i systemów przetwarzania dokumentów XML, nastąpił w ciągu ostatnich lat. Dla przykładu, specyfikacja języka XQuery¹² w wersji 1.0, o statusie jedynie propozycji w 2001 roku, stała się oficjalną rekomendacją W3C w 2007 roku, a jej najnowsza aktualizacja do wersji 3.1 pochodzi z 2017 roku. Ta ostatnia obejmuje już praktycznie wszystkie zaawansowane mechanizmy, z których możemy korzystać w języku SQL, co świadczy o osiągnięciu przez tzw. X-technologie¹³ etapu pełnej dojrzałości.

Podobny rozwój przeszły inne elementy piramidy technologicznej niezbędne do tworzenia aplikacji do publikacji i edycji źródeł w standardzie TEI. Przede wszystkim mówimy tutaj o ewolucji XML-owych baz danych w stronę platform znacznie wszechstronniejszych, obejmujących nie tylko organizację danych, ale również zintegrowaną platformę programistyczną¹⁴ oraz serwer aplikacji. Łatwo to zauważyć zwłaszcza w przypadku jednego z najpopularniejszych tego typu projektów, systemu eXist-db¹⁵.

Pierwsze próby stworzenia prototypu nowej wersji systemu dla *Projektu Dantyszek*, opartej w całości na X-technologiach, zostały podjęte w 2013 roku. Szybko okazało się,

że stworzenie pełnego mapowania między istniejącą relacyjną bazą metadanych a nagłówkiem *teiHeader* w TEI nie jest, jak przy wcześniejszych migracjach, jedynie problemem technicznym, ale wymaga zarówno kompetencji w zakresie edytorstwa, TEI, jak i umiejętności technicznych, nie wspominając o konsekwencjach dla systemu edycji, organicznie zrośniętego z relacyjną bazą danych. Z wielu powodów – i natury finansowej, i personalnej – ta operacja nie mogła zostać wtedy przeprowadzona. Jednak w tym przypadku czas pracował na naszą korzyść: rozwijały się nie tylko uniwersalne narzędzia, ale dojrzewały również idee bardziej zrównoważonego podejścia do kwestii edycji cyfrowych.

Jedna z autorek z zespołu pracującego nad *Projektem Dantyszek* przeszła do zespołu zajmującego się projektem TEI Simple¹⁶. Spośród jego celów na prowadzenie szybko wybił się pomysł stworzenia uniwersalnego modelu przetwarzania dokumentów XML wyrażonego w TEI. Główną ideą przyświecającą autorom było z jednej strony przywrócenie kontroli nad publikacją edytorom dzięki obniżeniu bariery technicznej, a z drugiej – krok w stronę „długiego trwania” dzięki stworzeniu standaryzowanej i możliwej do automatycznego przetwarzania formy wyrażania zamysłów edytorskich.

**Pierwsze próby stworzenia
prototypu nowej wersji systemu
dla *Projektu Dantyszek*
zostały podjęte w 2013 roku**

Owoc tego projektu pod nazwą TEI Processing Model został oficjalnie częścią standardu TEI w 2016 roku. Jeszcze zanim to nastąpiło, TEI Processing Model stał się kamieniem węgielnym projektu *open source* pod nazwą TEI Publisher¹⁷ przenoszącego te same idee *standards, interoperability, reusability and sustainability*, czyli w wolnym tłumaczeniu: „wykorzystanie standardów, współdziałanie systemów, ponowne użycie i zrównoważony rozwój”, z poziomu pojedynczego dokumentu do kompletnej aplikacji internetowej.

TEI Processing Model i TEI Publisher

Punktem wyjścia dla projektu TEI Simple była konstatacja, że większość badaczy swobodnie pracujących z dokumentami w standardzie TEI, gdy przychodzi do ich publikacji, musi ustąpić pola profesjonalistom z branży IT. Jednocześnie bardzo duża część wymagań co do prezentacji materiału w formie internetowej, drukowanej czy e-booka jest wspólna dla różnych projektów. Nie ma zatem powodu, dla którego edytorzy nie mogliby powszechnie korzystać z przygotowanej przez społeczność TEI bazowej specyfikacji dotyczącej przetwarzania dokumentów w tym standardzie.

Specyfikacja taka została stworzona w ramach projektu TEI Simple i udostępniona razem z prototypową implementacją TEI Processing Model. Istotne dla popularyzacji tego podejścia okazały się dwa czynniki: po pierwsze, TEI Processing Model wywodzi się z idei *literate programming*¹⁸ i jego reguły są wyrażone w TEI, czyli języku, którym edytorzy biegle posługują się w swojej codziennej praktyce; po drugie, w razie pojawienia się specyficznych potrzeb projektu edytorzy mogą te reguły dopasować do swoich wymagań nawet przy ograniczonych kompetencjach technicznych.

Jest bardzo istotne, że dokument ODD¹⁹ zawierający owe reguły przetwarzania stanowi dokładną dokumentację dotyczącą edytorskich zamysłów co do publikacji. Na podstawie jedynie ODD oraz danych źródłowych z dokumentów w TEI możemy zrekonstruować koncepcję publikacji nawet bez oprogramowania, co ma ważne znaczenie dla perspektywy długiego trwania zakończonych projektów.

Oczywiście pojedynczy dokument to jeszcze niekoniecznie kompletna edycja. Współtworzony przez jedną z autorek system publikacji pod nazwą TEI Publisher powstał jako rozszerzenie podejścia zapoczątkowanego przez TEI Processing Model w kontekście aplikacji internetowych. Podobnie jak reguły transformacji, również elementy funkcjonalne składające

się na publikację online są w dużym zakresie wspólne dla różnych projektów. TEI Publisher zatem proponuje, podobnie jak wcześniej TEI Simple, bazę, na której można się oprzeć, a jednocześnie dostosować wszystkie elementy do potrzeb konkretnej edycji. Dzięki modułowej konstrukcji proces tworzenia aplikacji odbywa się szybko i sprawnie, przy znacznie ograniczonych nakładach. Jednocześnie taki system nie tylko jest łatwiejszy w utrzymaniu dzięki minimalnej „nadbudowie” nad standardową bazą, ale dokłada następną samodokumentującą cegielkę, istotną dla perspektywy długiego trwania: szablony stron wyrażone w czytelnym kodzie HTML z wykorzystaniem komponentów z biblioteki TEI Publisher uzupełniają informacje z dokumentów źródłowych i ODD, a w efekcie umożliwiają pełną rekonstrukcję edycji, nawet w zupełnie odmiennym krajobrazie technologicznym, którego możemy się spodziewać w przyszłości.

Plany na przyszłość i długie trwanie

Sporo miejsca w tym artykule poświęciliśmy wyzwaniom związanym z tworzeniem, utrzymaniem i niezbędnymi migracjami systemów informatycznych. Wiedząc z całą pewnością, że zmiany są i pozostaną nieuniknione, jesteśmy przekonani, że tylko oparcie się na standardach pozwala na stosunkowo bezbolesne, najłatwiejsze (i najtańsze) przetrwanie projektu i zapewnienie jego dostępności w długiej perspektywie.

Z tego punktu widzenia migracja danych do standardu TEI/XML była krokiem milowym, który pozwoli na ewentualną rekonstrukcję rezultatów pracy zespołu edytorskiego, nawet gdyby wszystko inne przepadło. Co jednak rozumieliśmy, nasze aspiracje się tutaj nie kończą: chcielibyśmy widzieć *Korpus Tekstów i Korespondencji Jana Dantyszka* jako żywą, dostępną szerokiej publiczności i stale aktualizowaną kolekcję. Oparcie systemu publikacji na bazie TEI Publisher i rekomendacjach e-editiones²⁰ pozwala nam uzasadnić przekonanie, że przyszłe aktualizacje, integracja z innymi systemami czy rozbudowa zakresu funkcjonalnego publikacji internetowej będą przebiegały w optymalny sposób. Już teraz widzimy, że materiał zebrany w formacie TEI/XML w żaden sposób nie zestarzał się technologicznie, w szczególności może być bez zmian opublikowany poza systemem, dla którego został pierwotnie stworzony. Z tego względu priorytetem jest teraz pełna migracja danych do XML w celu zagwarantowania pełnej zrekonstruowalności naszej publikacji oraz możliwości współpracy z innymi systemami również w zakresie metadanych.

W spektrum *Projektu Dantyszek* znajdują się poza tym:

- integracja z systemem automatycznego rozpoznawania pisma ręcznego Handwritten Text Recognition (HTR, Transkribus);
- integracja z zewnętrznymi bazami danych dotyczącymi osób (internetowe słowniki biograficzne) i miejsc (*gazetteers*);
- integracja z innymi edycjami korespondencji²¹.

Key Words: Ioannes Dantiscus, Renaissance correspondence, TEI, XML, long lasting of digital scholarly editions

Abstract: The article is devoted to the works on the corpus constituted by the correspondence (over 6000 preserved letters, mainly in Latin and German) and other texts by Ioannes Dantiscus (1485–1548), a humanist and neo-Latin poet, a diplomat in the service of Polish rulers, bishop of Chełmno and Warmia. Since the 80's of the 20th century, the works have been carried out at Warsaw University under the project entitled *Registration and Publication of Ioannes Dantiscus' Correspondence (1485–1548)*, which involves online publication of the whole body of material in the form of *Corpus of Ioannes Dantiscus' Texts & Correspondence / Korpus Tekstów i Korespondencji Jana Dantyszka* (CIDTC) and a publication of the whole inventory of correspondence and a selection of letters in a printed series “Corpus Epistularum Ioannis Dantisci” (CEID), which has been appearing since 2004. The article primarily discusses issues related to the creation, maintenance and necessary migrations of information systems used in the project (from Fox Pro, through Access and MySQL to TEI/XML).

obejmowały również system publikacji internetowej, edycji transkrypcji w formacie XML oraz przygotowania materiału do publikacji drukiem.

¹⁰ Szczególnie wiele zawdzięczamy tutaj wytycznym do publikacji korespondencji projektu *DALF: Digital Archive of Letters in Flanders* (2003), a w późniejszym czasie również publikacji *Listów Vincenta Van Gogha* (2009).

¹¹ Popularna platforma serwerowa dla dynamicznych stron WWW obejmująca w przypadku *Dantyszka* zestaw Linux–Apache–MySQL–PHP wzbogacony procesorem XSLT – oprogramowaniem do przetwarzania dokumentów w formacie XML.

¹² Zob. diagram przedstawiający w ujęciu chronologicznym rozwój standardów dotyczących XML i Semantic Web – *XML and Semantic Web W3C Standards Timeline* w: N. Bikakis et al., *The XML and Semantic Web Worlds: Technologies, Interoperability and Integration. A Survey of the State of the Art*, <http://www.dblab.ntua.gr/~bikakis/XMLSemanticWebW3CTimeline.pdf> (dostęp: 5.06.2023).

¹³ XML i standardy pokrewne: XPath, XQuery, XSLT.

¹⁴ Platforma programistyczna, częściej z ang. *framework*, to system tworzenia aplikacji, który definiuje jej strukturę i ogólny mechanizm jej działania, a także dostarcza zestaw komponentów do wykonywania określonych zadań. Programistka rozbudowuje i dostosowuje poszczególne komponenty do wymagań realizowanego projektu, tworząc tak gotową aplikację.

¹⁵ *eXist: a NoSQL Document Database and Application Platform*, <http://www.exist-db.org/> (dostęp: 5.06.2023). Zob. też: *BaseX: The XML Framework*, <https://basex.org/> (dostęp: 5.06.2023).

¹⁶ *TEI Simple*, <https://teic.github.io/TEI-Simple/> (dostęp: 5.06.2023).

¹⁷ *TEI Publisher – The Instant Publishing Toolbox*, <https://teipublisher.com> (dostęp: 5.06.2023).

¹⁸ *Literate programming* (programowanie piśmienne) to styl programowania oparty na założeniu, że programy komputerowe powinny być pisane z naciskiem na czytelność kodu źródłowego dla ludzi, podobnie do dzieła literackiego.

¹⁹ ODD: *One Document Does It All*, specyficzny rodzaj dokumentu TEI, który zgodnie z ideą *literate programming* łączy opisową dokumentację projektu z formalną specyfikacją schematu znakowania (ang. *schema*) oraz regułami przetwarzania (*TEI Processing Model*).

²⁰ e-editiones – międzynarodowe stowarzyszenie zrzeszające indywidualnych badaczy i instytucje naukowe oraz kulturalne, którego celem jest wspieranie edycji naukowych przez budowanie i promowanie otwartych standardów, oprogramowania, koordynowanie współpracy między projektami i tworzenie rekomendacji dotyczących najlepszych praktyk, <https://e-editiones.org/> (dostęp: 5.06.2023).

²¹ W 2021 roku podstawowe metadane dotyczące całej zachowanej korespondencji *Dantyszka* wraz z linkami do naszego korpusu zostały zdeponowane w bazie *Early Modern Letters Online*, <http://emlo.bodleian.ox.ac.uk/> (dostęp: 5.06.2023).

¹ R. Marciniak, *Acta Tomiciana w kulturze politycznej Polski okresu Odrodzenia*, Warszawa–Poznań 1983.

² S. Grzybowski, *Teki Naruszewicza: „Acta regum et populi Poloni”*, Wrocław 1960.

³ A. Skolimowska, *Once again about the Origin and Afterlife of Files H.154 and H.155, Housed at the Uppsala University Library*, w: *Respublica Litteraria in Action. New Sources*, ed. K. Tomaszuk, Warsaw–Cracow 2016, series “Corpus Epistularum Ioannis Dantisci”, Vol. 5/3, s. 11–50.

⁴ *Acta Tomiciana*, Vol. 1–18, 1852–1999.

⁵ Biblioteka PAU–PAN, zbiór rękopisów, sygn. 8239–8251; A. M. Kurpiel, *Jan Dantyszek, życie i działalność. Cz. I. (1485–1519)*, w: *Sprawozdania z czynności i posiedzeń PAU 1910*, Kraków 1911, s. 3–8.

⁶ S. Skimina, *Twórczość poeticka Jana Dantyszka*, Kraków 1948, seria „Rozprawy Polskiej Akademii Umiejętności. Wydział Filologiczny” 68/1; *Ioannis Dantisci poetae laureati Carmina*, wyd. S. Skimina, Kraków 1950, seria „Corpus Antiquissimorum Poetarum Poloniae Latinorum” 7.

⁷ *Corpus of Ioannes Dantiscus' Texts & Correspondence / Korpus Tekstów i Korespondencji Jana Dantyszka*, dantiscus.al.uw.edu.pl, eds. A. Skolimowska, M. Turska, with collaboration of K. Jasińska-Zdun, publikacja w trakcie realizacji, pierwsza publikacja w internecie: 1.07.2010.

⁸ Ta wartość nie uwzględnia tabel przechowujących dane „słownikowe”, czyli możliwe wartości dla pól wyboru, np. dla pola „typ źródła” będą to: *oryginał*, *brudnopis* lub *kopia kancelaryjna*.

⁹ Trzeba zauważyć, że migracja z 1998 roku dotyczyła jedynie opracowania nowego schematu bazy danych, konwersji danych i przygotowania formularzy do edycji metadanych w systemie MS Access. Przygotowania do uruchomienia nowej wersji systemu w 2010 roku