ROCZNIKI HUMANISTYCZNE
Tom LXXIII, zeszyt 2 – 2025
ZESZYT SPECJALNY
DOI: https://doi.org/10.18290/rh25732.1s

BOGUMIŁ SZADY
ADAM ZAPAŁA

MODELOWANIE DANYCH W EDYTORSTWIE CYFROWYM*

Abstrakt. Artykuł analizuje znaczenie modelowania danych w edytorstwie cyfrowym, wskazując, że w badaniach historycznych dane zawarte w źródle są równie istotne jak sam tekst. Autorzy podkreślają, że standard TEI, choć dobrze sprawdza się przy opisie struktury i wariantów tekstu, jest niewystarczający do pełnego zamodelowania narracji źródła i odtworzenia relacji między faktami, osobami, miejscami czy wydarzeniami. Z tego względu proponują rozwiązania oparte na połączeniu standardu TEI XML z bazami relacyjnymi i grafowymi, które umożliwiają precyzyjniejsze odwzorowanie treści komunikatu źródłowego oraz jego warstwy informacyjnej. W artykule zaprezentowano przykłady zastosowań tych metod, podkreślając rolę edytora jako interpretatora oraz konieczność budowania modeli, które łączą wierne odwzorowanie warstwy znaku z możliwością wydobycia informacji istotnych dla badań historycznych.

Słowa kluczowe: modelowanie danych; naukowe edytorstwo cyfrowe; źródła historyczne; TEI (Text Encoding Initiative); bazy danych (RDF, SQL); dane badawcze

DATA MODELING IN DIGITAL SCHOLARLY EDITING

Abstract. The article analyzes the importance of data modeling in digital editing, emphasizing that in historical research the data contained in a source are just as crucial as the text itself. The authors argue that while the TEI standard is well suited for describing textual structure and variants, it is insufficient for fully modeling the narrative of a source and reconstructing the relations between

Dr hab. BOGUMIŁ SZADY – Ośrodek Badań nad Geografią Historyczną Kościoła w Polsce, Instytut Historii, Katolicki Uniwersytet Lubelski Jana Pawła II; adres do korespondencji: Al. Racławickie 14, 20-950 Lublin; e-mail: bogumil.szady@kul.pl; ORCID: https://orcid.org/0000-0003-0059-5596.

Dr ADAM ZAPAŁA – Instytut Historii im. Tadeusza Manteuffla Polskiej Akademii Nauk; adres do korespondencji: Instytut Historii im. Tadeusza Manteuffla PAN, Rynek Starego Miasta 29/31, 00-292 Warszawa; e-mail: azapala@ihpan.edu.pl; ORCID: https://orcid.org/0000-0003-3450-8286.

*Praca naukowa dofinansowana ze środków budżetu państwa w ramach programu realizowanego przez Ministra Edukacji i Nauki pod nazwą Narodowy Program Rozwoju Humanistyki (nr projektu NPRH/DN/SP/495771/2021/10).

facts, persons, places, or events. Therefore, they propose solutions based on on combining the TEI XML standard with relational and graph databases, which allow for a more precise representation of both the communicative content of the source and its informational layer. The article presents examples of applying these methods, highlighting the editor's role as an interpreter and the need to develop models that combine faithful representation of the textual layer with the extraction of information essential for historical research.

Keywords: data modeling; scholarly digital editing; historical sources; TEI (Text Encoding Initiative); databases (RDF, SQL); research data

WSTĘP. SPECYFIKA EDYTORSTWA CYFROWEGO

Wprowadzenie do edytorstwa źródeł historycznych formy cyfrowej oznacza nie tylko zmianę medium przekazu[1], lecz także dogłębną modyfikację metody opracowania źródła. Nie należy przy tym stawiać edytorstwa cyfrowego (elektronicznego) w opozycji do edytorstwa tradycyjnego (drukowanego). Edytorstwo cyfrowe jest twórczym rozwinięciem edytorstwa tradycyjnego. Do obowiązków edytora „tradycyjnego" należy odczytanie, zrozumienie i udostępnienie przekazu źródła wraz z towarzyszącym mu aparatem krytycznym (wstęp źródłoznawczy, przypisy, indeksy). Zadania te dotyczą także edycji cyfrowej, ale są realizowane przez komputerowe przetworzenie treści źródła do formy ustrukturyzowanej. Oznacza to przekształcenie treści źródła historycznego w zbiór danych badawczych w postaci cyfrowej[2]. Przetworzenie to może przyjąć różne formy, np. relacyjnej bazy danych, grafu wiedzy, bazy zaanotowanych dokumentów w formacie XML, musi jednak zawierać reprezentację źródła oraz jego krytyczne opracowanie[3]. Kluczowe znaczenie w tej kwestii ma łączenie tekstu źródła z danymi pozyskanymi na drodze jego semantycznego modelowania. Choć temat ten jest obecny w dyskusji naukowej od przeszło dekady, wciąż nie wypracowano rozwiązania, które mogłoby zostać przyjęte za standardowe. Celem niniejszego artykułu jest określenie miejsca modelowania danych we współczesnym edytorstwie

[1] O cechach paradygmatu cyfrowego w edytorstwie: Patrick Sahle, „What is Scholarly Digital Edition?", w *Digital Scholarly Editing. Theories and Practices*, red. Matthew James Driscoll, Elena Pierazzo (Cambridge: Open Book Publishers, 2016), 28–33, https://www.openbookpublishers.com/books/10.11647/obp.0095.

[2] Pełniejsze informacje na temat definicji danych badawczych w humanistyce zob. raport: Natalie Harrower, Maciej Maryl, Timea Biro, Beat Immenhauser, red., *ALLEA. Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities* (Berlin: ALLEA, 2020), 8–10; Maciej Maryl, Marta Błaszczyńska, Bartłomiej Szleszyński, Tomasz Umerle, „Dane badawcze w literaturoznawstwie", *Teksty Drugie. Teoria literatury, krytyka, interpretacja* 2 (2021): 13–22.

[3] Sahle, „What is Scholarly Digital Edition?", 23–5.

cyfrowym rozumianym jako proces badawczy. Służy temu dyskusja nad istniejącymi rozwiązaniami teoretycznymi oraz praktycznymi doświadczeniami w tym zakresie. Nie jest ambicją autorów sformułowanie jednoznacznych rekomendacji czy wskazanie najlepszych narzędzi i metod w modelowaniu danych towarzyszących edytorstwu źródeł historycznych. Tekst powinien być traktowany przede wszystkim jako głos w dyskusji nad ważnym i coraz ważniejszym – w opinii autorów – wyzwaniem, jakim jest sposób gromadzenia i przygotowania danych powstających w procesie edycji źródła, tak aby stały się one pełnowartościowymi danymi badawczymi, które mogą być wykorzystane przez historyków.

Choć słowo „modelowanie" często kojarzy się z pracą w środowisku cyfrowym, to ma ono także szersze, bardziej uniwersalne znaczenie. W ujęciu Jerzego Topolskiego modelowanie oznaczało konstruowanie obrazu przeszłej rzeczywistości jako jej reprezentacji, odzwierciedlającej cechy szczególne opisywanego zjawiska[4]. Zadaniem historyka było więc „rzeźbienie" lub „malowanie" tego obrazu za pomocą wiedzy źródłowej i pozaźródłowej. To teoretyczne i nieco abstrakcyjne ujęcie nabiera konkretnego i bardzo praktycznego znaczenia, gdy używamy komputera, czyli maszyny, której działanie opiera się na wykorzystaniu ustrukturyzowanych danych. W tym przypadku modelowanie przestaje być metaforycznym „malowaniem" i staje się umiejętnością przełożenia rzeczywistości na strukturę rozumianą oraz przetwarzaną przez „maszynę liczącą". Każde wykorzystanie komputera wiąże się z modelowaniem danych (nawet praca w najbardziej standardowym edytorze tekstu). Kwestią wielkiej wagi jest uświadomienie badaczom tego faktu i wypracowanie standardów, które pozwoliłyby robić to w naukowy sposób.

W środowisku cyfrowym język naturalny, czyli system komunikacji właściwy ludziom, nie jest najbardziej efektywnym sposobem przekazywania informacji. Dużo bardziej jednoznaczne i skuteczniejsze jest wykorzystanie języka w postaci ustrukturyzowanej. Oddzielenie znaczenia tekstu od jego tekstualnej formy pozwala na precyzyjniejsze opisywanie informacji, daje większe możliwości ich prezentacji, a także ułatwia ich zrozumienie i analizowanie. Odejście od linearnej formy tekstu książki i przetwarzanie tekstu w formie bazodanowej umożliwia definiowanie powiązań i relacji między poszczególnymi jego częściami, a także łączenie ich z informacjami pochodzącymi spoza źródła. Pozwala też na nakładanie na siebie wielu warstw informacji oraz na różnorodne sposoby prezentacji tego samego tekstu (np. w formie skróconej, rozszerzonej czy nawet w formie grafu).

Przekształcenie tekstu w dane wymaga od edytora wprowadzenia do tekstu dodatkowych znaczników niezwiązanych z jego naturalną formą. Edytor, który

[4] Jerzy Topolski, *Teoria wiedzy historycznej* (Poznań: Wydawnictwo Poznańskie, 1983), 255.

określa model danych i znaczniki oraz sposób ich wykorzystania, staje się więc pierwotnym interpretatorem wydawanego tekstu i ważnym pośrednikiem między tekstem a jego czytelnikiem. Odbiorca edycji powinien mieć zawsze świadomość, że zapoznaje się z tekstem zamodelowanym przez wydawcę. Cecha ta zwiększa znaczenie aparatu krytycznego (w rozumieniu interpretacji tekstu wyrażonej w formie danych) względem edycji tradycyjnej. Funkcjonalność edycji zależy bowiem właśnie od sposobu zamodelowania danych przez edytora i odzwierciedla przewidziane przez niego zainteresowania badawcze.

Modelowanie danych w edytorstwie cyfrowym mieści się w ogólnym zakresie modelowania danych humanistycznych. W tradycji polskiej i europejskiej można wyróżnić dwa przenikające się nurty: językowy, reprezentowany przez przedstawicieli nauk filologicznych i wydawców źródeł literackich, oraz historyczny, charakterystyczny dla badaczy przeszłości. Pierwszy nurt koncentruje się na modelowaniu języka i tekstu źródeł, natomiast drugi – na odtwarzaniu faktów i zjawisk historycznych przez nie opisanych. Literatura przedmiotu dotycząca edytorstwa cyfrowego jest w dużej mierze zdominowana przez podejście filologiczne. W konsekwencji problematyka dostosowania edytorstwa cyfrowego do potrzeb historyków jest często marginalizowana. Z jednej strony w opracowaniach dotyczących edytorstwa cyfrowego kwestia modelowania narracji źródła czy też opisanej w tekście rzeczywistości pojawia się sporadycznie i pozostaje na marginesie głównego nurtu dyskusji[5]. Z drugiej zaś większość przedsięwzięć edycji źródeł historycznych, nawet jeśli są one upowszechniane w Internecie, ignoruje możliwości narzędzi cyfrowych[6], co poważnie szkodzi całej dziedzinie.

[5] M.in. Justyna Rolińska, „Jak tworzyć cyfrowe edycje źródeł historycznych? Kilka uwag do nowej instrukcji wydawniczej", w *Edytorstwo źródeł od instrukcji do edycji*, red. Adam Perłakowski, t. 4 (Kraków: Towarzystwo Wydawnicze Historia Iagellonica, 2022); José Luis Losada Palenzuela, „Podstawy naukowej edycji cyfrowej", w *Od Gutenberga do Zuckerbergo. Wstęp do humanistyki cyfrowej*, red. Adam Pawłowski (Kraków: Universitas, 2023), 173–99; Anna Skolimowska, „Korespondencja Jana Dantyszka jako laboratorium edytorstwa cyfrowego", w *Człowiek twórcą historii*, t. 5: *Warsztat nowoczesnego humanisty historyka na progu XXI w.,* cz. 1, red. Cezary Kuklo, Wojciech Walczak (Białystok: Uniwersytet w Białymstoku, 2024), 71–96.

[6] Przykładem tego typu przedsięwzięcia jest monumentalna seria wydawnicza Folia Jagiellonica. Fontes, która zawiera kilkadziesiąt tomów źródeł opracowanych w tradycyjny sposób. Ten cenny materiał został przygotowany w bardzo krótkim czasie, w ramach jednego przedsięwzięcia. Rzuca on nowe światło na wiele aspektów epoki jagiellońskiej, jednak jego potencjał badawczy jest zasadniczo ograniczony ze względu na brak możliwości wspólnego przeszukiwania zasobu oraz filtrowania tekstów ze względu na zdefiniowane kryteria. Stworzenie nawet bardzo prostej bazy danych łączącej podstawowe metadane i indeksy wszystkich tomów oraz definiującej jedynie słowa kluczowe poszczególnych tekstów i ich fragmentów, znacząco ułatwiłoby korzystanie z tych edycji, a co za tym idzie – dałoby możliwość prowadzenia przekrojowych badań na tym materiale, co w obecnej formie jest w zasadzie niemożliwe. W podręczniku do edytorstwa źródeł historycznych z 2014 r. kwestie

Edycja cyfrowa zawiera bowiem wiele funkcjonalności niemożliwych do oddania w druku, a jednocześnie wysiłek „tradycyjnych" edytorów nie wpływa na udostępnienie ich pracy jako danych badawczych, które mogłyby być w stosunkowo łatwy sposób analizowane oraz porównywane z innymi bazami wiedzy.

MODELOWANIE DANYCH HUMANISTYCZNYCH – DEFINICJA, KONTEKST I CEL

Modelowanie danych humanistycznych, np. historycznych, było zawsze obecne w badaniach naukowych, także w tzw. epoce przedcyfrowej. Najwyraźniej widać to w rozwoju kartografii historycznej, gdyż u podstaw każdej mapy historycznej jako metody reprezentacji przeszłości leżał model konceptualny. Jego odzwierciedleniem jest legenda mapy. Wiele analiz z zakresu historii społecznej, w których autorzy korzystali z metod statystycznych (demografia historyczna, historia gospodarcza), opierało się na odpowiednich strukturach danych. Wprowadzenie metod i narzędzi cyfrowych wyraźnie rozwinęło i pogłębiło metodykę badań historycznych w tych obszarach. Refleksja metodyczna objęła z jednej strony podstawy teoretyczne i praktyczne modelowania danych, z drugiej zaś weryfikację i uściślenie pojęć funkcjonujących w istniejących modelach. Za przykład może posłużyć proces przetworzenia ustaleń Stanisława Litaka na temat sieci parafialnej Kościoła łacińskiego w Rzeczypospolitej przed I rozbiorem Polski do postaci bazy danych. Model danych wypracowany przez Litaka w latach 70. XX wieku wpisywał się w cały szereg prób o podobnym charakterze realizowanych w Instytucie Geografii Historycznej Kościoła w Polsce przy Katolickim Uniwersytecie Lubelskim[7]. Litak publikował wyniki badań nad strukturą parafialną epoki przedrozbiorowej w podobnym, ale nie tożsamym modelu danych w kolejnych monografiach, które zawierały wykazy opracowane według określonej struk-

cyfrowe ograniczają się do opisu wykorzystania narzędzi do automatyzacji, bez głębszej refleksji nad merytorycznymi zmianami wywołanymi pojawieniem się paradygmatu cyfrowego. Jerzy Tandecki, Krzysztof Kopiński, *Edytorstwo źródeł historycznych* (Warszawa: Wydawnictwo DiG, 2014), 248–66.

[7] Należy tutaj wspomnieć chociażby kwestionariusze dla gromadzenia danych badawczych w projektach takich, jak „Polonia Christiana" czy „Zakony męskie w Polsce w 1772 r.". Wyniki tych projektów ukazywały się w serii Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, np. *Zakony męskie w Polsce w 1772 roku*, red. Ludomir Bieńkowski, Jerzy Kłoczowski, Zygmunt Sułowski (Lublin: Towarzystwo Naukowe KUL, 1972) (Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, 1); Eugeniusz Wiśniowski, *Prepozytura wiślicka do schyłku XVIII wieku. Materiały do struktury organizacyjnej* (Lublin: Towarzystwo Naukowe KUL, 1976) (Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, 2.2); Lidia Müllerowa, *Sieć parafialna Kościoła katolickiego w Polsce w 1980 roku* (Lublin: Towarzystwo Naukowe KUL, 1982) (Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, 5).

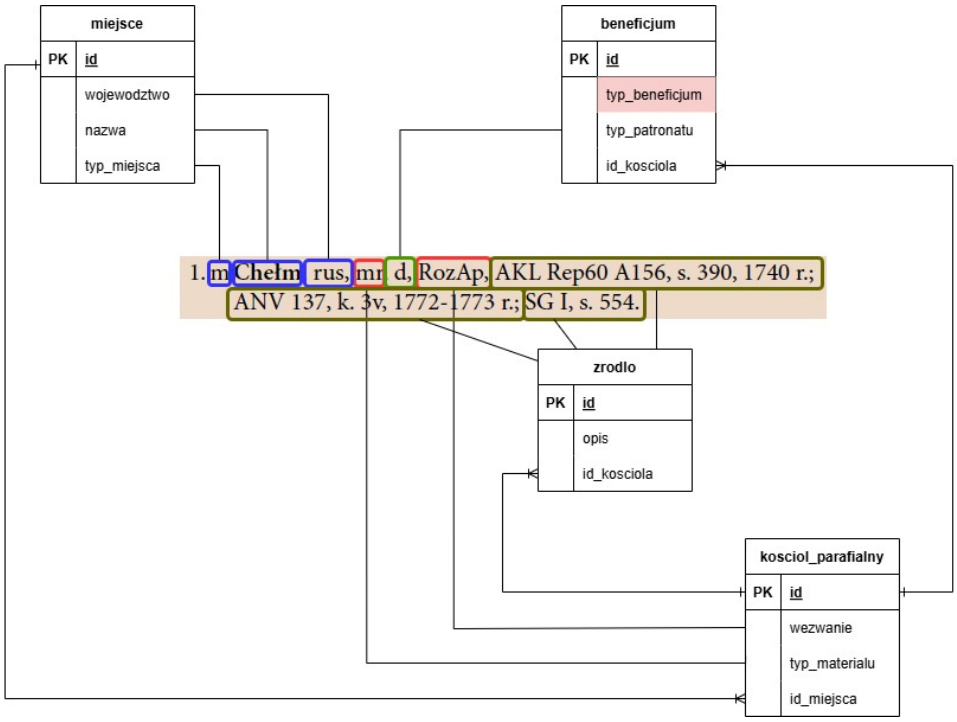
tury[8]. Zasadzała się ona na układzie hierarchicznym: metropolie, diecezje, archidiakonaty, dekanaty, a w ramach każdego dekanatu wykazy „kościołów parafialnych", a także „kościołów filialnych i innych nieparafialnych oraz kaplic publicznych". Wśród atrybutów każdego kościoła znajdowały się: typ i nazwa miejscowości, w której zlokalizowano świątynię, przynależność do województwa, typ materiału budowlanego, wezwanie kościoła oraz podstawa źródłowa. W przypadku kościołów filialnych po wezwaniu podawano również przynależność parafialną. Publikacje te zawierały ponadto indeksy miejscowości oraz wezwań.

W trakcie przekształcania wykazów Litaka do modelu danych, przygotowywanego na potrzeby relacyjnej bazy danych, pojawiła się kontrowersja dotycząca definicji samego przedmiotu wykazu – czy wykaz zawiera informacje odnoszące się do kościołów, czy do parafii. Drugim problemem okazał się sposób rejestracji podstawy źródłowej: przyjęty model zakładał zbiorcze podanie podstawy informacyjnej, co uniemożliwiało określenie podstawy dla poszczególnych atrybutów: wezwania, typu materiału budowlanego czy rodzaju prawa patronatu. Dodatkowo ten ostatni element nie był związany z kościołami, ale z beneficjami funkcjonującymi przy kościołach. Niektóre kościoły, jak kolegiaty czy katedry, posiadały wiele beneficjów, a wiele kościołów filialnych nie posiadało ich wcale. Spowodowało to, że typ patronatu pojawia się w wykazie tylko przy kolegiatach, gdzie prebendy kapitulne miały jednolity charakter patronatu, np. w Zamościu (szlachecki), pominięty zaś został przy kolegiatach, gdzie prebendy miały różnych patronów (szlachtę, miasto, króla, duchowieństwo), np. w Lublinie. Na pozór prosty i jasny drukowany wykaz w chwili przekształcania w bazę danych już na podstawowym poziomie – tylko dla kościoła parafialnego, bez przynależności do jednostek administracji kościelnej – zaczął się przeradzać w dość złożoną strukturę (ryc. 1). Model ujawnił też problem prawa patronatu jako zjawiska związanego z beneficjum, a nie kościołem, oraz mankament wiążący się ze sposobem zapisu podstawy źródłowej[9]. W docelowym schemacie i przy obecnym stanie wiedzy, dzięki projektowi Internetowy Atlas

[8] Stanisław Litak, *Struktura terytorialna Kościoła łacińskiego w Polsce w 1772 roku* (Lublin: Towarzystwo Naukowe KUL, 1980) (Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, 4); Stanisław Litak, *Kościół łaciński w Rzeczypospolitej około 1772 roku. Struktury administracyjne* (Lublin: Wydawnictwo KUL, 1996) (Wspólnoty religijne i narodowe w Rzeczypospolitej w drugiej połowie XVIII wieku, 1); Stanisław Litak, *Atlas Kościoła łacińskiego w Rzeczypospolitej Obojga Narodów w XVIII wieku* (Lublin: Towarzystwo Naukowe KUL, 2006) (Prace Instytutu Geografii Historycznej Kościoła w Polsce KUL, 10).

[9] Paweł Garbacz, Robert Trypuz, Bogumił Szady, Piotr Kulicki, Przemysław Grądzki, Marek Lechniak, „Towards a Formal Ontology for History of Church Administration", w *Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010)*, red. Antony Galton, Riichiro Mizoguchi (Amsterdam: IOS Press, 2010), 345–58 (Frontiers in artificial intelligence and applications, t. 209).

Polski Niepodległej, ostatnie z wymienionych na Rycinie 1 źródeł mogłoby zostać powiązane z bazą danych *Słownika geograficznego Królestwa Polskiego i innych krajów słowiańskich* (https://atlas1918.ihpan.edu.pl/)[10]. Nasuwa to naturalne pytanie o rolę modelowania danych w edycjach cyfrowych i – szerzej – w warsztacie badawczym historyka, który dokumentuje w bazach danych zjawiska o charakterze masowym i regularnym: spisy osób, urzędów, miejscowości, jednostek administracji państwowej i świeckiej itd.



Ryc. 1. Przykładowy model danych dla wykazu kościołów opublikowanego w pracy: Litak, *Atlas Kościoła łacińskiego w Rzeczypospolitej Obojga Narodów w XVIII wieku*, 10.

Źródło: Oprac. własne

[10] Projekt „Internetowy Atlas Polski Niepodległej" (numer projektu 01SPN 17 0008 18) pod kierunkiem Tadeusza Epszteina został zrealizowany w latach 2018-2021 przez Instytut Historii im. Tadeusza Manteuffla PAN oraz Wydział Geodezji i Kartografii Politechniki Warszawskiej dzięki finansowaniu w ramach programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „Szlakami Polski Niepodległej".

Niemal wszyscy specjaliści zajmujący się gromadzeniem danych humanistycznych odwołują się do ogłoszonych w 2016 r. i wciąż obowiązujących zasad FAIR (*Findable*, *Accessible*, *Interoperable* i *Reusable*)[11]. Doświadczenie ostatnich lat wskazuje, że nie da się w pełni stosować żadnej z zasad opisanych przez Marka D. Wilkinsona i jego współpracowników bez namysłu nad modelem i strukturą danych. Świadczy o tym szereg prac teoretycznych i praktycznych wysiłków, w szczególności na polu budowania ontologii dziedzinowych, słowników kontrolowanych oraz taksonomii dotyczących zjawisk społecznych i humanistycznych. W tym samym roku, w którym na łamach „Nature. Scientific Data" ogłoszono manifest na temat zasad FAIR, ukazały się także dwa ważne teksty dotyczące modelowania danych humanistycznych. Pierwszy, *Data Modelling*, autorstwa Julii Flanders i Fotisa Jannidisa, został opublikowany w podręczniku do humanistyki cyfrowej[12]. Jego rozwinięcie można znaleźć w dwóch pierwszych rozdziałach pracy zbiorowej poświęconej modelowaniu danych w humanistyce cyfrowej, opublikowanej w roku 2020[13]. Na podstawie tych artykułów, ale też innych opracowań z obszaru *digital humanities*, modelowanie danych należy rozumieć jako proces tworzenia schematu pojęciowego, który umożliwia opis i organizację danych przez definiowanie istotnych pojęć oraz relacji między nimi. W humanistyce modelowanie jest w znacznej mierze aktem interpretacyjnym – odzwierciedlającym sposób rozumienia i konceptualizacji badanych zjawisk. Modelowanie danych realizowane jako proces w świecie cyfrowym (komputerowym) opiera się na założeniu, że jego wynikiem będzie zawsze rodzaj niedoskonałej, uproszczonej i subiektywnej reprezentacji, otwartej na modyfikację (manipulację, zmianę) i ciągłą korektę wynikającą z rozwoju wiedzy o opisywanych zjawiskach[14]. Tymczasowość modelu jest wpisana w jego naturę, za to oczekiwanie i próba zbudowania ostatecznego i skończonego modelu

[11] Mark D. Wilkinson, i in., „The FAIR Guiding Principles for Scientific Data Management and Stewardship", *Scientific Data* 3, nr 1 (2016): 1–9.

[12] Julia Flanders, Fotis Jannidis, „Data Modeling", w *A New Companion to Digital Humanities*, red. Susan Schreibman, Ray Siemens, John Unsworth (Malden, MA–Chichester, West Sussex, UK: Wiley-Blackwell, 2016), 229–37.

[13] Julia Flanders, Fotis Jannidis, „Data Modeling in a Digital Humanities Context: An Introduction", w *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*, red. Julia Flanders, Fotis Jannidis (London–New York: Routledge, 2020), 1–20 (Digital Research in the Arts and Humanities); Julia Flanders, Fotis Jannidis, „A Gentle Introduction to Data Modeling", w *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*, red. Julia Flanders, Fotis Jannidis (London–New York: Routledge, 2020), 26–96. (Digital Research in the Arts and Humanities).

[14] Willard McCarty, „Humanities Computing" (Basingstoke–New York: Palgrave Macmillan, 2014), 26–7.

danych jakiegoś zjawiska przypomina iluzoryczną chęć napisania kompletnej historii wypraw krzyżowych lub podania ostatecznej interpretacji poezji Cypriana Kamila Norwida.

W 2016 r. problem modelowania danych humanistycznych podjęli Arianna Ciula i Øyvind Eide w artykule „Modelling in Digital Humanities”[15]. Uznali oni, że już sam proces tworzenia modelu danych humanistycznych ma walory eksploracyjne i przedstawili go w perspektywie semiotycznej. W ich rozważaniach można odnaleźć wiele treści wspólnych z dawniejszą refleksją Brygidy Kürbis na temat źródłoznawstwa: „Tekst rozpatrywany jako znak i jako wypowiedź jest nosicielem i przekazicielem więcej niż jednej warstwy znaczeń. Podstawowe znaczenie odsyła do znaczenia drugiego, symbolicznego, które «widzi się» za pośrednictwem samej warstwy informacyjnej”[16]. Podejście Ciuli i Eide jest bliskie założeniu, że zadaniem modelowania nie jest przede wszystkim przedstawienie relacji między faktami (osobami, miejscami, zjawiskami) opisanymi przez źródło, ale zrozumienie i pogłębienie wiedzy o samym przekazie źródłowym[17]. Ich refleksja stanowi zatem ważny punkt odniesienia dla modelowania danych w pracach edytorskich. Ponieważ model danych traktują jako platformę komunikacji między twórcą a odbiorcą, kluczowym elementem tego procesu staje się strona pojęciowa i definicyjna modelu, obarczona kontekstem kulturowym i intelektualnym zarówno jego twórcy, jak i odbiorcy[18]. Modelowanie danych humanistycznych jest trudne, gdyż musi uwzględniać zderzenie subiektywizmu twórcy i odbiorcy modelu z potencjalnie obiektywną relacją (podobieństwo, identyczność itp.) łączącą reprezentację oraz przedmiot modelowania. Taka perspektywa uświadamia trudności w realizacji zasad FAIR, zwłaszcza w wymiarze I (interoperacyjności) oraz R (ponownego użycia) danych.

Powyższe uwagi skłaniają do postawienia pytania o kontekst i cel modelowania danych w badaniach humanistycznych. Rozdział dotyczący modelowania danych w cytowanym wcześniej podręczniku do humanistyki cyfrowej został umieszczony w części dotyczącej analizy danych. Możliwość szerokiego stosowania narzędzi cyfrowych w warsztacie badawczym humanisty skłania do poszerzenia tej perspektywy przez propozycję pracy w podejściu wielomodelowym, w którym odrębne – ale powiązane ze sobą – modele danych są budowane na

[15] Arianna Ciula, Øyvind Eide, „Modelling in Digital Humanities: Signs in Context”, *Digital Scholarship in the Humanities*, 32 (2016): i33–i46.

[16] Brygida Kürbis, „Metody źródłoznawcze wczoraj i dziś”, *Studia Źródłoznawcze* 24 (1979): 91.

[17] Ciula, Eide, „Modelling in Digital Humanities”, i33–i46.

[18] Bogumił Szady, „Inżynieria ontologiczna w badaniach geograficzno-historycznych”, w *Człowiek twórcą historii*, t. 5: *Warsztat nowoczesnego humanisty historyka na progu XXI w.*, cz. 1, red. Cezary Kuklo, Wojciech Walczak (Białystok: Uniwersytet w Białymstoku, 2024), 97.

etapie gromadzenia, analizy (przetwarzania, wzbogacania) oraz prezentacji (udostępniania) danych[19]. Dotyczy to zarówno fazy modelowania konceptualnego, logicznego, jak i fizycznego[20], przy czym najważniejsze i warunkujące przebieg całego procesu tworzenia struktur danych jest modelowanie konceptualne. W jego ramach następuje identyfikacja oraz zdefiniowanie elementów modelu, przede wszystkim klas, atrybutów, własności (relacji) oraz reguł. Należy pamiętać, że model konceptualny powinien być tworzony niezależnie od docelowego schematu bazy danych (relacyjnej, grafowej) oraz od planowanych rozwiązań implementacyjnych modelu. Budowa modelu konceptualnego powinna wynikać wyłącznie z założeń i celów realizowanego projektu czy badań naukowych.

W konstruowaniu modelu konceptualnego dla danych humanistycznych występuje kilka podejść, które usystematyzował w 2014 r. Eide[21]. Pierwsze z nich zakłada przede wszystkim wykorzystanie istniejących standardów oraz ontologii dziedzinowych do opisania swoich danych. Podejście takie ma oczywisty walor polegający na uzgodnieniu modelu swojego zbioru z innymi, już funkcjonującymi zbiorami zorganizowanymi wokół podobnej struktury danych. Wadą natomiast są ograniczenia istniejących modeli danych i ontologii dziedzinowych, które zmuszają ich użytkownika do daleko idących kompromisów przez wprowadzanie uproszczeń, pomijanie danych trudnych do ujęcia w istniejące schematy lub wzbogacanie (indywidualizowanie) istniejących modeli, co może zaburzać interoperacyjność modelu. Szczególne znaczenie w modelowaniu danych humanistycznych w takim podejściu mają ontologie dziedzinowe, np. powszechnie wykorzystywana dla opisu elementów dziedzictwa kulturowego ontologia CIDOC CRM, która ma wiele rozszerzeń tematycznych, takich jak LRMoo (informacja bibliograficzna) czy CRMgeo (model czasowo-przestrzenny)[22]. W drugim podejściu tworzony jest model danych dedykowany dziedzinie lub tekstowi (tekstom), których dotyczą badania, oraz celowi, jaki ma być realizowany za pomocą zaplanowanej struktury danych. Zaletą takiego podejścia

[19] Bogumił Szady, „Technologie cyfrowe w badaniach historycznych", w *Od Gutenberga do Zuckerberga: Wstęp do humanistyki cyfrowej*, red. Adam Pawłowski (Kraków: Universitas, 2023), 240.

[20] Modelowanie konceptualne polega na identyfikacji i opisie encji (bytów) oraz relacji między nimi, często przedstawianych za pomocą różnych diagramów encja-relacja. Modelowanie logiczne polega na definiowaniu struktur danych zgodnie z modelem relacyjnym (np. tabele, kolumny, relacje) lub nierelacyjnym (np. klasy, własności), bez odniesienia do konkretnego systemu zarządzania bazą danych. Modelowanie fizyczne polega na implementacji modelu logicznego i optymalizacji bazy danych pod kątem wydajności w konkretnym środowisku implementacyjnym. Flanders, Jannidis, „A Gentle Introduction to Data Modeling", 82–3; Flanders, Jannidis, „Data Modeling", 230.

[21] Øyvind Eide, „Ontologies, Data Modeling, and TEI", *Journal of the Text Encoding Initiative* 8 (2014): 4.

[22] Https://cidoc-crm.org/collaborations.

jest możliwie najgłębsza i najwierniejsza reprezentacja zjawiska, natomiast jego mankamentem jest daleko posunięta specyfika modelu, który – jako mało generyczny – sprawia, że przygotowany zbiór danych staje się dość hermetyczny i trudny do wykorzystania w analizach opartych na „linked data". Tak scharakteryzowane dwa podejścia nie stanowią oczywiście kategorii rozłącznych. Najczęściej w praktyce dochodzi do dostosowywania istniejących modeli danych do potrzeb indywidualnego projektu lub tworzenia własności referujących do ważnych ontologii dziedzinowych w przypadku tworzenia modeli dedykowanych.

Opisane wyżej podejścia dość wyraźnie korespondują z podziałami zaproponowanymi przez Flanders i Jannidis, które można określić ogólnie jako modelowanie w podejściu *top-bottom* i *bottom-top*[23]. We wcześniejszej pracy tych autorów można znaleźć inne określenia odwołujące się nie tyle do kontekstu (sposobu) modelowania, ile do jego celu. Do pierwszej grupy użytkowników tworzących modele zaliczyli oni *curation-driven modelers* (odpowiednik *top-bottom*), do których zaliczyli np. archiwistów, bibliotekarzy czy stowarzyszenia historyczne (np. genealogiczne) – te grupy w gromadzeniu danych odwołują się do modeli najczęściej stosowanych i neutralnych (niekontrowersyjnych). Drugą grupę tworzą *research-driven modelers*, którzy tworzą dedykowane modele dla szczególnych obszarów badawczych, projektów lub wręcz pojedynczych tekstów. Flanders i Jannidis wskazują przy tym, że w praktycznym zastosowaniu zwycięża podejście *bottom-top* (*research-driven modelers*, *user tagging*, *egoisitic*)[24].

Praca nad ontologią dziedzinową 'ontohgis' (ontohgis.pl), która dotyczy historycznej typologii miejscowości oraz jednostek administracji państwowej i świeckiej na ziemiach polskich[25], ukazała skuteczność modelowania w metodyce zwinnej (iteracyjnej), która łączyła podejście *top-bottom* (według niego pracowali głównie inżynierowie ontologii) z podejściem *bottom-top* (które było bliższe ekspertom dziedzinowym). Po ogólnym i wstępnym określeniu specyfikacji wymagań przez historyków i geografów inżynier ontologii przygotował ontologię bazową. Była ona następnie rozwijana i testowana w procedurze iteracyjnej, z udziałem ekspertów dziedzinowych, którzy przedstawiali wyniki badań dotyczących analizowanych pojęć i terminów[26]. W ten sposób konceptualizacja

[23] Flanders, Jannidis, „Data Modeling in a Digital Humanities", 17–8.

[24] Flanders, Jannidis, „Data Modeling", 233–4.

[25] Projekt „Ontologiczne podstawy budowy historycznych systemów informacji geograficznej" (numer projektu 2bH 15 0216 83) pod kierunkiem Bogumiła Szadego został zrealizowany w latach 2016-2019 przez Instytut Historii im. Tadeusza Manteuffla PAN dzięki wsparciu Narodowego Programu Rozwoju Humanistyki.

[26] Agnieszka Ławrynowicz, „Metodyka budowy ontologii 'ontohgis'", w *Metodologia tworzenia czasowo-przestrzennych baz danych dla rozwoju osadnictwa oraz podziałów terytorialnych*, red. Bogumił Szady (Warszawa: Instytut Historii PAN, 2025), 405–7, https://doi.org/10.5281/zenodo.3751265.

i implementacja ontologii docelowej jest wynikiem uzgodnienia między ogólnymi ramami dwóch (standardowych) ontologii dziedzinowych wyższego rzędu (CIDOC CRM, BFO) z wiedzą ekspertów, którzy w swoich badaniach wychodzili od języka źródeł i przekazów historycznych. Praca w metodyce zwinnej sprawiła, że opracowana ontologia dziedzinowa jest zgodna z modelami wyższego rzędu, dzięki czemu dane zgromadzone w strukturach opisanych przez tę ontologię można odnosić do dużych zasobów zewnętrznych opartych na podobnym aparacie pojęciowym. Zastosowane podejście sprawia, że zbudowany model jest wciąż otwarty na uzupełnienia i uściślenia ekspertów specjalizujących się w geograficzno-historycznych badaniach osadniczych oraz administracyjnych. Z jednej strony realizuje więc założenie *model of*, jako reprezentacja czegoś w celach badawczych (poznawczych, zrozumienia, wyjaśnienia), z drugiej zaś realizuje on założenia *model for*, jako struktura umożliwiająca wytworzenie nowej wiedzy na temat przemian osadniczych lub administracyjnych, a ta z kolei pozwala udoskonalić *model of*[27]. Taka „gra modelami", polegająca na manipulowaniu składnikami tych modeli, czyli klasami, własnościami oraz atrybutami, przestawianiu i różnym łączeniu tych elementów, wykazuje wiele analogii z reinterpretacją opisu zjawisk w narracji historycznej w związku z rozwojem stanu wiedzy, paradygmatów i metod badawczych. Nie można jednak dopuścić, żeby to manipulowanie modelami danych humanistycznych zastąpiło istotę badań naukowych[28], których celem wciąż pozostaje poznanie, a przynajmniej dążenie do poznania prawdy przez wyjaśnianie zjawisk utrwalonych w przekazach źródłowych. Cel modelowania danych historycznych nie powinien tkwić w samym modelowaniu, ale w dążeniu do odkrycia prawideł rozwoju procesu dziejowego.

W świetle znaczenia źródeł historycznych w procesie ustalania i wyjaśniania zjawisk z przeszłości pytanie, jak modelować edycje naukowe jako systemy informacji cyfrowej („How to model scholarly editions as digital information systems"[29]), nabiera szczególnego znaczenia i wydaje się wykraczać poza istotę samego edytorstwa. Edycje cyfrowe stają się bazą analiz algorytmów komputerowych i w ten sposób odgrywają zasadniczą rolę w budowaniu wiedzy naukowej bez udziału człowieka. Sposób modelowania danych w edycjach cyfrowych będzie miał wpływ na kształt odpowiedzi udzielanych przez narzędzia AI.

[27] "Modelling of something readily turns into modelling for better or more detailed knowledge of it; similarly, the knowledge gained from realizing a model for something feeds or can feed into an improved version", McCarty, „Humanities Computing", 27.

[28] „Science is no longer the quintessence of knowledge and of what is worth knowing, but a way", Hans-Georg Gadamer, *Reason in the Age of Science*, tłum. Frederick G. Lawrence (Cambridge, MA: MIT Press, 1982), 69–70.

[29] Cyt. za: Flanders, Jannidis, „Data Modeling in a Digital Humanities", 19.

W dotychczasowej refleksji na temat edytorstwa cyfrowego dominowała dyskusja o charakterze filologicznym i tekstologicznym, skupiona na strukturze językowej przekazu, rzadziej zaś na jego wymiarze informacyjnym, który polega na identyfikacji i opisie zjawisk, wydarzeń, osób, miejsc oraz relacji ich łączących. Edytorstwo cyfrowe umożliwia pogodzenie obu tych nurtów, przy czym obecnie należy pogłębić znaczenie warstwy informacyjnej w wydawanych tekstach. Przekreślenie nazwiska podatnika i kwoty zapłaconego podatku w rejestrze z XVI wieku nie może być tylko znakiem graficznym odnotowanym w edycji cyfrowej, ale powinno nieść także walor informacyjny z odniesieniem (*linked data*) do innego rejestru, gdzie ten sam podatnik jest zarejestrowany jako osoba, która już podatek opłaciła, oraz do bazy danych identyfikującej tę osobę, jak również pełnione przez nią funkcje. Pytanie o przedmiot modelowania, o którym będzie mowa w następnej części artykułu, powinno uwzględniać charakter samego źródła (dokumentowe, narracyjne) i umieszczać je w centrum procesu historycznego – jako „agenta", który może występować w dwóch przenikających się rolach. Pierwsza przypomina aktywnego aktora, który współkształtuje, tworzy i zmienia stan rzeczy, druga zaś biernego świadka, który raczej informuje o przebiegu wydarzeń historycznych. W pierwszej roli tekst – jako ściślej powiązany z wydarzeniami – wyjaśnia ich przebieg, w drugiej zaś opisuje zjawiska, podając pewne fakty, miejsca, zdarzenia i osoby, które biorą udział w procesie oraz są mniej lub bardziej „uwikłane" w to źródło. Taka konstrukcja może być zachętą do spojrzenia na źródło, a tym samym na model danych, w kontekście relacji tekstu do wydarzeń, osób i miejsc, zamiast bezpośredniego modelowania relacji między zjawiskami zarejestrowanymi w przekazie źródłowym, np. miejscem a instytucją, instytucją a osobą. Podejście takie, w którym źródło historyczne staje się swoistego rodzaju wydarzeniem w procesie historycznym, może odwoływać się do modelowania w metodyce *event based*, która jest bliższa ontologii CIDOC CRM, a nie *object based*. W ten sposób ważną częścią modelu w edytorstwie cyfrowym byłoby stworzenie serii relacji źródła do osób, miejsc, instytucji, pojęć itp., które w nim występują lub które towarzyszyły jego powstaniu, a nie tylko opisu bytów występujących w tekście.

PRZEDMIOT MODELOWANIA W EDYTORSTWIE CYFROWYM

Jako że edycja cyfrowa nie jest jedynie odwzorowaniem tekstu, ale także stworzoną na jego podstawie bazą danych, to wymaga od edytora zastosowania się do prawideł dotyczących modelowania danych. W tym przypadku chodzi

przede wszystkim o stworzenie zgodnego ze sztuką modelu danych, który pozwoli na przedstawienie jakiegoś określonego obszaru rzeczywistości. Działanie to wymaga namysłu nad kwestią, która w erze przedcyfrowej mogła się wydawać oczywista, a mianowicie nad pytaniem, „co jest przedmiotem modelowania?". Pytanie to nie jest jednak trywialne. W trakcie tworzenia bazy danych na podstawie tekstu edytor może bowiem skupić się na różnych jego aspektach. Może skoncentrować się na zastosowaniu języka i na poziomie ustrukturyzowanych danych zawrzeć informacje o wykorzystanych przypadkach, kolokacjach czy też metaforach. Może też próbować zamodelować, w jaki sposób tekst zmieniał się w czasie – wówczas warstwa informacyjna będzie skoncentrowana przede wszystkim na wskazaniu odmianek zawartych w różnych przekazach. Inną możliwością jest skupienie się na modelowaniu warstwy narracyjnej tekstu lub nawet opisywanej w nim rzeczywistości. Edytor może starać się połączyć różne z wymienionych wyżej podejść, jednak ostatecznie będzie musiał dokonać wyboru. Opisanie na poziomie danych wszystkich aspektów tekstu byłoby bowiem niezmiernie pracochłonne, a przez to przeciwskuteczne. Różne możliwości wymagają refleksji nad tym problemem. Decyzje powinny natomiast być podejmowane przede wszystkim z myślą o odbiorcach dzieła i badaniach, które mogliby oni prowadzić. Najważniejszą cechą edycji jest bowiem jej użyteczność.

Oczywiście problematyka wyboru sposobu, w jaki przetwarzany jest zabytek na potrzeby edycji, nie jest nowa. Jest ona kontynuacją dyskusji o różnicy między metodą filologiczną a historyczną w edytorstwie – dyskusji, która w polskiej literaturze przedmiotu toczy się wyraźnie co najmniej od lat pięćdziesiątych XX wieku[30]. Odróżnienie pracy filologów, którzy starają się przede wszystkim odtworzyć oryginalne brzmienie tekstu i dążą do jego oczyszczenia oraz wskazania jego odmianek, od historyków, którzy traktują źródło jako fakt historyczny i starają się odtworzyć jego „życie" (kontekst powstania, funkcję, sposób wykorzystywania) oraz wyzyskać z niego informacje o rzeczywistości historycznej, było wyraźne już w epoce przedcyfrowej. Zwiększenie znaczenia aparatu krytycznego w edytorstwie cyfrowym[31] wynosi jednak tę kwestię na nowy, wyższy poziom. W praktyce warstwy informacji modelowane przez literaturoznawców, językoznawców czy historyków są na tyle odmienne, że często okazują się nieprzystające do badań prowadzonych w innych dziedzinach.

[30] Brygida Kürbis, „Osiągnięcia i postulaty w zakresie metodyki wydawania źródeł historycznych", *Studia Źródłoznawcze* 1 (1957): 54–7; Gerard Labuda, „Próba nowej systematyki i nowej interpretacji źródeł historycznych", *Studia Źródłoznawcze* 1 (1957): 3–52.

[31] Marina Buzzoni, „A Protocol for Scholarly Digital Editions? The Italian Point of View", w *Digital Scholarly Editing: Theories and Practices*, red. Matthew James Driscoll, Elena Pierazzo (Cambridge: Open Book Publishers, 2016), 64–6.

Mówiąc o modelowaniu danych w edytorstwie, należy zwrócić uwagę, że tekst funkcjonuje przynajmniej w dwóch wymiarach. Pierwszym z nich jest sfera fizyczna, którą można określić również jako sferę „znaku". W tym wymiarze zabytek jest konkretnym przedmiotem (np. książką), składającym się z nośnika (np. papieru) oraz naniesionych na niego znaków (np. liter). Przedmiot ten stanowi materialną manifestację tekstu w jego idealnym rozumieniu, powstałą w określonym miejscu i czasie. W wymiarze idealnym tekst jest komunikatem mającym swego nadawcę, treść oraz zamierzonego odbiorcę. Komunikat ma też jednoznacznie zdefiniowany w umyśle nadawcy cel, np. przekazanie informacji, wywołanie wrażenia, zamanifestowanie treści czy ustanowienie lub zmianę normy prawnej. Edytor (bądź badacz) nie jest uczestnikiem tej komunikacji. Ma do niej dostęp, ale nie bierze w niej udziału. Jest jej zewnętrznym obserwatorem. Wydając tekst, stara się przekazać sferę znaku zabytku, nie jest jednak w stanie odtworzyć jego komunikatu – ten jest udziałem wyłącznie pierwotnych członków komunikacji. Edytor tworzy zatem nowy komunikat, w którym sam jest nadawcą, natomiast użytkownicy edycji stają się jego odbiorcami. Treścią komunikatu jest wzbogacona przez wiedzę pozaźródłową wydawcy interpretacja sfery znaku pierwotnego komunikatu. W edycjach tradycyjnych – ze względu na ograniczenia druku – sfery idei i znaku mocno się przenikały, w edytorstwie cyfrowym zaś powinny one być jak najmocniej odseparowane, tak aby użytkownik edycji zdawał sobie sprawę, jaka część odbieranego komunikatu wynika z fizycznej formy oryginalnego zabytku, a jaka jest udziałem wtórnego nadawcy – edytora. Strukturyzacja informacji ze źródła jest zawsze elementem interpretacji. W edytorstwie cyfrowym nie ma więc miejsca na odziedziczoną po historyzmie manierę, która traktuje badacza oraz źródła jako czyste lustra, obiektywnie odbijające rzeczywistość taką, „jaką była".

Historyków w przekazie źródłowym najbardziej interesuje możliwość pozyskania informacji o przeszłości, przekazywanych zarówno przez jego formę zewnętrzną, jak i treść. Przedmiotem ich dociekań nie jest więc sam tekst, ale przeszłość, do której daje on dostęp. Z tego też względu w edytorstwie historycznym wykształciła się i zyskała dużą popularność idea edycji „regestowej"[32]. W tego typu dziełach „edytorzy" nie wydają tekstu, tylko streszczają pozyskane z niego informacje w formie regestu. Ignorują więc zupełnie sferę znaku, przekazując jedynie komunikat w sferze idei. Rezygnując jednak z przekazania

[32] Przykłady wydawnictw, które porzuciły tradycyjną edycję tekstów na rzecz udostępniania informacji o nich, można by wymieniać dziesiątkami, poczynając od monumentalnych publikacji, takich jak *Regesta Imperii*, *Documenta res Hungaricas tempore regum Andegavensium illustrantia*, *Ut per litteras apostolicas*, czy też polskiego *Bullarium Poloniae*.

sfery znaku, ograniczają możliwość wnioskowania odbiorcy na temat pierwotnej komunikacji, stają się uzurpatorem, dając sobie prawo mówienia w imieniu źródła. Powszechne przyjęcie tej formy „edycji" przez badaczy przeszłości pokazuje najdobitniej, w jak fundamentalny sposób różni się podejście historyków od podejścia filologów. Różnica ta jest bardzo wyraźna także w przypadku pracy ze źródłem w erze cyfrowej. Przejawia się ona tym, że filologów bardziej interesuje opisanie poszczególnych części tekstu, gdy tymczasem historyków – baza danych z niego wynikająca. Z tego też względu badacze przeszłości więcej wysiłku poświęcają modelowaniu danych czy też tworzeniu edycji asertywnych[33] niż anotacji tekstu. Inaczej niż edycje drukowane, technologie cyfrowe pozwalają na różnorodne prezentowanie opracowywanego tekstu, umożliwiają więc do pewnego stopnia pogodzenie obydwóch tych podejść.

Kwestią zasadniczą w modelowaniu danych w edytorstwie cyfrowym jest wyraźne oddzielenie prezentacji sfery „znaku" od sfery „idei". Odpowiadają temu dwie warstwy modelowania: warstwa źródłowa (tekstowa), która oddaje elementy, własności i atrybuty zawarte w źródle, oraz warstwa krytyczna (informacyjna), zawierająca dodatkowe informacje o rzeczonych elementach, własnościach i atrybutach. Celem tego pierwszego jest więc jak najdokładniejsze odwzorowanie i opisanie fizycznej formy zabytku. W tej części powinien się więc znaleźć opis wymiarów, opraw, mocowań itp., a także jak najdokładniejsze skopiowanie znaków odpowiadających za przekazanie tekstu wraz z ich układem. W edytorstwie tradycyjnym ze względów użytkowych przyjęło się publikowanie transkrypcji, a zatem tekstu w zmodernizowanej formie. Ponieważ modernizacja pisowni jest jednak elementem interpretacji edytora, nie powinna się ona znajdować na poziomie modelu źródłowego. Znacznie lepszym rozwiązaniem w tym wypadku jest zastosowanie transliteracji. Miejsce na modernizację czy objaśnienie znaczenia poszczególnych znaków i słów jest w aparacie krytycznym, który w edycjach cyfrowych daje znacznie większe możliwości niż w przypadku druku. W wielu edycjach cyfrowych dostęp do poziomu znaku źródła zapewniany jest dzięki dołączeniu skanów. Pomagają one użytkownikowi lepiej zapoznać się z oryginałem, nie mogą jednak zastąpić ustrukturyzowanych danych na tym poziomie.

Jak ważne jest rozgraniczenie modelu źródłowego i analitycznego, wskazuje doświadczenie pracy z materiałami kartograficznymi, gdzie sfery „znaku"

[33] Koncept ten przedstawił: Georg Vogeler, „The 'Assertive Edition'. On the Consequences of Digital Methods in Scholarly Editing for Historians", *International Journal of Digital Humanities* 1, nr 2 (2019): 309–22, https://doi.org/10.1007/s42803-019-00025-5. W Polsce ideę tę propagowali Marek Słoń i Michał Słomski, zob. „Edycje cyfrowe źródeł historycznych", w *Jak wydawać teksty dawne*, red. Karolina Borowiec, Dorota Masłej, Tomasz Mika, Dorota Rojszczak-Robińska (Poznań: Wydawnictwo Rys, 2017), 65–84.

i „znaczenia" są szczególnie widoczne, ponieważ mapa operuje powiązanymi ze sobą językami tekstu (np. nazwy) i symbolu (np. sygnatury). Między innymi z tego powodu w procedurze digitalizacji map województw Karola Perthéesa – o czym będzie jeszcze mowa – zdecydowano się oddzielić prostą rejestrację sygnatur i nazw od etapu łączenia ich w obiekty geograficzne. W wielu przypadkach relacja między sygnaturami a nazwami była bowiem niejednoznaczna, a zatem utworzenie obiektu, np. miejscowości, zawierało element interpretacji historyka-geografa i odzwierciedlało jego subiektywną opinię[34].

Figure 3. Fragment of the 'special map' of Kalisz palatinate depicting the complexity of the drawing. Going from the left, there is "Turska m." which is a parochial village (proper name and sign are overlapped by the red polygon), then there is "Boguszow" village with an inn (proper name and two signs are overlapped), a separate mill and two other villages. The polygons are features is the database as allow for flexible linking between sometimes ambiguously represented annotations and signs.

Ryc. 2. Odrębne oznaczenie nazw miejscowości i sygnatur obiektów topograficznych w procedurze rejestrowania danych z mapy województwa kaliskiego Karola Perthéesa z II połowy XVIII wieku. Źródło: Panecki, „Mapping Imprecision", 9.

W modelu analitycznym powinna znaleźć się interpretacja zabytku. Tworząc model analityczny, edytor nakłada na odwzorowane „znaki" swoją wizję przeszłości wynikającą z wiedzy pozaźródłowej. Modeluje więc przeszłość, jak rozumiał to Topolski[35]. Definiując klasy obiektów, ich możliwe atrybuty oraz relacje między nimi, wychodzi poza odwzorowanie czy też komentowanie tekstu, jak ma to miejsce w przypadku edycji drukowanej, i tworzy *de facto* konkretny model rzeczywistości. Na model ten wpływa zarówno treść przekazu źródłowego, jak i wiedza pozaźródłowa edytora. Edytor cyfrowy na poziomie analitycznym

[34] Tomasz Panecki, „Mapping Imprecision: How to Geocode Data from Inaccurate Historic Maps", *ISPRS International Journal of Geo-Information* 12, nr 4 (2023): 149.

[35] Topolski, *Teoria wiedzy historycznej*.

interpretuje treść źródła w znacznie większym stopniu niż robił to edytor tradycyjny np. w przypisach rzeczowych. Jest mu zdecydowanie bliżej do badacza sporządzającego regesty dokumentów, jednak możliwość jednoczesnego udostępnienia zarówno przeanalizowanych przez niego danych, jak i odwzorowania sfery znaku pozwala użytkownikowi edycji zapoznać się nie tylko z informacjami ze źródła, lecz także z samym źródłem. Aby jednak uniknąć „mówienia w imieniu źródła", konieczne jest odpowiednie zdefiniowanie relacji między tekstem przekazu a danymi oraz świadome wyznaczenie granic między wiedzą źródłową i pozaźródłową.

W tym pierwszym przypadku zasadniczym pytaniem jest, czy bardziej zależy nam na opisaniu znaczenia poszczególnych części tekstu, czy też raczej na zamodelowaniu jego treści. Uchodzący od ponad dekady za standard do tworzenia edycji cyfrowych model danych, wypracowany w ramach inicjatywy Text Encoding Initiative, wywodzi się z filologicznego sposobu myślenia. Nadaje się on bardzo dobrze do definiowania znaczenia poszczególnych słów, a także modelowania odmianek tekstu ze wskazaniem kontaminacji i najlepszej jego wersji. Na poziomie metadanych pozwala przede wszystkim zdefiniować różnych świadków tekstu, a także opisać ich fizyczną formę. Sprawdza się więc w oddawaniu sfery „znaku" tekstu, jest natomiast zdecydowanie mniej adekwatny do oddawania jego znaczenia. W standardzie tym jest możliwe tylko definiowanie hierarchicznej relacji między słowami, w której struktura mniej ogólna musi się w całości zawierać w strukturze bardziej ogólnej. Nie jest natomiast możliwe definiowanie niehierarchicznych relacji pomiędzy słowami czy zjawiskami. Ograniczenie to uniemożliwia anotowanie semantyki tekstu, a w konsekwencji – treści komunikatu. Możemy więc oznaczyć informację, że dane słowo jest imieniem osoby, a inne nazwą organizacji, jednak nie dostajemy satysfakcjonujących narzędzi, które pozwalałyby na zdefiniowanie związku między nimi. Podobnie nie mamy możliwości anotacji czynności i wydarzeń (te można definiować tylko na poziomie metadanych). Wynika stąd, że za pomocą TEI możemy stworzyć cyfrowy indeks umożliwiający odnajdywanie bytów poszczególnych kategorii, nie możemy jednak zbudować bazy danych pozwalającej zrozumieć przekaz tekstu. Podstawowe zastosowanie TEI służy więc przede wszystkim modelowaniu tekstu jako takiego, nie jest natomiast odpowiednie do modelowania tekstu jako źródła historycznego. Nawet dodanie możliwości referowania do zasobów zewnętrznych (atrybut @ref) czy wprowadzenie modelowania wydarzeń w metadanych (tag <event>) nie zmienia zasadniczo tego stanu rzeczy. Być może właśnie z tych ograniczeń modelu danych TEI wynika fakt, że refleksja historyczna nad modelowaniem danych zmierza raczej w kierunku baz relacyjnych lub grafowych.

Możliwość modelowania relacji daje przede wszystkim baza grafowa. Kluczowe w tym aspekcie jest definiowanie relacji między dwoma bytami. W przypadku baz grafowych (np. wykorzystujących standard RDF) relacja definiowana jest przez orzeczenie w układzie podmiot–orzeczenie–dopełnienie. Możliwość łączenia dwóch bytów za pomocą relacji oznaczającej czynność pozwala w sposób ustrukturyzowany odwzorować informację zawartą w komunikacji źródła. Najbardziej rozbudowany przykład zastosowania tego podejścia można zaobserwować w systemie CASTEMO zastosowanym w projekcie DISSINET[36]. W jego ramach zdania dzielone są na poszczególne wyrazy, które są ze sobą wzajemnie łączone relacjami. Każdy wyraz ma więc jednoznacznie zdefiniowaną własną funkcję w komunikacie. Takie modelowanie informacji ze źródła pozwala w sposób pełny odwzorować narrację w formie ustrukturyzowanej. Z tego też względu daje ono badaczowi możliwość zarówno odnajdywania poszczególnych faktów historycznych, jak i umieszczenia ich w strukturze komunikatu (możemy znaleźć nie tylko informację, że dana osoba była heretykiem, lecz także sprawdzić, jak na ten temat wypowiadali się poszczególni twórcy komunikatów). Takie przygotowanie tekstu wydaje się satysfakcjonujące dla każdych badań, jest ono jednak niezmiernie czasochłonne. Kluczowa dla wypracowania standardu modelowania danych w historycznym edytorstwie cyfrowym wydaje się możliwość uzyskania najważniejszych funkcjonalności powyżej opisanego podejścia, przy jednoczesnym ograniczeniu czasochłonności analizy tekstu.

Fundamentalną kwestią wydaje się oddanie relacji między poszczególnymi wyrazami oraz opisanie ich znaczenia w kontekście komunikatu. W tym przypadku edytor ma dwie możliwości. Może starać się odtworzyć relację wewnątrz tekstu, tzn. zamodelować informację opisaną w tekście, lub też wyjść poza tekst i starać się zamodelować fakty historyczne, które z niego wynikają. Dla historyka, którego interesują przede wszystkim fakty, bardziej naturalne wydaje się to drugie podejście. Baza danych uwzględniająca fakty historyczne będzie dużo bardziej czytelna, jednak może być zwodnicza. Bez uwzględnienia niepewności i względności oraz kontekstu wypowiedzi łatwo utracić rzeczywiste znaczenie komunikatu. Najlepiej kwestię tę pokazać na przykładzie przekazu kronikarskiego. Ze zdania: „Iohannes eciam interim Quinqueecclesiensis episcopus, dolore tribulacionum pressus, diem obiit vel, ut quibusdam placuit, venano intoxicatus”[37], na poziomie danych, nie możemy powiedzieć, że biskup Jan został otruty.

[36] David Zbíral, Robert L. J. Shaw, Tomáš Hampejs, Adam Mertel, *Model the Source First! Towards Source Modelling and Source Criticism 2.0* (Zenodo, 2021), https://doi.org/10.5281/zenodo.5218926.

[37] *Joannis Dlugossii Annales seu Cronicae Incliti Regni Poloniae: Liber Duodecimus 1462–1480* (Cracoviae: Polska Akademia Umiejętności, 2005), 286 (1472 r.).

Musimy dać możliwość zamodelowania względności wynikającej z wtrącenia „vel, ut quibusdam placuit". Co więcej, musimy umożliwić czytelnikowi zrozumienie całego kontekstu wypowiedzi, a dotyczy ona rzekomej odpowiedzi Macieja Korwina na spiskowanie jego poddanych z Jagiellonami. Długosz w tym zdaniu w pośredni sposób oskarżył Macieja Korwina o morderstwo polityczne. Jakiekolwiek zamodelowanie tej informacji bez uwzględnienia niuansów tej wypowiedzi będzie więc zwodnicze.

Problem kontekstu i „zwodniczości" komunikatu źródła, tak jasno widoczny w przypadku kronik, dotyczy jednak każdego tekstu źródłowego. Pierwszym problemem jest język komunikatu, który nigdy nie jest obiektywny, zawsze niesie ze sobą ukryte znaczenie. Dobrym tego typu przykładem jest używane w papieskiej dyplomatyce sformułowanie „carissimus in Christo filius noster" jako tytułu władców w papieskich dokumentach. Bez znajomości papieskiej dyplomatyki moglibyśmy zrozumieć to zdanie jako świadczące o szczególnie bliskiej relacji, tymczasem było ono używane niezależnie od relacji między władcą a papieżem. Tytułowano w ten sposób nawet władców, którzy działali wbrew Stolicy Apostolskiej, jak np. Kazimierza Jagiellończyka w trakcie wojny trzynastoletniej[38]. Niezrozumienie tej konwencji i oddanie jej na poziomie danych w sposób dosłowny wprowadziłoby więc w błąd użytkowników edycji.

Drugim problemem jest brak obiektywizmu samego nadawcy. W przypadku wszystkich komunikatów zapisanych w celu przekazania jakiejś informacji innym osobom nadawca zawsze opisuje swój punkt widzenia. Świadomie lub nieświadomie manipuluje on zatem odbiorcą informacji przez dobór przekazanych informacji oraz sposób ich przedstawienia. Nawet źródła, które za cel miały wierne opisanie rzeczywistości (jak źródła prawa czy rejestry instytucji państwowych), nie są w pełni neutralne. W ich przypadku należy bowiem zawsze brać pod uwagę kontekst ich wytworzenia, a także niewiedzę i ograniczenia ich twórców. Dobrym przykładem w tym względzie są staropolskie rejestry poborowe. Nie mogą one zostać bezkrytycznie przyjęte jako źródło opisujące gospodarkę państwa. Kolektor lub pisarz odnotowywali bowiem tylko to, co uważali za słuszne. Z tego też względu sposób opisywania podatku, choć w teorii powinien być ujednolicony i opisywać te same kategorie, różni się między poszczególnymi ziemiami, nawet dla tego samego poboru. Aby odpowiednio zinterpretować źródło, badacz musi sięgnąć więc zarówno po uniwersały ustalające kategorie dóbr podlegających

[38] *Vetera Monumenta Poloniae et Lithuanie Gentiumque Finitimarum Historiam Illustrantia*, red. Augustin Theiner, t. 2, nr 159 (Romae: Typis Vaticanis, 1861), 116.

opłatom, jak i praktykę danego poborcy[39]. Z tego względu celem modelu danych edycji powinno być zarówno przekazanie komunikatu źródła, jak i informacji o rzeczywistości przezeń opisanej.

Edytor musi więc umiejętnie wyczuć i zamodelować granicę między opisem i komentowaniem źródła a rzeczywistością przezeń opisaną. W erze druku postulowano przygotowanie, poza wstępem do edycji, także artykułu źródłoznawczego[40], w którym edytor miałby możliwość przedstawienia własnej analizy opracowywanego tekstu[41]. W przypadku edycji cyfrowej jest na to miejsce na poziomie modelu analitycznego. Edytor może, a nawet powinien w tym modelu poddać tekst analizie i wzbogacić o swoją wiedzę pozaźródłową. Może to zrobić na dwa sposoby: po pierwsze – poprzez opisywanie bytów za pomocą klas i relacji z innymi bytami, po drugie – poprzez referowanie do zewnętrznych baz wiedzy (baz referencyjnych). Nie da się tego osiągnąć wyłącznie przy użyciu standardu TEI; konieczne jest połączenie go z innym środowiskiem bazodanowym. To z kolei rodzi pytanie: „jak i gdzie modelować?".

MODEL I DANE W EDYCJI CYFROWEJ – RAZEM CZY OSOBNO

Uwagi dotyczące metodyki modelowania danych w edycji cyfrowej sformułowane w tej części mają wymiar teoretyczny, odwołujący się do literatury przedmiotu, oraz praktyczny, który odnosi się do prac prowadzonych w Instytucie Historii Polskiej Akademii Nauk oraz w Katolickim Uniwersytecie Lubelskim Jana Pawła II. W ramach Zakładu Atlasu Historycznego IH PAN, w którym funkcjonuje Pracownia Historii Cyfrowej, prowadzony jest projekt „Kartografia w służbie reform państwa epoki stanisławowskiej – krytyczne opracowanie «Geograficzno-statystycznego opisania parafiów Królestwa Polskiego» oraz map województw koronnych Karola Perthéesa" (NPRH, nr 11H 18 0122 87,

[39] Tomasz Związek, „Mechanizmy poboru podatków nadzwyczajnych na początku XVI wieku na przykładzie kaliskich rejestrów poborowych", *Kwartalnik Historii Kultury Materialnej* 70, nr 1 (2022): 13–29.

[40] Brygida Kürbis, Gerard Labuda, „Nowa seria Monumenta Poloniae Historica", *Studia Źródłoznawcze* 1 (1957): 216.

[41] Przykładami edycji opatrzonych obszernym wstępem źródłoznawczym i będących *de facto* monografiami edytowanego zabytku są np.: *Privilegia Casimiriana. Przywileje króla Kazimierza IV Jagiellończyka dla Gdańska z okresu wojny trzynastoletniej*, wyd. Marcin Grulkowski, t. 1 (Gdańsk: Muzeum Gdańska, 2023), czy też Adam Kozak, *Księga sądowa gnieźnieńskich wikariuszy generalnych Sędka z Czechla i Jana z Brzóstkowa (1449–1453, 1455). Studium źródłoznawcze i edycja krytyczna* (Poznań–Warszawa: Poznańskie Towarzystwo Przyjaciół Nauk–Polskie Towarzystwo Historyczne, 2023).

https://perthees.ihpan.edu.pl/), którego integralną częścią jest krytyczne opracowanie rękopiśmiennej kartoteki parafii oraz map województw koronnych z II połowy XVIII wieku. Z kolei Laboratorium Humanistycznych Danych Badawczych KUL wspiera projekt edycji dokumentów watykańskich realizowany przez zespół Marka Daniela Kowalskiego w Centrum Studiów Mediewistycznych KUL (https://csm.kul.pl/vaticana/). Zróżnicowany charakter materiału źródłowego oraz różne cele stawiane przed zespołami projektowymi umożliwiają pewną generalizację kwestii dotyczących tego, jak modelować edycje naukowe jako cyfrowe systemy informacyjne[42]. Wykorzystane zostało także doświadczenie metodyczne edycji hybrydowej dwóch kościelnych rejestrów podatkowych z 1561 r. (http://geo-ecclesiae.kul.pl/regestra-1561).

Rozwój metodyczny naukowej edycji cyfrowej dość wyraźnie poszerza jej zakres, wychodząc poza modelowanie samego tekstu źródła. Przez wiele lat, od 1987 r., gdy narodził się standard TEI, głównym zadaniem wydawcy było opisanie tekstu źródła odpowiednimi znacznikami w notacji SGML/XML. Każda edycja cyfrowa, podobnie jak „analogowa", stanowiła odrębną jednostkę, która mogła być łączona z innymi edycjami przede wszystkim na płaszczyźnie wspólnego standardu znaczników TEI. Jego zaletą jest duży stopień generalizacji oraz prostota, co ułatwia integrację np. danych o osobach i miejscach, przy czym integracja ta do 2007 r. była możliwa głównie na poziomie rodzajów informacji ustalonych przez standard TEI, przykładowo przez porównywanie wartości tekstowych oznaczonych jako <persName>, <surname>, <forename>, <roleName>, <placeName>, <settlement> czy <country>. Istotny przełom przyniósł rok 2007 i opublikowanie standardu TEI w wersji 5, która umożliwia łączenie znaczników TEI oraz wyrazów przez nie oznaczonych z zewnętrznymi zbiorami referencyjnymi – zarówno na poziomie elementów modelu danych, jak i samych danych[43]. Postawiło to przed wydawcami źródeł historycznych problem uczestnictwa w tworzeniu lub modyfikowaniu taksonomii i ontologii dziedzinowych, a także zbiorów referencyjnych. Edycje cyfrowe przestały być zamkniętymi – chociaż połączonymi standardem – kolekcjami i stały się pełnowartościowymi elementami „linked data", a zarazem integralną częścią baz wiedzy w naukach humanistycznych. Wprowadzenie połączeń do zbiorów zewnętrznych umożliwia zapis części elementów edycji cyfrowej i jej aparatu krytycznego poza plikami XML, a także otwiera problem modelowania danych w bardziej rozbudowanych strukturach baz grafowych oraz relacyjnych. Stąd postawione w tytule

[42] Flanders, Jannidis, „Data Modeling in a Digital Humanities", 19.

[43] Christian Wittern, Arianna Ciula, Conal Tuohy, „The Making of TEI P5", *Literary and Linguistic Computing* 24, nr 3 (2009): 281–96.

podrozdziału pytanie o relację edycji cyfrowej i danych badawczych oraz miejsce danych w edycji cyfrowej – razem czy osobno? Dodatkowe komplikacje, ale zarazem szanse, stwarza możliwość uwzględnienia w edycji także łączenia obrazu źródła z danymi przez anotowanie (znacznikowanie) nie tylko tekstu, ale też fragmentów skanów, i udostępnianie tych połączeń w Internecie.

Wymienione wyżej przykłady edycji reprezentują odmienne architektury – zarówno pod względem metodyki modelowania danych, zastosowanych rozwiązań, jak i konstrukcji samych edycji. W dalszej części tekstu – dla ułatwienia odbioru i skrócenia opisu – edycja dokumentów watykańskich jest określana jako „Vaticana", edycja kartoteki parafii i map Karola Perthéesa jako „Perthées", a edycja hybrydowa spisów podatkowych z 1561 r. – jako „Regestra 1561".

Tab. 1. Architektura prezentowanych edycji cyfrowych

	Model danych	Środowisko danych	Środowisko edycji
Perthées	relacyjny	PostgreSQL	Geoserver-INDXR
Vaticana	grafowy	Wikibase	XML TEI Publisher
Regestra 1561	grafowy relacyjny	PostgreSQL Wikibase Nodegoat	XML TEI Publisher Geoserver-INDXR

Źródło: Oprac. własne

W przypadku opracowania materiałów Perthéesa przyjęto architekturę wielomodelową, a wszystkie dane są przechowywane w relacyjnej przestrzennej bazie danych PostgreSQL/PostGIS, która jest połączona ze skanami źródeł przez system anotacji wykorzystujący układ warstw i standardy GIS. Kartograficzny charakter źródeł, który sprawia, że dużą rolę odgrywa schemat treści, a elementy graficzne (sygnatury, symbole) przeplatają się z tekstowymi (podpisy, etykiety, opisy tekstowe), uniemożliwił praktyczne zastosowanie przy opracowaniu tego materiału standardu XML TEI, który jest dedykowany głównie źródłom tekstowym, a grafika stanowi jedynie element towarzyszący. Biorąc pod uwagę cel opracowania, jakim było przygotowanie mapy osadnictwa i podziałów administracyjnych, oraz objętość materiału (ponad 2 tys. kart parafii i 12 arkuszy map w skali 1:225 000), podjęto decyzję o przygotowaniu modelu edycji obejmującego tylko część treści zamieszczonej w tzw. kartotece parafii. Wykorzystanie standardów przestrzennych GIS Web Map Service (WMS) oraz Web Feature Service (WFS) opracowanych przez Open Geospatial Consortium (OGC)

umożliwiło ustrukturyzowaną anotację treści kart parafii oraz map, przy której wykorzystano metodykę digitalizacji treści map dawnych[44]. Przy podejściu wielomodelowym struktura danych jest mocno zdeterminowana przez etap pracy badawczej. W ten sposób model danych służący bezpośredniemu gromadzeniu informacji ze źródeł w procesie anotowania fragmentów kart parafii i map może być określony jako *source driven* lub *bottom-top*. Modele pośrednie, odrębne dla kart parafii i map, które mogą być określone jako krytyczne lub analityczne, uwzględniają zarówno strukturę źródeł, jak i potrzeby badawcze związane z ustaleniem sieci osadniczej i jednostek podziału administracyjnego w II połowie XVIII wieku. Na ostatnim poziomie znajduje się model prezentacyjny określony potrzebami wizualizacji kartograficznej oraz aplikacji internetowej. Jest on powiązany z modelami analitycznymi dla kart parafii i map, gdyż jest zasilany danymi w sposób automatyczny na podstawie algorytmów porównujących wcześniej opracowane dane[45]. Aplikacja internetowa ukazująca opracowane materiały źródłowe, która ukaże się w 2026 r., będzie zawierała dane pochodzące ze wszystkich trzech poziomów pracy nad źródłami z odniesieniem do ich interaktywnych faksymil. Zbliżona metodyka była wykorzystana przy opracowaniu *Słownika geograficznego Królestwa Polskiego i innych krajów słowiańskich* oraz map Wojskowego Instytutu Geograficznego w skali 1:100 000 (https://atlas1918.ihpan.edu.pl). Na skutek trudności interpretacyjnych przekazu kartograficznego model danych w projekcie Perthéesa głębiej oddzielił warstwę graficzną (symbole, etykiety) od warstwy obiektowej (*features*), która zawiera już element subiektywnej interpretacji geografa-historyka[46].

Jak wspomniano wcześniej, rozwinięcie standardu TEI, tak aby umożliwić referencję modeli i danych edycji ze zbiorami zewnętrznymi, zmienia charakter samych edycji. Z tej możliwości w pełni korzysta zespół Vaticana, który podzielił architekturę swojej edycji cyfrowej na dwie części: tekstową (XML TEI) oraz bazodanową (RDF)[47]. Problem relacji między modelem danych „zaszytym" w głównym pliku lub plikach edycji w formacie XML oraz modelami zewnętrznymi szeroko omówił Eide.

---

[44] Tomasz Panecki, „Cyfrowe edycje map dawnych: perspektywy i ograniczenia na przykładzie mapy Gaula/Raczyńskiego (1807–1812)", *Studia Źródłoznawcze. Commentationes* 58 (2020): 199–202.

[45] Bogumił Szady, Tomasz Panecki, „Source-Driven Data Model for Geohistorical Records' Editing: A Case Study of the Works of Karol Perthées", *Miscellanea Geographica. Regional Studies on Development* 26, nr 1 (2022): 52–62.

[46] Zob. s. 23; Panecki, „Mapping Imprecision", 8–9.

[47] Więcej szczegółów na temat metodyki pracy oraz modelu danych edycji źródeł watykańskich odnajdzie czytelnik w odrębnym tekście tego tomu: „Między tekstem a danymi. Metodologia opracowywania cyfrowej edycji na przykładzie *Monumenta vaticana res gestas polonicas illustrantia*", 69–92.

Tab. 2. Rodzaje relacji między TEI i zewnętrznymi ontologiami

	Links from TEI	Links to TEI
TEI in body (method A)	Elements in the body of the TEI document point to an external ontology.	Links from external ontology to elements in the body of the TEI documents.
Non-TEI in header (method B)	Elements in the body of the TEI document point to elements in another namespace in the header. Links from the header elements point to an external ontology.	Links from external ontology point to elements in another namespace in the header of the TEI document.
TEI in header (method C)	Elements in the body of the TEI document point to TEI elements in the header. Links from the header elements point to an external ontology.	Links from ontology point to TEI elements in the header of the TEI document.

Źródło: Eide, „Ontologies, Data Modeling, and TEI", 12.

Metodyka pracy oraz architektura edycji źródeł watykańskich wskazują, że zastosowanie znalazło tutaj połączenie podejścia A i B, w którym poszczególne elementy w tekście edycji powiązano z zewnętrznym zbiorem danych oraz zewnętrzną ontologią. Nie jest to jednak klasycznie rozumiana ontologia dziedzinowa opisana w języku OWL, ponieważ w tym przypadku została ona umieszczona w postaci modelu danych w środowisku Wikibase razem z danymi (https://va.wiki.kul.pl). Odniesienia dotyczą więc nie tylko modelu danych (klasy, własności, atrybuty), lecz także elementów, takich jak konkretne osoby, miejsca, instytucje czy dokumenty. W ten sposób, przykładowo, w nagłówku edycji dokumentu jego metadane opisano w postaci zestawu odniesień (np. <catRef scheme="formularz" target="va-Q354"/>) do elementów opisanych w bazie va.wiki.kul.pl (https://va.wiki.kul.pl/wiki/Item:Q354). Edycja źródeł watykańskich składa się więc z dwóch części: anotowanego w standardzie TEI pliku XML i towarzyszącej mu bazy danych, która zastąpiła opisy wszystkich elementów edycji w nagłówku pliku XML oraz otworzyła możliwość modelowania, wzbogacania, a także wykorzystywania zgromadzonych danych poza tekstem źródła.

Ważną kwestią w tym świetle jest określenie relacji między modelowaniem w pliku XML oraz modelowaniem tej części danych, która została umieszczona w odrębnym środowisku, w bazie danych. Analiza anotacji w pliku XML oraz elementów i własności w bazie danych wskazuje, że baza danych, mimo iż zawiera kluczowe informacje dla odczytania i interpretacji edycji w pliku XML, ma wobec niej charakter wtórny. Główna różnica wobec stosowanych do tej pory rozwiązań polega na tym, że ma charakter dziedzinowy, a model danych został ściśle związany z celem, jakim jest edycja źródeł watykańskich. Jej model tym samym jest uwarunkowany raczej źródłowo niż tematycznie (*source driven*). W samej edycji odnośniki do bazy danych dotyczą głównie obiektów (osób, instytucji) zapisanych w schemacie: prefix (va-) oraz identyfikator elementu w bazie Wikibase (QXXX). O ile referencje dla obiektów (osób, miejsc czy instytucji) stworzone przez atrybut @ref=va-QXXX trwale wiążą edycję zapisaną w XML z obiektami w bazie danych, o tyle atrybuty (własności, role) oznaczone jako @ana=’tekst’, np. <rs ref="va-Q249" ana="wystawca"> stanowią tylko źródło informacji dla przygotowania odpowiednich elementów w bazie danych (‘wystawca’ jako ‘rola w dokumencie’ jest opisany w bazie danych z identyfikatorem https://va.wiki.kul.pl/wiki/Item:Q536). Powstaje tutaj istotne pytanie, jak głęboko powinna postępować integracja edycji zapisanej w formacie XML z bazą danych RDF i czy np. atrybut @ana=’xxx’ nie powinien mieć także charakteru odniesienia do bazy danych (w tym przypadku do właściwości P).

Rozwój „linked data” oraz różnego typu ontologii dziedzinowych, taksonomii i referencyjnych baz danych oraz narzędzi je wykorzystujących wskazuje, że będzie rosła skala obiektywizacji elementów edycji przez tworzenie referencji do już istniejących zasobów danych. Tym bardziej że z poziomu bazy danych, np. przez podanie źródła informacji, można wskazać ich źródło w konkretnej edycji, a nawet w konkretnym dokumencie lub miejscu w dokumencie. W ten sposób w bazie danych, która staje się cyfrowym aparatem naukowym edycji, można oznakować źródło informacji i odróżnić te wiadomości, które pochodzą z tekstu źródła, od wiedzy pozaźródłowej.

Wskazane wyżej podejścia do modelowania danych w edytorstwie zostały eksperymentalnie zestawione w przygotowanej edycji hybrydowej rejestru podatku zebranego od duchowieństwa diecezji poznańskiej w 1561 r., który zachował się w dwóch rękopisach (https://geo-ecclesiae.kul.pl/regestra-1561). Wydawca przygotował w tym przypadku artykuł naukowy, zawierający wstęp źródłoznawczy, któremu towarzyszy strona internetowa oraz dostępne on-line:

– tradycyjna edycja źródłowa w formacie PDF;

– aplikacja internetowa INDXR, która zawiera skany rękopisów oraz dane na temat osób, miejsc i instytucji wymienionych w obu rejestrach;

– cyfrowa edycja pełnotekstowa (XML) w aplikacji TEI-Publisher;

– aplikacja internetowa w środowisku *Nodegoat*, która wizualizuje dane badawcze zgromadzone w trakcie opracowania obu rękopisów;

– repozytorium danych w postaci aplikacji Wikibase, która umożliwa dostęp do danych za pomocą SPARQL-endpointa;

– pliki z danymi do pobrania wraz z opisem modelu danych, umieszczone w repozytorium Zenodo.

Miejsce danych badawczych w każdej z prezentowanych form edycji jest inne, i to właśnie ta relacja je różnicuje. W przypadku edycji tradycyjnej funkcję danych badawczych spełnia aparat naukowy, czyli przypisy oraz indeksy geograficzny i osobowy. Są one umieszczone w edycji, jednak łatwo je wyodrębnić – przypisy znajdują się na dole każdej strony, natomiast indeksy zostały umieszczone na końcu, po tekście źródła. W przypadku aplikacji przygotowanej za pomocą aplikacji INXDR[48], która umożliwia anotowanie treści ze skanów, a następnie ich prezentację razem z danymi, cała treść znajduje się w relacyjnej bazie danych, która zawiera zarówno informacje ze źródła (model źródłowy), jak również komentarz krytyczny i dane uzupełniające (model krytyczny). W opinii niektórych specjalistów, ze względu na brak oddania tekstu źródła w pełnym brzmieniu, taka forma nie spełnia kryteriów edycji naukowej. W przypadku edycji cyfrowej pełnotekstowej dane badawcze znajdują się w dwóch połączonych ze sobą miejscach: w pliku XML oraz w Wikibase. Jak wynika z powyższego, miejsce i rola danych badawczych względem treści źródła w edycji tradycyjnej oraz edycji cyfrowej są bardzo podobne. Podstawową różnicę wprowadzają w tym przypadku dwa elementy: 1) struktura i forma danych badawczych w edycji analogowej mają formę narracyjną, wyłącznie informacyjną, której nie można poddać procedurom analitycznym jakościowym i ilościowym, co jest możliwe w przypadku danych badawczych w edycji cyfrowej; 2) możliwość uzupełniania i rozwijania danych badawczych w przypadku edycji cyfrowej pełnotekstowej także po jej opublikowaniu (udostępnieniu) – zastosowanie środowiska Wikibase czyni z edycji cyfrowej pełnotekstowej edycję otwartą z możliwością rejestracji zmian oraz autorstwa zmian wprowadzanych w aparacie naukowym, co jest niemożliwe w przypadku edycji tradycyjnej.

Jedną z ważnych różnic w metodyce opracowania danych badawczych dla edycji rejestrów podatkowych z 1561 r. w porównaniu z edycją dokumentów

[48] Https://ihpan.edu.pl/dariah-lab/indxr/ [dostęp: 2.06.2025].

watykańskich jest sposób przygotowania bazy danych Wikibase. W przypadku Regestra 1561 baza danych Wikibase, mimo że ma inny model danych, została zasilona danymi przez eksport z wcześniej przygotowanej relacyjnej bazy danych PostgreSQL/PostGIS. W przypadku źródeł watykańskich nie jest opracowywana relacyjna baza danych i dane do towarzyszącej tej edycji bazy Wikibase są wprowadzane w dwóch fazach: w pierwszej fazie ręcznie – poprzez stworzenie odpowiednich elementów (np. dokumentów, osób, miejsc, instytucji), a w drugiej fazie – poprzez skrypt w języku Python, który wydobywa dane z edycji XML i tworzy odpowiednie twierdzenia (*statements*) w bazie Wikibase.

*

Określenie miejsca modelowania danych we współczesnym edytorstwie cyfrowym, rozumianym jako proces badawczy, wymagało odniesienia do podstaw teorii modelowania z odwołaniem do konkretnych problemów ilustrowanych przykładami użycia. Konieczna była pogłębiona refleksja o przedmiocie i metodyce modelowania, przy czym zagadnienia teoretyczne ograniczono do zakresu koniecznego dla wyjaśnienia lub uzasadnienia wysuwanych propozycji. Pominięto pogłębione, często filozoficzne i kognitywistyczne, rozważania dotyczące teorii danych i informacji, źródeł historycznych i wiedzy pozaźródłowej. W zakresie ogólnej teorii reprezentacji wiedzy historycznej oraz modelowania danych w humanistyce cyfrowej odsyłamy do obszernej literatury przedmiotu[49].

Modelowanie danych w edytorstwie cyfrowym w prostej linii nawiązuje do elementów edytorstwa tradycyjnego, a więc pełnotekstowych edycji źródeł opatrzonych aparatem naukowym, przypisami rzeczowymi i tekstowymi oraz indeksami. Wszystkie te elementy występują również w edytorstwie cyfrowym – przypisy tekstowe reprezentowane są przez elementy modelu dotyczące struktury dokumentu oraz jego grafii i formy, a przypisy rzeczowe są oddawane przez znaczniki (tagi) lub rekordy w bazie danych zawierających informacje o osobach, miejscach czy wydarzeniach. Treści ujęte w tych znacznikach czy w polach baz danych mogą zostać w sposób automatyczny zagregowane i przekształcone w indeksy geograficzne, osobowe lub rzeczowe (wraz z odnośnikami).

[49] Ciula, Eide, „Modelling in Digital Humanities"; Flanders, Jannidis, *Data Modeling*; Flanders, Jannidis, *The Shape of Data in the Digital Humanities*; Arianna Ciula, Øyvind Eide, Cristina Marras, Patrick Sahle, *Modelling Between Digital and Humanities: Thinking in Practice*, Cambridge (UK): Open Book Publishers, 2023.

Za punkt wyjścia przyjęto założenie o potrzebie modelowania danych historycznych, niezależnie od dyskusji na temat naturalnego napięcia między danymi i modelami oraz występowaniem jednostkowych zjawisk, które wymykają się wszelkim strukturom czy matrycom. Model danych historycznych będzie zawsze uproszczeniem złożonej rzeczywistości i kompromisem idącym dalej niż opis tej rzeczywistości w języku naturalnym. W tym kontekście model danych historycznych poddaje się tym samym rygorom krytycznej oceny, co tekst historyczny – autor modelu danych podlega podobnym uwarunkowaniom i ograniczeniom jak autor opracowania historycznego. Inne są tylko metody, forma oraz narzędzia wyrażenia własnej interpretacji zjawisk historycznych. W przedstawionej tutaj perspektywie modelowanie danych nie służy wyłącznie reprezentacji informacji ze źródeł historycznych, ale także wspiera odkrywanie i właściwe definiowanie relacji (własności) między zjawiskami w przeszłości.

Różnicą w stosunku do istniejących wcześniej podejść w modelowaniu danych humanistycznych jest przyjęcie założenia wielo- lub multimodelu, który odpowiada kolejnym etapom (elementom) przygotowania edycji cyfrowej: opracowaniu tekstu, przygotowaniu aparatu rzeczowego i tekstowego, ustaleniu kompozycji graficznej edycji lub odmiennych form publikacji (w tym aplikacji internetowych). W istniejących ujęciach model danych edycji przeważnie odwołuje się do ustalonych standardów ogólnych, które służą do zbudowania generalnej struktury tekstu (TEI). Zbudowanie bardziej złożonych schematów, uwzględniających specyfikę źródła i tematu, wymusza zastosowanie przynajmniej dwóch typów modelowania: źródłowe (*bottom-top*, *source driven*), którego celem jest rekonstrukcja struktury informacji w źródle, oraz krytyczne (*top-bottom*, *problem driven*), które nadaje zgromadzonym danym strukturę uwzględniającą standaryzację danych oraz łączenie z innymi zbiorami referencyjnymi, dziedzinowymi czy pochodzącymi z innych edycji. Dyskusję wokół modelowania danych w edytorstwie należy odnieść także do koncepcji jezior danych (*data lake*) oraz hurtowni danych (*data warehouse*). Za organizowaniem danych humanistycznych w postaci rozproszonej w odrębnych zbiornikach (*silo*) przemawia ich specyfika – rozproszenie źródeł informacji, w tym edycji cyfrowych, wysoki stopień granulacji oraz niejednorodność struktur, niepewność oraz niekompletność informacji historycznych. Należy zrezygnować z budowania trudnych w modelowaniu złożonych struktur obejmujących całe kompleksy zjawisk, co wymusza duży stopień generalizacji, która pociąga za sobą daleko idące uproszczenia w prezentowanych danych. Lepszym rozwiązaniem jest budowanie małych (na ile to możliwe) i modułowych zbiorów danych (silosów), które dzięki właściwym referencjom zastosowanych własności i klas oraz metadanym będą

wzbogacały zasób *linked data* i staną się dostępne zarówno dla badaczy, jak i narzędzi AI[50].

Przeniesienie aparatu naukowego i objaśnień kontekstowych do formy ustrukturyzowanej jest tylko pozornym zubożeniem edycji cyfrowej w porównaniu z edycją analogową. Najważniejsze kwestie interpretacyjne i wyjaśniające dotyczące procesów oraz zjawisk opisanych w źródle niezmiennie powinien zawierać wstęp źródłoznawczy. Można uznać, że poziom głębi interpretacyjnej w edycji cyfrowej jest nawet większy niż w edycji tradycyjnej, gdyż edycja cyfrowa wymaga zbudowania całego złożonego modelu dla prezentowanych przez źródło informacji, a nie jedynie przygotowania – często oderwanych od siebie w odrębnych przypisach – objaśnień dla miejsc, osób czy wydarzeń. Fakt, że modelowanie danych ma charakter otwarty i rozwojowy sprawia, że odbiorca zyskuje większą swobodę w interpretacji faktów źródłowych i nie jest ograniczony do jednej, podanej przez edytora wersji interpretacyjnej. Narzędzia cyfrowe umożliwiają szereg operacji o charakterze analitycznym (np. wyszukiwanie różnic, agregowanie danych, wykonywanie różnego rodzaju obliczeń statystycznych) na danych opracowanych w procedurze edycji, które może zdefiniować samodzielnie historyk, odbiorca edycji.

Na obecnym etapie rozwoju edytorstwa cyfrowego, w którym przodują nauki filologiczne, bardziej rozwinięty i dojrzały, lepiej ustandaryzowany, jest sposób opisu tekstu i jego własności niż informacji, którą ten tekst niesie. Powszechnie stosowany standard XML TEI spełnia swoje zadanie na poziomie opisu metadanych, własności i struktury tekstu. Nie pozwala jednak w sposób efektywny modelować informacji, którą ten tekst zawiera, zwłaszcza w kwestii rozwinięcia aparatu naukowego o pozaźródłowe informacje dotyczące osób, miejsc, instytucji oraz relacji, które je łączyły w przeszłości, a które nie są zawarte wprost w treści opracowywanego źródła. Proste informacje dodatkowe o elementach tekstu można oczywiście uwzględnić w nagłówku edycji, ale trudno zbudować w ten sposób bardziej złożone struktury danych. Na skutek tych trudności narodziła się naturalna potrzeba łączenia możliwości, jakich dostarcza standard XML TEI, z możliwościami modelowania danych w bazach danych o strukturze relacyjnej lub grafowej. Dodatkową korzyścią płynącą z takiego podejścia jest możliwość wzbogacania danych poza tekstem edycji oraz wiązanie ich z innymi zbiorami danych przy wykorzystaniu technologii informacyjnych. Ze względu

[50] Jérôme Darmont, Cécile Favre, Sabine Loudcher, Camille Noûs, "Data Lakes for Digital Humanities", w *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, red. Everardo Reyes-García, Gérald Kembellec, Fatma Siala-Kallel, Lilia Sfaxi, Malek Ghenima, Saleh Imad (New York, NY: Association for Computing Machinery), 2020.

na elastyczność operowania modelem i danymi poza formatem XML TEI należy poważnie rozważyć zastosowanie w edytorstwie źródeł historycznych rozwiązań, które łączyłyby dwa środowiska: XML oraz środowisko bazodanowe. W ten sposób można ograniczyć zakres tagowania tekstu w XML-u do aparatu krytycznego oraz odwołań do bazy danych (tag <rs>). Informacje ze źródła byłyby zamodelowane na poziomie analitycznym w innym środowisku (np. RDF, SQL). Przy takim rozwiązaniu edycja cyfrowa składałaby się z dwóch powiązanych ze sobą części – tekstu zaanotwanego w standardzie TEI oraz bazy danych zawierającej informacje (odpowiednik tradycyjnych przypisów rzeczowych). Baza danych mogłaby mieć charakter otwarty, być rozwijana i powiązana z zewnętrznymi zbiorami danych dziedzinowych i referencyjnych (*linked data*). Model danych edycji ma w tym przypadku postać hybrydową (XML TEI, zewnętrzna baza danych). W tym kontekście kluczowa dla rozwoju edytorstwa cyfrowego staje się trwałość zbiorów danych referencyjnych i dziedzinowych, która jest budowana na podstawie stałych identyfikatorów[51] (*persistent identifiers*) oraz wersjonowania „pełnych" edycji cyfrowych, które będą zawierały zarówno warstwę tekstową, jak i warstwę informacyjną pobieraną z baz danych zewnętrznych.

BIBLIOGRAFIA

Bieńkowski, Ludomir, Jerzy Kłoczowski, Zygmunt Sułowski, red. *Zakony męskie w Polsce w 1772 roku*. Lublin: Towarzystwo Naukowe KUL, 1972.

Buzzoni, Marina. „A Protocol for Scholarly Digital Editions? The Italian Point of View". W *Digital Scholarly Editing. Theories and Practices*, red. Matthew James Driscoll, Elena Pierazzo, 64–6. Cambridge: Open Book Publishers, 2016.

Ciula, Arianna, Øyvind Eide. „Modelling in Digital Humanities: Signs in Context". *Digital Scholarship in the Humanities* 32 (2016): i33–i46.

Ciula, Arianna, Øyvind Eide, Cristina Marras, Patrick Sahle. *Modelling Between Digital and Humanities*. Cambridge, UK: Open Book Publishers, 2023.

Darmont, Jérôme, Cécile Favre, Sabine Loudcher, Camille Noûs. „Data Lakes for Digital Humanities". W *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress (DTUC '20)*, red. Everardo Reyes-García, Gérald Kembellec, Fatma Siala-Kallel, Lilia Sfaxi, Malek Ghenima, Saleh Imad, 1–4. New York, NY: Association for Computing Machinery, 2020.

Driscoll, Matthew James, Elena Pierazzo, red. *Digital Scholarly Editing. Theories and Practices*. Cambridge: Open Book Publishers, 2016, https://www.openbookpublishers.com/books/10.11647/obp.0095.

[51] Dla polskich danych historycznych taką bazą jest stworzona w ramach projektu „Cyfrowa infrastruktura badawcza dla humanistyki i nauk o sztuce DARIAH-PL". Nr projektu: POIR.04.02.00-00-D006/20 baza WikiHum: wikihum.lab.dariah.pl [dostęp: 8.06.2025].

Eide, Øyvind. „Ontologies, Data Modeling, and TEI'. *Journal of the Text Encoding Initiative* 8 (2014): 1–22.

Flanders, Julia, Fotis Jannidis. „A Gentle Introduction to Data Modeling". W *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*, red. Julia Flanders, Fotis Jannidis, 26–95. London–New York: Routledge, 2020.

Flanders, Julia, Fotis Jannidis. „Data Modeling". W *A New Companion to Digital Humanities*, red. Susan Schreibman, Ray Siemens, John Unsworth, 229–37. Malden, MA–Chichester, West Sussex, UK: Wiley-Blackwell, 2016.

Flanders, Julia, Fotis Jannidis. „Data Modeling in a Digital Humanities Context: An Introduction". W *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*, red. Julia Flanders, Fotis Jannidis, 1–20. London–New York: Routledge, 2020.

Flanders, Julia, Fotis Jannidis, red. *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*. London–New York: Routledge, 2020.

Gadamer, Hans-Georg. *Reason in the age of science*, trans. Frederick G. Lawrence. Cambridge, MA: MIT Press, 1982.

Garbacz, Paweł, Robert Trypuz, Bogumił Szady, Piotr Kulicki, Przemysław Grądzki, Marek Lechniak. „Towards a Formal Ontology for History of Church Administration". W *Formal Ontology in Information Systems: Proceedings of the Sixth International Conference (FOIS 2010)*, red. Antony Galton, Riichiro Mizoguchi, t. 209, 345–58. Amsterdam: IOS Press, 2010.

Harrower, Natalie, Maciej Maryl, Timea Biro, Beat Immenhauser, red. *ALLEA. Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities*. Berlin: ALLEA, 2020.

Joannis Dlugossii Annales seu Cronicae Incliti Regni Poloniae: Liber Duodecimus 1462-1480. Kraków: Polska Akademia Umiejętności, 2005 (1472 r.).

Kozak, Adam. *Księga sądowa gnieźnieńskich wikariuszy generalnych Sędka z Czechla i Jana z Brzóstkowa (1449–1453, 1455). Studium źródłoznawcze i edycja krytyczna*. Poznań–Warszawa: Poznańskie Towarzystwo Przyjaciół Nauk–Polskie Towarzystwo Historyczne, 2023.

Kürbis, Brygida. „Metody źródłoznawcze wczoraj i dziś". *Studia Źródłoznawcze* 24 (1979): 83–96.

Kürbis, Brygida. „Osiągnięcia i postulaty w zakresie metodyki wydawania źródeł historycznych". *Studia Źródłoznawcze* 1 (1957): 54–7.

Kürbis, Brygida, Gerard Labuda. „Nowa seria Monumenta Poloniae Historica". *Studia Źródłoznawcze* 1 (1957): 210–18.

Labuda, Gerard. „Próba nowej systematyki i nowej interpretacji źródeł historycznych". *Studia Źródłoznawcze* 1 (1957): 3–52.

Litak, Stanisław. *Atlas Kościoła łacińskiego w Rzeczypospolitej Obojga Narodów w XVIII wieku*. Lublin: Towarzystwo Naukowe KUL, 2006.

Litak, Stanisław. *Kościół łaciński w Rzeczypospolitej około 1772 roku. Struktury administracyjne*. Lublin: Wydawnictwo KUL, 1996.

Litak, Stanisław. *Struktura terytorialna Kościoła łacińskiego w Polsce w 1772 roku*. Lublin: Towarzystwo Naukowe KUL, 1980.

Losada Palenzuela, José Luis. „Podstawy naukowej edycji cyfrowej". W *Od Gutenberga do Zuckerberga. Wstęp do humanistyki cyfrowej*, red. Adam Pawłowski, 199–231. Kraków: Universitas, 2023.

Ławrynowicz, Agnieszka. „Metodyka budowy ontologii ‘ontohgis’”. W *Metodologia tworzenia czasowo-przestrzennych baz danych dla rozwoju osadnictwa oraz podziałów terytorialnych*, red. Bogumił Szady, 405–7. Warszawa: Instytut Historii PAN, 2025. https://doi.org/10.5281/zenodo.3751265.

Maryl, Maciej, Marta Błaszczyńska, Bartłomiej Szleszyński, Tomasz Umerle, „Dane badawcze w literaturoznawstwie”. *Teksty Drugie*. *Teoria literatury, krytyka, interpretacja* nr 2 (2021): 13–44.

McCarty, Willard. *Humanities Computing*. Basingstoke–New York: Palgrave Macmillan, 2014.

Müllerowa, Lidia. *Sieć parafialna Kościoła katolickiego w Polsce w 1980 roku*. Lublin: Towarzystwo Naukowe KUL, 1982 (Materiały do Atlasu Historycznego Chrześcijaństwa w Polsce, 5).

Panecki, Tomasz. „Cyfrowe edycje map dawnych: perspektywy i ograniczenia na przykładzie mapy Gaula/Raczyńskiego (1807–1812). *Studia Źródłoznawcze. Commentationes* 58 (2020): 185–206.

Panecki, Tomasz. „Mapping Imprecision: How to Geocode Data from Inaccurate Historic Maps”. *ISPRS International Journal of Geo-Information* 12, nr 4 (2023): 149.

Rolińska, Justyna. „Jak tworzyć cyfrowe edycje źródeł historycznych? Kilka uwag do nowej instrukcji wydawniczej”. W *Edytorstwo źródeł od instrukcji do edycji*, red. Adam Perłakowski, t. 4, 275–325. Kraków: Towarzystwo Wydawnicze Historia Iagellonica, 2022.

Sahle, Patrick. „What is Scholarly Digital Edition?”. W *Digital Scholarly Editing. Theories and Practices*, red. Matthew James Driscoll, Elena Pierazzo, 19–39. Cambridge: Open Book Publishers, 2016.

Skolimowska, Anna. „Korespondencja Jana Dantyszka jako laboratorium edytorstwa cyfrowego”. W *Człowiek twórcą historii*, t. 5: *Warsztat nowoczesnego humanisty historyka na progu XXI w.*, cz. 1, red. Cezary Kuklo, Wojciech Walczak, 71–96. Białystok: Uniwersytet w Białymstoku, 2024.

Słoń, Marek. „Pryncypia edytorstwa cyfrowego w dobie cyfrowej”. *Studia Źródłoznawcze* 53 (2015): 155–61.

Słoń, Marek, Michał Słomski. „Edycje cyfrowe źródeł historycznych”. W *Jak wydawać teksty dawne*, red. Karolina Borowiec, Dorota Masłej, Tomasz Mika, Dorota Rojszczak-Robińska, 65–84. Poznań: Wydawnictwo Rys, 2017.

Spadini, Elena, Francesca Tomasi, Georg Vogeler, red. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*. Norderstedt: Books on Demand, 2021.

Szady, Bogumił. „Inżynieria ontologiczna w badaniach geograficzno-historycznych”. W *Człowiek twórcą historii*, t. 5: *Warsztat nowoczesnego humanisty historyka na progu XXI w.*, cz. 1, red. Cezary Kuklo, Wojciech Walczak, 97–117. Białystok: Uniwersytet w Białymstoku, 2024.

Szady, Bogumił. „Technologie cyfrowe w badaniach historycznych”. W *Od Gutenberga do Zuckerberga. Wstęp do humanistyki cyfrowej*, red. Adam Pawłowski, 235–63. Kraków: Universitas, 2023.

Szady, Bogumił, Tomasz Panecki. „Source-driven Data Model for Geohistorical Records’ Editing: A Case Study of the Works of Karol Perthées”. *Miscellanea Geographica* 26, nr 1 (2022): 52–62.

Tandecki, Jerzy, Krzysztof Kopiński. *Edytorstwo źródeł historycznych*. Warszawa: Wydawnictwo DiG, 2014.

Topolski, Jerzy. *Teoria wiedzy historycznej*. Poznań: Wydawnictwo Poznańskie, 1983.

Vogeler, Georg. „The ‘Assertive Edition’. On the Consequences of Digital Methods in Scholarly Editing for Historians”. *International Journal of Digital Humanities* 1, nr 2 (2019): 309–22, https://doi.org/10.1007/s42803-019-00025-5.

Wilkinson, Mark D., i in. „The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific Data* 3, nr 1 (2016): 1–9.

Wiśniowski, Eugeniusz. *Prepozytura wiślicka do schyłku XVIII wieku. Materiały do struktury organizacyjnej*. Lublin: Towarzystwo Naukowe KUL, 1976.

Wittern, Christian, Arianna Ciula, Conal Tuohy, „The Making of TEI P5". *Literary and Linguistic Computing* 24, nr 3 (2009): 281–96.

Zbíral, David, Robert L. J. Shaw, Tomáš Hampejs, Adam Mertel. *Model the Source First! Towards Source Modelling and Source Criticism 2.0*. Zenodo, 2021. DOI: 10.5281/zenodo.5218926, 2021.

Związek, Tomasz. „Mechanizmy poboru podatków nadzwyczajnych na początku XVI wieku na przykładzie kaliskich rejestrów poborowych". *Kwartalnik Historii Kultury Materialnej* 70, nr 1 (2022): 3–33.