

ROZDZIAŁ 1
Ewolucja infrastruktury informacyjnej

Realizacja zadań publicznych a *Big data*

W rozdziale omówiono pojęcie oraz zadania infrastruktury informacyjnej nowoczesnego państwa, metodę modelowania najważniejszego jej składnika, czyli zasobów informacyjnych, istotę oraz cechy idei *Big data*. Po analizie kierunków zmian zadań infrastruktury informacyjnej przedstawiono, problem stosowania metod *Big data* w procesach realizacji zadań publicznych. W szczególności celem niniejszego opracowania jest udzielenie odpowiedzi na pytanie czy istnieje potrzeba oraz warunki praktycznego wykorzystania tych metod.

I. Wprowadzenie

Zdolność administracji publicznej do realizacji zadań publicznych drogą elektroniczną jest ważnym czynnikiem jej oceny oraz warunkiem uzyskania postępu w budowie społeczeństwa informacyjnego. Bez spójnej i nowoczesnej infrastruktury informacyjnej administracji publicznej nie będzie możliwe interoperacyjne, bezpieczne i wiarygodne współdziałanie systemów informacyjnych (tj. ludzi, systemów, usług) jednostek administracji publicznej, we wzajemnych relacjach i w relacjach z osobami fizycznymi, w szczególności obywatelami oraz podmiotami nienależącymi do administracji publicznej, w szczególności przedsiębiorcami w procesach realizacji zadań publicznych. Szczególną odpowiedzialność za stan infrastruktury informacyjnej ponosi administracja rządowa będąca gestorem zasobów informacyjnych o znaczeniu ogólnokrajowym. Z uwagi na to, że nie ma jednej powszechnie uznanej definicji infrastruktury informacyjnej państwa na użytek artykułu przyjęto poniższą, opisową definicję:

Infrastruktura informacyjna administracji publicznej (państwa) to zbiór procedur, modeli, systemów, zasobów informacyjnych oraz ludzkich, których zadaniem jest zgodne z **prawem** gromadzenie, przechowywanie, przetwarzanie i udostępnianie informacji w celu realizacji zadań publicznych¹.

¹ G. Bliźniuk, B. Szafrąński (red.), Interoperacyjność i bezpieczeństwo systemów informatycznych administracji publicznej, Katowice 2006, s. 11.

Odnosząc się do tej definicji należy postawić pytanie: czy w sytuacji coraz powszechniejszego wykorzystania rozwijanych poza administracją publiczną metod analizy danych, opartych na koncepcji *Big data*, powyższe rozumienie infrastruktury informacyjnej państwa nadal obowiązuje i czy jest ono poprawne i uzasadnione?

W celu udzielenia odpowiedzi na tak sformułowane pytanie tekst został podzielony na trzy zasadnicze części, w których będą omówione:

- 1) problem modelowania zasobów informacyjnych administracji publicznej;
- 2) istota oraz najważniejsze cechy metod *Big data*;
- 3) uwarunkowania wykorzystania metod *Big data* w procesach realizacji zadań publicznych.

W zakończeniu przedstawiono wnioski z przeprowadzonych rozważań, stanowiące jednocześnie odpowiedź na pytanie sformułowane we wprowadzeniu.

II. Modelowanie zasobów informacyjnych państwa

Rozwój i funkcjonowanie infrastruktury informacyjnej państwa zależne są od wielu podmiotów (jednostek administracji publicznej, placówek naukowych, organizacji pozarządowych), ale muszą być koordynowane przez centralną administrację rządową. Przedsięwzięcia koordynacyjne obejmują przede wszystkim:

- 1) wprowadzanie zasad identyfikacji i klasyfikacji obiektów informacyjnych o znaczeniu ogólnokrajowym;
- 2) wprowadzanie norm i reguł interoperacyjności;
- 3) inicjowanie regulacji prawnych w zakresie informacyjnych kontaktów obywateli i przedsiębiorców z administracją publiczną;
- 4) tworzenie stosownych rozwiązań techniczno-technologicznych i organizacyjnych infrastruktury informacyjnej;
- 5) określenie zasad udostępniania publicznych zasobów informacyjnych jednostkom administracji publicznej oraz obywatelom i przedsiębiorcom.

Celem modelowania infrastruktury informacyjnej państwa jest stworzenie metodologicznych i merytorycznych podstaw dla:

- 1) działań koordynacyjnych w obszarach, za które odpowiada rząd;
- 2) identyfikacji istotnych problemów decyzyjnych oraz określenia podmiotów odpowiedzialnych za przygotowanie wariantów oraz podmiotów i formy podejmowania decyzji;
- 3) określenia bazy pojęciowej i klasyfikacyjnej, a następnie stosowania jej w określaniu kierunków dalszych prac prowadzących do modernizacji infrastruktury informacyjnej państwa, w tym przede wszystkim gwarantujących wyższy poziom interoperacyjności i integralności infrastruktury informacyjnej państwa.

W celu racjonalnego ukierunkowania rozwoju infrastruktury informacyjnej państwa konieczne jest tworzenie i sukcesywne rozwijanie modeli zasobów informacyjnych administracji, zaczynając od zasobów, które ze względu na własności referencyjności mają przesądzające znaczenie dla realizacji zadań publicznych. W pracach prowadzo-

nych w ramach Zespołu Komitetu Rady Ministrów ds. Informatyzacji i Telekomunikacji² działającego pod nazwą Zespołu Koordynacji Rozwoju Infrastruktury Informacyjnej Państwa dla określenia tego typu modelu przyjęto nazwę – Centralny Model Danych (w skrócie CMD). Z uwagi na ograniczony czas w pierwszym etapie procesu modelowania Zespół ten skupił się na fragmencie modelu (nazwanym jądrem modelu CMD) obejmującym informacje o osobie fizycznej (obywatelu), podmiocie grupowym (przedsiębiorcy) i informacji o terenie (obiektach przestrzennych). Odnosząc prace nad jądrem CMD do definicji infrastruktury informacyjnej należy stwierdzić, że dotyczą one najważniejszego aspektu tej infrastruktury, lecz oczywiście nie wyczerpują zakresu koniecznych prac. Wyczerpująca relacja z wielomiesięcznych prac Zespołu została przedstawiona i poddana dyskusji na posiedzeniu plenarnym 26.3.2009 r. wspomnianego wyżej Komitetu Rady Ministrów³. Odpowiedni fragment protokołu z tego posiedzenia, na którym Zespół przedstawił istotę podejścia oraz główne wnioski jest następujący: „**komitet przyjął, iż ustalenia Zespołu prowadzą do następujących konkluzji:**

I. Rozwój i funkcjonowanie infrastruktury informacyjnej państwa powinny być koordynowane przez Komitet Rady Ministrów ds. Informatyzacji i Łączności – jako organ wewnętrzny RM – w zakresie:

- 1) wprowadzania zasad identyfikacji i klasyfikacji obiektów informacyjnych o znaczeniu ogólnorozwojowym;
- 2) wprowadzania norm i reguł interoperacyjności;
- 3) inicjowania regulacji prawnych dotyczących informacyjnych kontaktów obywateli i przedsiębiorców z podmiotami publicznymi;
- 4) tworzenia stosownych rozwiązań techniczno-technologicznych i organizacyjnych infrastruktury informacyjnej;
- 5) określenia zasad udostępniania publicznych zasobów informacyjnych (w tym w szczególności rejestrów publicznych) jednostkom administracji publicznej oraz obywatelom i przedsiębiorcom.

II. Modernizacja funkcjonowania podmiotów publicznych musi zapewnić:

- 1) dla każdego zadania publicznego realizowanego przez podmioty publiczne możliwości pełnego lub częściowego zrealizowania (lub co najmniej zainicjowania) tego zadania drogą elektroniczną;
- 2) »odmiejscowiony« i realizowany zgodnie z zasadą »braku wykluczenia« dostęp do zadań publicznych realizowanych drogą elektroniczną;
- 3) jeden punkt elektronicznego kontaktu« dla usług realizowanych drogą elektroniczną;
- 4) zakaz żądania danych, które już znajdują się w zasobach informacyjnych administracji, bez względu na istniejące podziały organizacyjne i kompetencyjne administracji;

² B. Szafrński/KRMI, Stanowisko w sprawie Założeń Centralnego Modelu Danych dla Infrastruktury Informacyjnej Państwa, Prace Zespołu ds. Koordynacji Rozwoju Infrastruktury Informacyjnej Państwa, protokół z posiedzenia KRMI z 17.4.2009 r.

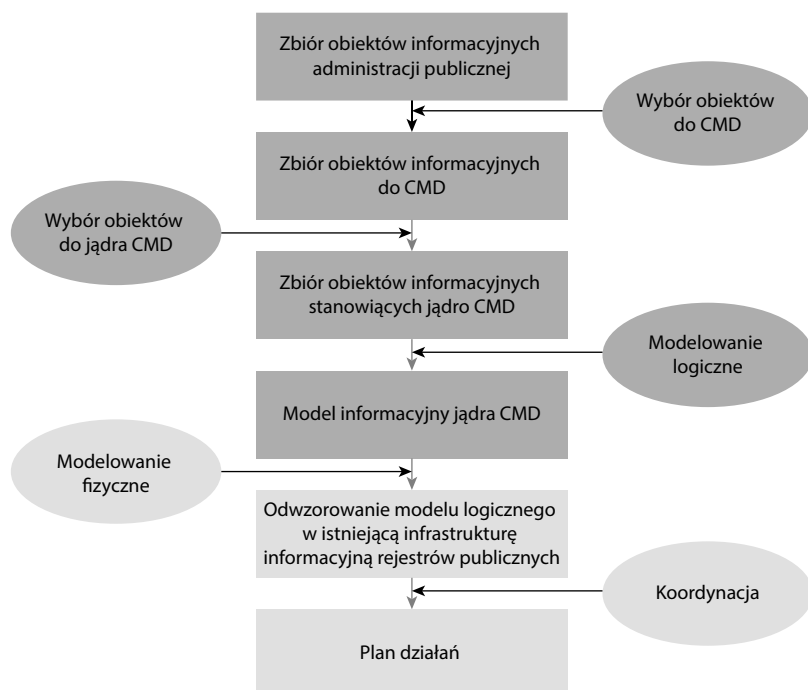
³ *Ibidem*.

- 5) stan, w którym wszystkie dane niezbędne do realizacji zadań publicznych będą dostępne elektronicznie.
- III. Prace Zespołu potwierdziły zasadność tworzenia, utrzymywania i sukcesywnego rozwoju centralnego modelu danych infrastruktury informacyjnej Państwa.
- IV. Konieczne jest stworzenie mechanizmu zapewnienia regularnej oceny stanu rejestrów, obowiązku utrzymania pożądanego stanu oraz prowadzenia, w razie takiej potrzeby, koniecznych działań »naprawczych« przez dysponentów tych rejestrów wg wskazanej metody.
- V. Niezbędne prace powinny koncentrować się przede wszystkim na najbardziej referencyjnych zasobach z punktu widzenia infrastruktury informatycznej Państwa.
- VI. Istnieje pilna konieczność uwzględnienia skutków implementacji dyrektywy INSPIRE w pracach z zakresu modelowania infrastruktury informacyjnej.
- VII. Należy podjąć prace nad modelowaniem infrastruktury informacyjnej państwa w znacznie szerszym zakresie niż to wykonano w ramach prac Zespołu; prace te powinny być prowadzone z wykorzystaniem metod naukowo-badawczych oraz praktycznej wiedzy z dziedziny funkcjonowania administracji publicznej i nowoczesnych technik informacyjnych”.

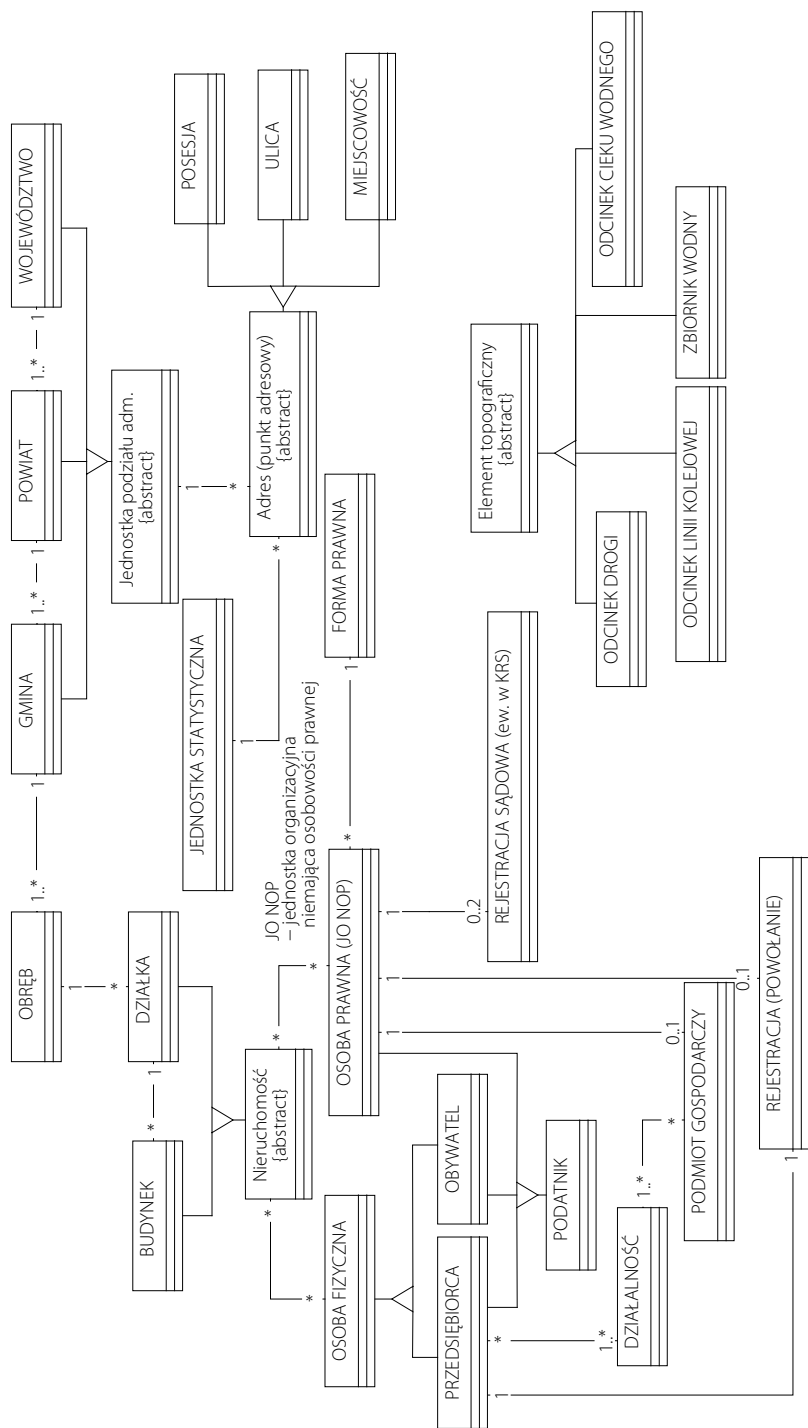
W swych pracach Zespół przyjął kilka podstawowych założeń metodycznych (znaczący udział w opracowaniu tych założeń miał dr inż. *Leopold Żurek*). **Po pierwsze** – proces modelowania rozłożono na dwa poziomy: poziom modelowania logicznego i poziom modelowania fizycznego. Poziom modelowania logicznego jest realizowany przed modelowaniem fizycznym, jako pierwszy. W jego ramach należy wybrać zakres obiektów i atrybutów pełniących rolę referencyjną w infrastrukturze informacyjnej państwa. Na tym etapie należy abstrahować od konkretnych, praktycznie funkcjonujących rejestrów publicznych, lecz kierować się przewidywaną rolą referencyjną atrybutów i charakterem powiązań informacyjnych pomiędzy obiektami. Poziom modelowania fizycznego ma za zadanie przyporządkować wybrane referencje konkretnym rejestrów publicznym (lub wypracować inne decyzje). To założenie zapewnia realną możliwość dokonania zobiektywizowanych ocen i oderwania na poziomie logicznym rozważań od istniejącego, kształtowanego na przestrzeni wielu lat, stanu zasobów informacyjnych. **Po drugie** – należy ograniczyć licznosc obiektów informacyjnych oraz ich atrybutów, aby zwiększyć szanse realizacji tak złożonego przedsięwzięcia, czyli przedsięwzięcia, którego celem jest doprowadzenie do spójności atrybutów rejestrów publicznych z rejestrami określonymi jako referencyjne. Racjonalne ograniczenie zakresu informacyjnego objętego modelowaniem powinno wynikać z analizy wykonywanej w ramach spełniania roli koordynacyjnej w procesie zarządzania rozwojem infrastruktury informacyjnej państwa. Warunkiem koniecznym w tym względzie jest to, by wybrane obiekty informacyjne i ich atrybuty tworzyły „konstrukcję nośną” funkcjonowania infrastruktury informacyjnej państwa, czyli należały do tzw. bazowych zasobów informacyjnych państwa (rejestrów bazowych). **Po trzecie** – należy początkowo skoncentrować uwagę na bytach typu: „istniejący podmiot”, odkładając na następny etap obiekty informacyjne o naturze zdarzeń. Wyjątkiem mogą być zdarzenia, które stano-

wią ważną cechę informacyjną wybranego uprzednio podmiotu (np.: zarejestrowanie działalności gospodarczej przez osobę fizyczną). Rejestry tego typu (tzn. ewidencjonujące istniejące podmioty) mają cechę większej stabilności i ich udział w realizacji zadań publicznych trwa przez dłuższy czas w porównaniu z rejestrami zdarzeniowymi. Stąd, ich bardziej długotrwały wpływ na infrastrukturę informacyjną, w której funkcjonują. **Po czwarte** – należy na preferencyjnych warunkach traktować rejestry klasyfikacyjne (klasyfikatory). Klasyfikatory, zgodnie z ich przeznaczeniem, spełniają główną rolę porządkującą infrastrukturę informacyjną. **Po piąte** – przy wyborze referencyjnych atrybutów należy uznać szczególną rolę identyfikatorów poszczególnych obiektów informacyjnych. Identyfikatory, szczególnie identyfikatory systemowe posiadające cechę pełnej jednoznaczności, odpowiadają za spójność informacyjną wszystkich rejestrów funkcjonujących w infrastrukturze informacyjnej. **Po szóste** – dla potrzeb koordynacyjnych należy budować model na poziomie konceptualnym, abstrahującym od implementacyjnych cech fizycznych obiektów informacyjnych. Założenie to wynika z konieczności zapobieżenia dominacji aspektów technologicznych nad aspektami informacyjnymi.

Poniżej dla zilustrowania metody omówiono przykładowe wyniki otrzymane w procesie modelowania infrastruktury informacyjnej na poziomie logicznym. Podział na poziomy modelowania zobrazowany jest na poniższym schemacie:



Rys. 1. Schemat podziału zadań modelowania na poziomy: logiczny (kolor ■) i fizyczny (kolor □)

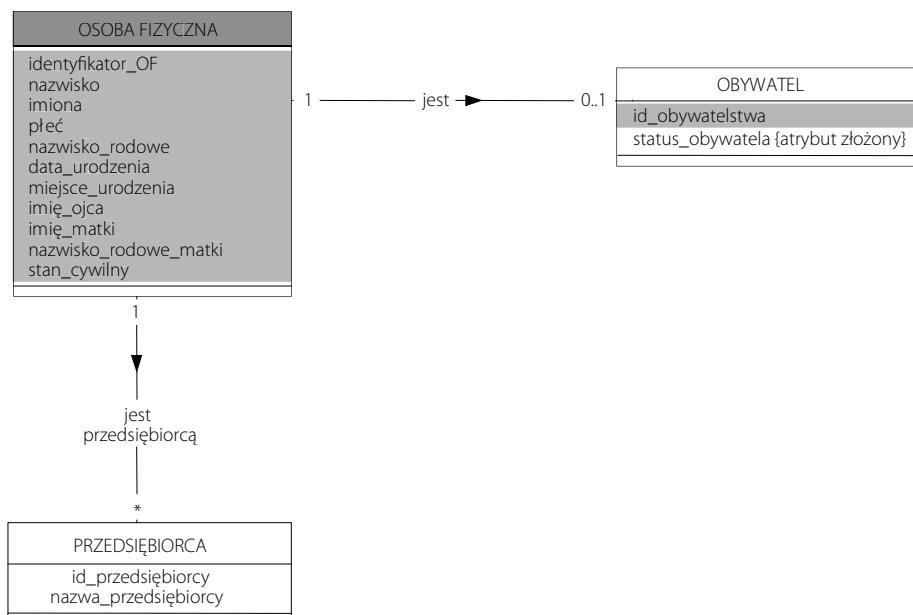


Rys. 2. Ogólny schemat powiązań informacyjnych, jako podstawa rozbudowy CMD

W ramach poziomu modelowania logicznego zrealizowano kolejno:

- 1) analizę podstawowych zasobów informacyjnych w infrastrukturze informacyjnej państwa, w szczególności analizę potrzeb informacyjnych przy realizacji usług publicznych z planowanym wsparciem elektronicznym (analizą zasobów informacyjnych w infrastrukturze informacyjnej państwa objęto przede wszystkim te obszary tematyczne, których dotyczą podstawowe projekty obowiązującego wtedy Planu Informatyzacji Państwa);
- 2) ustalenie bazowych obiektów informacyjnych, od których rozpoczęto modelowanie, a następnie sekwencyjną rozbudowę CMD na podstawie powiązań informacyjnych z kolejnymi obiektami [w wyniku analiz infrastruktury informacyjnej, dyskusji i uzgodnień przyjęto, jako obiekty bazowe – osobę fizyczną, osobę prawną (inaczej podmiot grupowy) oraz obiekt przestrzenny (jak: element topograficzny, nieruchomość, adres, jednostka podziału terytorialnego)];
- 3) określenie zasad selekcji elementów CMD do priorytetowego wdrażania w infrastrukturze informacyjnej państwa, a następnie zastosowanie tych zasad do ograniczenia zakresu CMD; równoległe do procesu rozbudowy CMD w wyniku rozpatrywania kolejnych ról obiektów bazowych, realizowany powinien być proces ograniczający [inaczej: dokonywania wyboru obiektów do „jądra CMD”), który ma za zadanie chronić model przed nadmiernym wzrostem. W tym przypadku zaleca się wziąć pod uwagę co najmniej stopień referencyjności danego obiektu czy atrybutu, czyli jego znaczenie dla spójności i interoperacyjności infrastruktury informacyjnej oraz „skalę” jego wykorzystywania w procesach realizacji zadań publicznych].

Proces rozbudowy CMD musi być iteracyjny. Można go przedstawić jako uwzględnianie w procesach analizy i modelowania obszarów tematycznych (ról), w których biorą udział wybrane główne obiekty. Zaznaczone obszary są efektem uwzględnienia w modelu skutków aktywności funkcjonalnej obiektów bazowych. Schematycznie zobrazowano to na rys. 2 w zakresie kilku wybranych ról (takich, jak: przedsiębiorca, obywatel, podatnik) oraz na rys. 3 – w zakresie powiązań informacyjnych obiektu bazowego osoba fizyczna występująca tu w roli obywatela.



Rys. 3. Powiązania informacyjne obiektów: osoba fizyczna, obywatel

III. Zadania infrastruktury informacyjnej państwa

Zakładana powszechna dostępność zadań publicznych drogą elektroniczną wymusi konieczność dynamicznego zwiększania udziału automatycznego (tj. bez udziału człowieka) dostępu do zasobów informacyjnych administracji publicznej w procesach realizacji tych zadań. Tak rozumiana modernizacja funkcjonowania administracji publicznej musi ponadto zapewnić:

- 1) dla każdego zadania publicznego realizowanego przez administrację istnienie możliwości pełnego lub częściowego zrealizowania (lub co najmniej zainicjowania) tego zadania drogą elektroniczną;
- 2) „odmiejscowiony” i realizowany zgodnie z zasadą „braku wykluczenia” dostęp do zadań publicznych realizowanych drogą elektroniczną;
- 3) „jeden punkt elektronicznego kontaktu” dla usług realizowanych drogą elektroniczną;
- 4) brak żądań danych, które już znajdują się w jej zasobach informacyjnych, bez względu na istniejące podziały organizacyjne i kompetencyjne administracji;
- 5) stan, w którym wszystkie dane niezbędne do realizacji zadań publicznych będące w posiadaniu administracji będą dostępne elektronicznie.

Takie spojrzenie na zadania infrastruktury informacyjnej państwa obowiązywało do momentu dostrzeżenia faktu, że najnowsze technologie informacyjne oparte na wykorzystaniu metod i narzędzi informatyki i telekomunikacji stworzyły warunki do zwiększenia udziału obywateli w procesach informacyjnych państwa. W szczególności