

**Łukasz Żyła**

Fundacja 3.0

l.zyla@media30.pl

## Formaty i opowieści oparte na danych

**Streszczenie:** Otwarte dane w ostatnich latach stanowią słowo-wytrych, poprzez które często próbuje się nakładać na działania instytucji publicznych czy podmiotów wykorzystujących dane płaszczyznę nowoczesności i innowacyjności. Nierzadko jednak kładzie się akcenty na otwarte dane w miejscach o znikomym znaczeniu. Obecnie, w czasach doświadczania świata i komunikacji poprzez bazy danych, bardziej liczy się zakres dostępnych zasobów w minimalnie dostępnym formacie aniżeli mniejszy zakres danych w formacie w pełni przetwarzalnym, np. poprzez API. W niniejszym artykule prezentuję aktualne sposoby otwierania danych i popularnych formatów, biorąc pod uwagę także wymogi prawne oraz przybliżam czytelnikom aspekty narracyjne w formach kulturowych opartych na otwartych danych.

**Słowa kluczowe:** otwarte dane, otwarty format, opowieści oparte na danych, data storytelling

Rozpoczynając ten artykuł od definicji, które są niezbędne do prawidłowego poruszania się w tematyce otwartych danych, skupię się na tych, które według mnie są często mylnie interpretowane i niejednokrotnie nadużywane znaczeniowo. Niewłaściwa terminologia, jak wskażę w tym krótkim artykule (zdaję sobie sprawę, że tekst ten nie wyczerpuje w żaden sposób opisów i analizy wskazywanych zjawisk i tematyki otwartych danych), przyczynia się do realnych działań, mających negatywny wpływ na rozwój społeczeństwa informacyjnego.

Czym są zatem otwarte dane? Zgodnie z definicją zawartą w Ustawie z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego<sup>1</sup> *otwarte dane – informacje sektora publicznego udostępniane lub przekazywane w postaci elektronicznej, bezwarunkowo lub z uwzględnieniem warunków, o których mowa w rozdziale 3, kompletne, aktualne, w wersji źródłowej, w otwartym i niezastrzeżonym formacie przeznaczonym do odczytu maszynowego, które są przeznaczone do bezpłatnego ponownego wykorzystywania na tych samych zasadach dla każdego użytkownika, bez konieczności potwierdzania tożsamości przez użytkownika.* Zwracam uwagę na określenie: *kompletne, aktualne, w wersji źródłowej, w otwartym i niezastrzeżonym formacie przeznaczonym do odczytu maszynowego.* Tak należy rozumieć otwarte dane. W ogólnym znaczeniu raczej nie są informacją tylko sektora publicznego.

Kolejne określenie istotne z punktu widzenia otwartych danych to plik odczytywany maszynowo. Większość danych znajduje się w plikach (jeśli pomijamy moment np. przesyłania ich pomiędzy serwerami), zaś otwartość plików przede wszystkim opiera się na możliwości odczytywania ich maszynowo. W tym miejscu należy wyjaśnić, dlaczego przywiązuje się tak wielką wagę do formatów przetwarzalnych maszynowo. W tym celu cofnę się nieco w czasie. Nasze sposoby percepcji informacji zmienił całkowicie rozwój i powstanie cyfrowych maszyn przetwarzających dane. Pierwsze mechaniczne maszyny liczące powstawały już wiele wieków przed rewolucją przemysłową

---

<sup>1</sup> Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego. Dz.U. 2021, poz. 1641.

(mechanizm z Andikithiry<sup>2</sup> datowany na lata 150–100 p.n.e.), jednak to powstanie komputera w XX w. spowodowało, że zdołano przekształcić informację w sygnał cyfrowy. Rozwój komputerów sprawił, że człowiek coraz więcej czynności związanych z odbiorem i przetwarzaniem informacji scedował na maszynę. Dalszy rozwój maszynowego uczenia i sztucznej inteligencji spowodował z kolei powstanie zewnętrznych agentów, które analizują informacje za człowieka, a w przyszłości będą także za człowieka podejmować decyzje (składać oświadczenie woli?)<sup>3</sup>. W 2016 r. kancelaria prawnicza Baker & Hostetler wykorzystwała w swej pracy system sztucznej inteligencji „Ross” jako „adwokata”. W oparciu o superkomputer Watson firmy IBM system analizujący teksty prawne oraz informacje wejściowe, dokonywał interpretacji prawnych dla działu zajmującego się postępowaniami upadłościowymi<sup>4</sup>. Aby jednak systemy przetwarzania informacji, którym człowiek deleguje coraz więcej zadań, funkcjonowały prawidłowo, dane wejściowe powinny być przygotowane w odpowiednich formatach – formatach do odczytu maszynowego (*machine-readable data*)<sup>5</sup>. Stąd też rozwój tzw. API, czyli interfejsów programistycznych aplikacji.

Ostatnią definicją, na której się skupię jest format danych. Format to inaczej reguły określające sposób zakodowania informacji w pliku. Sposób ten zależy zaś od zastosowanej aplikacji/programu komputerowego. Specyfikacja, inaczej standard, struktury takiego formatu może być powszechnie znana i udostępniona publicznie lub utajniona przez właściciela formatu i producenta programu komputerowego<sup>6</sup>. Potocznie nazywa się je odpowiednio formatem otwartym i formatem zamkniętym. Formatem zamkniętym są np. DOC stworzony i opatentowany przez firmę Microsoft czy PSD stworzony i opatentowany przez firmę Adobe. Pliki z formatami zamkniętymi w większości otwierane i edytowane mogą być jedynie w specjalnie do tego przygotowanych programach. W przypadku pliku DOC jest to program Microsoft Word, w przypadku PSD – program Adobe Photoshop. Oczywiście, właścicielami programów są firmy, które stworzyły dane formaty. W przypadku wymiany pomiędzy dwoma urządzeniami informacji zawartych w pliku aby móc się z nimi zapoznać, na obu urządzeniach musi być zainstalowane odpowiednie oprogramowanie. Format otwarty, w przeciwieństwie do formatu zamkniętego, ma jawną ogólnodostępną specyfikację oraz strukturę nieograniczoną żadnym prawem związanym z licencjonowaniem, patentami czy znakami towarowymi. Format taki może być otwierany poprzez dowolne oprogramowanie obsługujące dany format. Przykładem formatu z otwartym standardem jest np. ODT (stworzony w ramach standardu Open Document Format), który może być otwierany, edytowany i zapisywany przez bezpłatne programy, jak Open Office czy Libre Office oraz inne komercyjne edytory tekstu.

<sup>2</sup> [Przyp. red.] Starożytny przyrząd mechaniczny pełniący rolę astronomicznego kalendarza. Zob.: B. Sieja. „Komputer” sprzed 2 tys. lat. Czym jest i jak działał mechanizm z Antykithiry? *Komputerświat.pl* [online]. [Dostęp 21.10.2021]. Dostępny w: <https://www.komputerwiat.pl/artykuly/redakcyjne/komputer-sprzed-2-tys-lat-czym-jest-i-jak-dzialal-mechanizm-z-antykithiry/jltc4y1>.

<sup>3</sup> STYLEC-SZROMEK, P. Sztuczna inteligencja – prawo, odpowiedzialność, etyka. *Zeszyty Naukowe. Organizacja i Zarządzanie / Politechnika Śląska* 2018, z. 123, s. 501-509. ISSN 1641-3466.

<sup>4</sup> WIERCZYŃSKI, G., WIEWIÓRKOWSKI, W.R. *Informatyka prawnicza: nowoczesne technologie informacyjne w pracy prawników i administracji publicznej*. Wyd. 4. Warszawa: Wolters Kluwer, 2016, s. 89. ISBN 9788380925991.

<sup>5</sup> WILLIAMS, M. E. The impact of machine-readable data bases on library and information services, *Information Processing & Management* [online]. 1977, vol. 13, z. 2, s. 95–107. [Dostęp 21.10.2021]. ISSN 0306-4573. Dostępny w: [https://doi.org/10.1016/0306-4573\(77\)90043-7](https://doi.org/10.1016/0306-4573(77)90043-7).

<sup>6</sup> BŁASZCZYK, A. *Encyklopedia rozszerzeń i formatów plików*. Gliwice: Helion, 2001. ISBN 83-7197-370-5.

## Instytucja

Często mówi się, że otwarte dane (*open data*) to dane, które po prostu mogą być dowolnie wykorzystywane, ponownie użyte i rozpowszechniane przez kogokolwiek<sup>7</sup>. Sam termin otwartości danych od samego początku swym zakresem semantycznym obejmował potencjalne możliwości wykorzystania otwartych danych w budowaniu interoperacyjności<sup>8</sup>. Różnego rodzaju dążenia do otwierania danych publicznych powiązane są z kolei z modelem prowadzenia polityki rządu jako open government. Model ten zakłada właściwe regulacje i działania państwa w zakresie zwiększenia transparentności poprzez odpowiednie udostępnianie danych i informacji publicznych, dając możliwość ponownego ich wykorzystania<sup>9</sup>. Tym samym obecnie otwartość danych często traktuje się jako wyjście naprzeciw obywatelom, pokazanie się w sposób transparentny, otwierając szklane drzwi do gabinetu prezydenta, burmistrza lub wójta. Ta błędna definicja otwartych danych sprawia, że umyka nam coś zasadniczego. Otóż przyjmując techniczną definicję otwartych danych, bez aspektów związanych z teleologicznymi naleciałościami, trzeba przyjąć, że powstaje tutaj pole do manipulacji, dezinformacji czy też inaczej prowadzenia pewnej narracji marketingu politycznego. Nie ulega wątpliwości, że każdy z nas dostrzega zalew stron internetowych instytucji publicznych informacjami, mającymi charakter propagandowy, tworzonymi przez departamenty i wydziały zajmujące się marketingiem politycznym<sup>10</sup>. Instytucje skwapliwie wykorzystują terminologię przyciągającą uwagę, jak „nowoczesne open data”, „innovacyjne rozwiązania smart city”, jako słowa wytrychy, umożliwiające budowania wizerunku otwartej i nowoczesnej instytucji. Tymczasem o tym, co zostanie udostępnione w formacie otwartym jako otwarte dane decyduje sama instytucja. A dane, które obywatele ciekawią mogą być skrzętnie ukrywane.

## Obywatel

Zakładając jednak, że instytucja publiczna nie ma złej woli, zależy jej na rozwoju społeczeństwa informacyjnego i oddaje swoje zasoby informacyjne obywatelom, musi wziąć pod uwagę bardzo ważną kwestię, a mianowicie potrzeby informacyjne swoich potencjalnych odbiorców. I tutaj wiąższość instytucji niestety popełnia błędy nie przeprowadzając żadnych badań rynku, ankiet czy prostych zapytań w mediach społecznościowych. Otwarte dane są tylko wtedy otwarte, jeżeli mają swojego użytkownika. Instytucja otwierając swoje zbiory danych, najpierw powinna rzetelnie poznać zainteresowania i potrzeby potencjalnych odbiorców. Oczywiście nie jest konieczne prowadzenie jakichkolwiek badań, jeżeli instytucja zamierza udostępnić wszystkie swoje dane, choć nawet wtedy warto zapytać: „Jak mamy to zrobić najlepiej?”, „Gdzie powinniśmy zamieścić dane?”

<sup>7</sup> What is Open Data? W: *Open Data Handbook* [online]. [Dostęp 20.10.2021]. Dostępny w: <https://opendatahandbook.org/guide/en/what-is-open-data/>.

<sup>8</sup> YU, H., ROBINSON, D.G. The New Ambiguity of „Open Government”. W: *59 UCLA Law Review Discourse* [online]. 2012, s. 189. [Dostęp 20.10.2021]. Dostępny w: <https://doi.org/10.2139/ssrn.2012489>.

<sup>9</sup> Należy zwrócić uwagę, że popularne ostatnio działania na rzecz otwartych danych czy otwartego rządu wcale nie muszą świadczyć o zwiększaniu dostępności do informacji publicznej, ponieważ ciągle to instytucja publiczna decyduje, jakie dane udostępnia, tworząc przy tym narrację „otwartości”, jednocześnie udostępniając np. dwie bazy danych.

<sup>10</sup> MAZUR, M. *Marketing polityczny: studium porównawcze*. Warszawa: Wydaw. Naukowe PWN, 2019, s.17-18. ISBN 9788301147372.

## Co jest najważniejsze?

Niestety, w procesie udostępniania danych następuje także mylne interpretowanie samych danych. Zwłaszcza w polskich instytucjach publicznych wyższego szczebla publikuje się przede wszystkim ustrukturyzowane dane numeryczne, traktując je jako otwarte dane. Tymczasem oprócz danych numerycznych istnieją także dane nienumeryczne, np. tekstowe, zaś oprócz danych ustrukturyzowanych istnieją dane nieustrukturyzowane, np. pod postacią ciągu liczb z różnego rodzaju czujników. Skupianie się na danych pod postacią tabel zerojedynkowych znacznie ogranicza zakres danych, które instytucje powinny upubliczniać. Co więcej, instytucje te skupiając się na danych ustrukturyzowanych i numerycznych, starają się wejść na możliwie najwyższy poziom ich otwartości (oznaczany gwiazdkami)<sup>11</sup>, udostępniając je np. przez API, a jednocześnie nie dając możliwości ich zwykłego pobrania, np. w formie xlsx. Często zdarza się, że pęd ku nowoczesności i chęć pochwalenia się kolejnymi gwiazdkami przystania cele, jakim ma służyć otwieranie danych. Na przykład, jaki jest sens tworzenia za olbrzymie środki finansowe zbiorów dostępnych przez API, jeśli z takiego rozwiązania skorzysta tylko niewielu programistów, z których część i tak ekonomicznie nie wykorzysta tych zasobów z powodów innych zewnętrznych okoliczności biznesowych. Tymczasem podstawowe dane publiczne nieustrukturyzowane i nienumeryczne o najważniejszych osobach w państwie, czyli oświadczenia majątkowe posłów są publikowane w formie skanów dokumentów papierowych. Na stronach internetowych urzędów miast i gmin znajdziemy wiele zeskanowanych plików PDF, w których osadzone są tabele z Excela. Jednocześnie można zaobserwować pojawiające się tu i ówdzie dedykowane portale open data miast, na których, publikowane są informacje, np. iloma żłobkami miasto dysponuje, czy też na stronie rządowej [Otwarte Dane](#)<sup>12</sup>, jakie było najpopularniejsze imię dla dziewczynek w kwietniu, natomiast o liczbie przeprowadzonych testów na Covid-19 w danym powiecie już się nie dowiemy. Stosowanie niewłaściwych formatów w przypadku istotnych zbiorów danych, a z drugiej strony udostępnianie innych mniej istotnych danych poprzez interfejs API, chwając się poziomem otwartości, to nie tylko kolokwialnie pójście na łatwiznę, ale często część wyrachowanej narracji marketingu politycznego.

## Czyja narracja?

Otwarte dane, zwiększając dostęp do informacji, z pewnością mają wpływ na budowanie i rozwój społeczeństwa informacyjnego. Można powiedzieć, że bez informacji i danych nie istniałoby obecnie wiele gałęzi życia. Starłem się dotychczas wskazać na niektóre pęknięcia, które mogą się pojawiać/pojawiają się już obecnie w zakresie otwierania danych. Co jednak powinno przykuć uwagę badaczy, to zakres korzystania z danych przez społeczeństwo oraz komunikowania czy bardziej doświadczania świata. Jakiś czas temu Lev Manovich<sup>13</sup> wskazywał na olbrzymi rozwój baz danych, podkreślając jednocześnie, że tak jak kiedyś narracja dominowała jako forma doświadczania i opisywania świata, tak teraz wobec natłoku informacji w obrębie nowych mediów swoistym genre (gatunkiem, stylem) komunikacji staje się baza danych, konstytuując bazodanową formację kulturową. W medium tekstu pierwsze oznaki tego typu interfejsu można odnaleźć w *Ulissesie* Jamesa Joyce'a, gdzie dzięki zastosowanym środkom literackim, wątki fabularne nakładają się na siebie

<sup>11</sup> Zob. 5 \* *Open Data* [online]. [Dostęp 24.10.2021]. Dostępny w: <https://5stardata.info/en/>.

<sup>12</sup> Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 19.10.2021 r.

<sup>13</sup> MANOVICH, L. *Język Nowych Mediów*. Warszawa: Wydawnictwa Akademickie – Oficyna Wydawnicza Łośgraf, 2011. ISBN 9788362726349.

i symultanicznie przenikają, tworząc załączki hipertekstualności. Dzisiaj wielokrotnie w poszukiwaniu informacji w rzeczywistości wchodzimy do baz danych i wybieramy z nich te informacje, które nas interesują. Oczywiście na podstawie bazy danych można tworzyć kolejne opowieści, łącząc je z innymi. Warto zaznaczyć, że tak budowane narracje oparte na bazach danych, czyli de facto „stara” forma kulturowa narracji, wykorzystując formę baz danych, daje olbrzymie możliwości zaangażowania odbiorcy, który zatracą się w nowym medium (przykład wykorzystywania przez firmę Facebook algorytmu, który promował nienawistne komentarze czy fake newsy tak, aby zwiększyć ruch reklamowy).

Niestety, sposoby tworzenia opowieści na danych (data storytelling) nie są jeszcze dostatecznie zbadane, mimo że wiele lat temu Lev Manovich twierdził, iż tzw. nowa info-estetyka powinna zdecydowanie zainteresować badaczy. Nie ulega jednak wątpliwości, że im więcej danych publicznych będzie naprawdę otwartych, tym komunikacja i narracje będą zdrowsze dla społeczeństwa (a już zdecydowanie bardziej transparentne niż narracje wielu instytucji publicznych).

## Bibliografia

1. 5 \* *Open Data* [online]. [Dostęp 24.10.2021]. Dostępny w: <https://5stardata.info/en/>.
2. BEREZIŃSKI, M., HOŁUBIEC, J.W., WAGNER, D. Hierarchiczna struktura poznania : piramida wiedzy. *Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą* 2009, nr 19, s. 5–17. ISSN 1732-324X.
3. BERNACZYK, M. *Prawo do informacji publicznej w Polsce i na świecie*. Warszawa: Wydaw. Sejmowe, 2014. ISBN 978-83-7666-344-9
4. BERNSTEIN, J.H. The Data-Information-Knowledge-Wisdom Hierarchy and its antithesis. *CUNY Academic Works* [online]. 6.2009. [Dostęp 19.10.2021]. Dostępny w: [https://academicworks.cuny.edu/kb\\_pubs/11/](https://academicworks.cuny.edu/kb_pubs/11/).
5. BŁASZCZYK, A. *Encyklopedia rozszerzeń i formatów plików*. Gliwice: Helion, cop. 2001. ISBN 83-7197-370-5.
6. CAMPBELL-KELLY, M., ASPRAY, W., ENSMINGER, N., YOST, J.R., *Computer: a history of the information machine*. Boulder, CO: Westview Press, a member of the Perseus Books Group, 2013. ISBN 978-0-8133-4590-1.
7. CASTELLS, M. *The rise of the network society*. 2nd ed. Chichester, West Sussex: Wiley-Blackwell, 2010. ISBN 9781405196864.
8. GOBAN-KLAS, T., SIENKIEWICZ, P. *Spółeczeństwo informacyjne: szanse, zagrożenia, wyzwania*. Kraków: Wydaw. Fundacji Postępu Telekomunikacji, 1999. ISBN 8386476192.
9. GRAHAM, M., DUTTON, H.W. *Society and the Internet: how networks of information and communication are changing our lives*. Oxford; New York: Oxford University Press, 2014. ISBN 9780199662005.
10. GRYSZCZYŃSKA, A. *Rejestry publiczne: jawność i interoperacyjność*. Warszawa: Wydaw. C.H. Beck, 2016. ISBN 9788325586843.
11. HETMAŃSKI, M. *Świat informacji*. Warszawa: Difin, 2015. ISBN 9788379306671.
12. JACOBSON, D., BRAIL, G., WOODS, D. *Interfejs API: strategia programisty*. Gliwice: Helion, cop. 2015. ISBN 9788328305557.
13. JANOWSKI, J. *Transformacja w kierunku cywilizacji informacyjnej: od etycznej do technicznej metody życia społecznego*. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2018. ISBN 9788378148418.
14. MANOVICH, L. *Język Nowych Mediów*. Warszawa: Wydawnictwa Akademickie – Oficyna Wydawnicza Łośgraf, 2011. ISBN 9788362726349.
15. MAZUR, M. *Marketing polityczny: studium porównawcze*. Warszawa: Wydaw. Naukowe PWN, 2019. ISBN 9788301147372.
16. OLEŃSKI, J. *Infrastruktura informacyjna państwa w globalnej gospodarce*. Warszawa: Uniwersytet Warszawski. Wydział Nauk Ekonomicznych, 2006. ISBN 8387374407.
17. SIEJA, B. „Komputer” sprzed 2 tys. lat. Czym jest i jak działał mechanizm z Antykithiry? *Komputerświat.pl* [online]. 9.10. 2021 [dostęp 20.10.2021]. Dostępny w: <https://www.komputerświat.pl/artykuly/redakcyjne/komputer-sprzed-2-tys-lat-czym-jest-i-jak-dzialal-mechanizm-z-antykithiry/jlrc4y1>.
18. STYLEC-SZROMEK, P. Sztuczna inteligencja – prawo, odpowiedzialność, etyka. *Zeszyty Naukowe. Organizacja i Zarządzanie / Politechnika Śląska* 2018, z. 123, s. 501–509. ISSN 1641-3466.

19. Ustawa z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego. Dz.U. 2021, poz. 1641.
20. WANIEK, D. *Dylematy ładu medialnego RP: standardy europejskie a praktyka polityczna* [online]. Kraków: Oficyna Wydawnicza AFM, 2007. [Dostęp 19.10.2021]. ISBN 9788389823434. Dostępny w: <https://repozytorium.ka.edu.pl/handle/11315/18791>.
21. WEBER, R. Applying artificial intelligence in the science & technology cycle. *Information Services & Use* [online]. 2019, vol. 39, nr 4, s. 303–318. [Dostęp 19.10.2021]. ISSN 1875-8789. Dostępny w: <https://content.iospress.com/articles/information-services-and-use/isu190062>.
22. What is Open Data? W: *Open Data Handbook* [online]. [Dostęp 20.10.2021]. Dostępny w: <https://opendatahandbook.org/guide/en/what-is-open-data/>.
23. WIERCZYŃSKI, G., WIEWIÓRKOWSKI, W.R. *Informatyka prawnicza: nowoczesne technologie informacyjne w pracy prawników i administracji publicznej*. Wyd. 4. Warszawa: Wolters Kluwer, 2016. Seria Akademicka. ISBN 9788380925991.
24. WILLIAMS, M.E. The impact of machine-readable data bases on library and information services, *Information Processing & Management* [online]. 1977, vol. 13, z. 2, s. 95–107. [Dostęp 24.10.2021]. ISSN 0306-4573. Dostępny w: [https://doi.org/10.1016/0306-4573\(77\)90043-7](https://doi.org/10.1016/0306-4573(77)90043-7).
25. YU, H., ROBINSON, D.G. The New Ambiguity of „Open Government”. W: *59 UCLA Law Review Discourse* [online]. 2012, s. 178-208. [Dostęp 20.10.2021]. Dostępny w: <https://doi.org/10.2139/ssrn.2012489>.