

**ARCHIWIZACJA DANYCH I OPROGRAMOWANIE**

Zbigniew Sawiński

## UWAGI NA TEMAT ROLI ARCHIWÓW DANYCH W NAUKACH SPOŁECZNYCH

*Wypowiedź nie ma charakteru systematycznej analizy roli archiwum danych w naukach społecznych. W kilku punktach staram się jedynie zasygnalizować wybrane problemy, które wiążą się z rozwojem archiwizacji danych z badań społecznych w Polsce.*

### DWIE STRATEGIE ARCHIWIZACJI DANYCH

Można wyróżnić dwa podejścia do archiwizacji danych. Pierwsze z nich opiera się na założeniu, że ciężar merytorycznego przygotowania danych spoczywa na instytucji, która dostarcza dane do archiwizacji. Zgodnie z tą strategią archiwum nie ingeruje w dostarczone zasoby danych, lecz dystrybuuje je w takiej postaci, w jakiej je otrzymało. W tym sensie archiwum pośredniczy jedynie w kontaktach pomiędzy instytucjami dostarczającymi dane a ich odbiorcami.

Drugie podejście zakłada odpowiedzialność archiwum za postać dystrybuowanych danych. Z jednej strony, przybiera ona formę skodyfikowanych wymogów, jakie dostarczane dane powinny spełniać, aby mogły zostać włączone do zasobów archiwum. Z drugiej zaś strony, w przypadku niektórych zbiorów danych włączanych do zasobów, archiwum bierze na siebie trud ich merytorycznego przygotowania do archiwizacji i dystrybucji. Dotyczy to przede wszystkim danych cennych dla społeczności naukowej, które z różnych przyczyn nie zostały odpowiednio przygotowane przez instytucję je gromadzącą (np. dane realizowane we wcześniejszych latach i dostępne jedynie na kartach perforowanych).

Strategie reprezentowane przez istniejące na świecie archiwa są z reguły wypadkową tych dwóch podejść. W większości przypadków dominuje jednakże podejście pierwsze, zgodnie z którym archiwum jedynie pośredniczy w wymianie danych. Rozwiązanie takie przyjmuje się na ogół ze względu na ograniczone możliwości archiwum. Dla przykładu, największe archiwum danych dla nauk społecznych no świecie, czyli ICPSR (Inter-university Consortium for Political and Social Research, University of Michigan), przyjmuje rocznie do archiwizacji około 250 zbiorów danych, to jest w przybliżeniu jeden zbiór dziennie. Mimo faktu, że archiwum zatrudnia prawie 40 osób, nie jest w stanie poddać merytorycznej analizie zawartości dostarczanych danych. Środki, którymi archiwum dysponuje, pozwalają jedy-

nie na przygotowanie technicznej dokumentacji zbioru, obejmującej informacje o zakresie tematycznym badania, charakterystyki jego realizacji (typ próby, rodzaje stosowanych narzędzi badawczych), charakterystyki plików danych (jak sposób zapisu, rodzaje dokumentacji maszynowej) i inne tego rodzaju podstawowe informacje, bez których użytkownik nie byłby w stanie ocenić, czy dane będą przydatne dla jego potrzeb.

W nielicznych przypadkach archiwa podejmują pracę nad merytorycznym opracowaniem zbiorów danych. Na ogół realizuje się to w ramach odrębnych projektów badawczych, ponieważ koszty finansowe takich przedsięwzięć są niebłahe. Prace podjęte przez archiwum obejmują wtedy skompletowanie danych i dokumentacji badawczej, kontrolę plików danych (kontrola zakresów i logiczną) wraz z ewentualnym poprawieniem błędów bądź przynajmniej ujednoczeniem reguł zapisu, opracowanie maszynowej dokumentacji plików danych (jak np. *codebook* w formacie SPSS lub OSIRIS) oraz – na ogół – przygotowanie dokumentacji w formie publikacyjnej. Jako przykład tego rodzaju przedsięwzięć można podać opracowywanie danych kolejnych edycji Eurobarometru przez ICPSR, lub danych programu ISSP (International Social Survey Program) przez Zentralarchiv für Empirische Sozialforschung w Kolonii (Niemcy).

## ARCHIWUM A ŚRODOWISKO BADACZY

Z przedstawionego obrazu wynika, że rola archiwum w sferze dokumentowania danych jest z natury rzeczy ograniczona. W związku z tym archiwizacja danych ma mocne podstawy wyłącznie w tych społecznościach naukowych, które posiadają tradycję i doświadczenia w realizacji badań i opracowywaniu ich wyników z zastosowaniem standardów komputerowego przetwarzania danych. Postać danych dostarczanych do archiwum jest wtedy pochodną pewnych wspólnych doświadczeń i zasad pracy z danymi tych instytucji, które dane gromadzą, i tych, które z danych korzystają.

Z odmienną sytuacją mamy do czynienia w Polsce. W naszym kraju badacze przez całe lata mieli trudności w dostępie do sprzętu komputerowego i oprogramowania, a także do środków gwarantujących realizację wieloletnich programów badawczych. Prowadziło to do dwóch niekorzystnych zjawisk. Po pierwsze, do rozdzielenia roli badacza – zajmującego się przygotowaniem narzędzi badawczych oraz opracowaniem rezultatów badania – oraz informatyka, który zajmował się obróbką danych na komputerze. Po drugie, do braku kumulacji stosowanych reguł zapisu i przechowywania danych z badań. Stosowane rozwiązania miały na ogół charakter doraźny, zaś dane niekiedy – po ich jednorazowym wykorzystaniu do wykonania konkretnych obliczeń – odstawiano na półkę.

Trudno się dziwić, że w takiej sytuacji globalnej tymczasowości nie odczuwano potrzeby powstania wyspecjalizowanej instytucji, która zajmowałaby się archiwizowaniem danych z badań. Archiwum daje bowiem badaczom wymierne korzyści przede wszystkim wtedy, gdy nauka rozwija się w sposób stabilny i kumulatywny. Przekazując dane do archiwum badacz musi być przekonany, że spełnione zostaną trzy warunki. Po pierwsze, dane nie będą wykorzystane w sposób niezgodny z jego wolą (zaufanie do instytucji archiwizującej dane). Po drugie, dostarczone

dane nie będą odbiegały na niekorzyść od standardów innych archiwizowanych badań (zaufanie do własnych standardów). Po trzecie, dane otrzymane za pośrednictwem archiwum będą reprezentowały należyty poziom metodologiczny, co jest warunkiem ich użyteczności (zaufanie do środowiska).

Pierwszy z warunków wydaje się dość oczywisty, aczkolwiek w Polsce jest chyba nie do końca zinternalizowany. Można się bowiem niekiedy spotkać z przekonaniem, że przekazanie danych do archiwum jest równoważne ich publicznemu udostępnieniu, co powoduje utratę kontroli nad ich wykorzystaniem. Nic bardziej mylnego. Archiwum jest instytucją służebną wobec nauki i nie może z założenia prowadzić działalności sprzecznej z zasadami przyjętymi w środowisku. Każdy badacz ma prawo do opracowania zebranych przez siebie danych w pierwszej kolejności, toteż może wymagać od archiwum okresu karencji, w którym przekazane dane nie zostaną nikomu udostępnione lub też będą udostępniane wyłącznie po każdorazowym otrzymaniu zgody badacza. Złożenie danych w archiwum jest natomiast uzasadnione ze względów praktycznych. Badacz uzyskuje gwarancję, że przypadkowo lub na skutek niekorzystnych okoliczności (np. rozwiązanie instytucji, w której realizował badanie) nie utraci zebranych danych i nawet po latach będzie miał do nich dostęp.

Dwa pozostałe warunki (zaufanie do własnych standardów oraz zaufanie do środowiska) też muszą być spełnione, aby można było mówić o zasadności istnienia archiwum. Jeśli badacz w realizowanych przez siebie badaniach dba o zachowanie najwyższych standardów metodologicznych, to może uważać, że przydatność danych gromadzonych przez innych będzie dla niego mniejsza. Może w tej sytuacji podświadomie unikać wymiany danych za pośrednictwem archiwum. Z drugiej strony, badacz może nie być do końca przekonany o poprawności stosowanych przez siebie rozwiązań i odczuwać obawę, że w przypadku udostępnienia danych narazi się na uzasadnioną krytykę bądź wręcz nastąpi zakwestionowanie sformułowanych przez niego wniosków. Trudno wyrokować, na ile tego rodzaju postawy (będące w konflikcie z fundamentalną zasadą intersubiektywnej kontrolowalności rezultatów naukowych) są w naszym środowisku powszechne. Z pewnością jednak przez wiele lat istniały warunki, aby mogły się one utrwalić.

## NOWE OCZEKIWANIA WOBEC ARCHIWUM

W ostatnim czasie zaistniało wiele okoliczności, które na problem archiwizacji danych w Polsce nakazują spojrzeć z nowej perspektywy. Są to przede wszystkim: ograniczona dostępność informacji na temat istniejących i prowadzonych badań, wzrastająca rola kooperacji międzynarodowej oraz zwiększający się zakres możliwości w zakresie wyboru rozwiązań realizacyjnych podczas prowadzenia badań.

W Polsce liczbę realizowanych badań empirycznych można opisać za pomocą krzywej wykładniczej. Prawie całkowity ich brak w latach czterdziestych i pięćdziesiątych spowodował, że możemy mówić o 20–30-letniej, a więc stosunkowo krótkiej perspektywie czasowej. W latach sześćdziesiątych i siedemdziesiątych projekty empiryczne były stosunkowo nieliczne i uzyskanie orientacji co do zakresu i tematyki prowadzonych aktualnie badań nie było trudne. Zupełnie nowa sytuacja po-

wstała natomiast w latach osiemdziesiątych, zwłaszcza w ich drugiej połowie, gdy rozpoczęto realizację badań opinii publicznej na skalę masową.

W sytuacji, gdy zasoby informacyjne niektórych instytucji doszły do kilkuset zbiorów danych, powstały trudności w sprawnym dostępie do poszukiwanych informacji. Stało się oczywiste, że dotychczasowy system oparty na interpersonalnej komunikacji (mniej więcej było wiadomo, kto nad czym pracuje i jakimi danymi dysponuje) musi zostać zastąpiony przez komputerowe systemy wyszukiwawcze. Wydaje się rzeczą naturalną, że wiodącą rolę w tym zakresie powinno przejąć archiwum danych. Istnieje bowiem potrzeba jednej, centralnej instytucji, w której badacz miałby możliwość uzyskania w miarę kompletnych i aktualnych informacji, gdzie, kiedy oraz czy w ogóle dane na określony temat były zbierane.

Drugą okolicznością nakładającą na archiwum nowe funkcje jest wzrastająca rola kooperacji międzynarodowej. Coraz więcej badaczy dochodzi do przekonania, że efektywna współpraca międzynarodowa wymaga wspólnego języka, również w sferze gromadzenia, przygotowania i prezentacji danych. Partykularny i fragmentaryczny charakter większości dotychczasowych rozwiązań rodzi często potrzebę uzyskania wiedzy na temat tego, w jaki sposób gromadzić i przygotować dane tak, aby móc z nimi wejść do międzynarodowej współpracy. Wydaje się, że archiwum jest właściwą instytucją, w której wiedzę taką można byłoby uzyskać – bądź w formie specjalnie przygotowanych publikacji, bądź poprzez uczestnictwo w szkoleniach i seminariach.

Kolejną sferą funkcjonowania archiwum może być merytoryczne doradztwo w zakresie niektórych decyzji związanych z realizacją badań oraz opracowywaniem ich wyników. Z jednej strony, wiąże się to powstaniem i dynamicznym rozwojem rynku usług badawczych w Polsce. Badacz ma obecnie możliwość zlecenia realizacji badania konkurującym ze sobą instytucjom i firmom, z których każda charakteryzuje się pewną specyfiką. Byłoby dla środowiska sytuacją korzystną, gdyby istniała możliwość uzyskania w tym zakresie pomocy ze strony jednej instytucji, zorientowanej w aktualnej konfiguracji rynku usług badawczych. Drugim ważnym aspektem jest nabierający znaczenia problem legalności oprogramowania używanego do gromadzenia, przetwarzania i analizy danych. Wobec konieczności wydatkowania znacznej niekiedy kwoty na zakup oprogramowania badacz musi dokonać wyboru w ramach dość szerokiej oferty. Wydaje się rzeczą naturalną, że pomocy w tym zakresie powinien szukać w archiwum, które na co dzień operuje zróżnicowanymi zasobami danych i posiada doświadczenia w zakresie zalet i wad dostępnych produktów.

Wymieniłem tutaj trzy sfery działalności archiwum, które być może nie są w pierwszej chwili łączone z tą instytucją. Uważam jednak, że archiwum badań społecznych powinno być instytucją otwartą na wszelkie problemy, z którymi borykają się badacze realizujący projekty empiryczne. Owa otwartość może stać się w naszych warunkach ważną płaszczyzną propagowania idei archiwizowania i wymiany danych, co w istocie stanowi konstytutywną sferę działalności każdego archiwum. Z wymienionych wcześniej powodów idee te nie znalazły w naszym środowisku takiego zrozumienia, jak to miało miejsce w wielu innych krajach. Z pewnością nie jest to sytuacja korzystna dla dalszego rozwoju nauk społecznych w Polsce.