

**Piotr Krajewski**

piotr.krajewski@pg.edu.pl

ORCID: <https://orcid.org/0000-0002-1422-1993>

**Magdalena Szuflita**

magdalena.szuflita@pg.edu.pl

ORCID: <https://orcid.org/0000-0002-1490-8234>

Biblioteka Politechniki Gdańskiej

## Znaczenie i proces certyfikacji zaufanych repozytoriów danych badawczych – MOST Wiedzy Open Research Data Catalog i certyfikat CoreTrustSeal

**Streszczenie:** W artykule omówiono zagadnienia związane z certyfikacją tzw. zaufanych repozytoriów danych badawczych, ze szczególnym uwzględnieniem certyfikatu CoreTrustSeal, uznawanego za jeden z kluczowych standardów jakości. Proces aplikacji o certyfikat CoreTrustSeal jest szczegółowy i obejmuje gruntowną weryfikację przez ekspertów, co zwiększa zaufanie użytkowników do repozytorium oraz zapewnia długoterminową dostępność i bezpieczeństwo danych. Przykładem polskiego repozytorium spełniającego najwyższe standardy jest MOST Wiedzy Open Research Data Catalog na Politechnice Gdańskiej.

**Słowa kluczowe:** zaufane repozytoria, certyfikacja repozytoriów, CoreTrustSeal

### Wprowadzenie

Otwartość w nauce i wiążące się z nią wymogi udostępniania rezultatów badań, czy to w postaci artykułów, czy to jako zbiory danych (*datasety*), stanowi nieodłączny już element wszelkich procesów badawczych i publikacyjnych. Wydaje się jednak, że zbyt ni nacisk kładziemy na natychmiastowe efekty tych działań, zapominając o celach długofalowych. Oczywiście, kwestie poprawy jakości i wiarygodności wyników, oszczędności finansowej czy też niwelowanie barier w dostępie do informacji o najnowszych osiągnięciach nauki są niezwykle istotne. Jednak równie ważne jest zapewnienie trwałości wyników badań naukowych, przede wszystkim w postaci danych surowych, tak aby mogły one w przyszłości posłużyć innym badaczom. Trzeba mieć tu na uwadze również takie procesy badawcze, których metody jeszcze nie zostały nawet opracowane. Znamienny przykład wykorzystania danych w badaniach o całkowicie innych niż pierwotne założeniach stanowi mapowanie transkryptomów różnych gatunków roślin oraz zwierząt. Chociaż pierwotnym celem gromadzenia tych sekwencji DNA były badania nad ekspresją genów, to prędko dane te okazały się niezwykle przydatne w badaniach ewolucyjnych i śledzeniu filogenetycznych powiązań między poszczególnymi grupami organizmów. Choć często podkreśla się wagę długotrwałego przechowywania danych, można niejednokrotnie odnieść wrażenie, że dane są istotne dla badaczy jedynie do czasu publikacji bazującego na nich artykułu, ewentualnie do momentu rozliczenia grantu. W polskich realiach nie pomaga też tłumaczenie z angielskiego pojęcia „data preservation”. Zwykło się stosować określenie „archiwizowanie danych” które kojarzy się raczej ze składowaniem elementów, które jedynie dokumentują zakończone procesy.

W związku z powyższym badacze bardzo często nie skupiają się na wyborze odpowiedniego repozytorium dla zdeponowania danych. A przecież ta właśnie decyzja stanowi o tym, czy dane będą mogły zostać wykorzystane w przyszłości. Aby pomóc naukowcom w jej podjęciu, powstały różnego rodzaju rejestry dostępnych repozytoriów. Jednym z nich jest baza [re3data.org](https://re3data.org)<sup>1</sup> stanowiąca wykaz wszelkich serwisów internetowych umożliwiających zdeponowanie danych badawczych. Należy tu jednakże podkreślić, że indeksacja repozytorium w tej bazie nie gwarantuje jakości oferowanych przez nie usług.

Innym serwisem indeksującym repozytoria danych badawczych jest założony na Uniwersytecie Oksfordzkim [FAIRsharing.org](https://fairsharing.org), któremu przyświeca idea pomocy naukowcom w wyborze odpowiedniego miejsca do publikowania i dzielenia się wynikami badań w zgodzie z zasadami FAIR (akronim angielskich słów *findable, accessible, interoperable* oraz *reusable*). W tym celu FAIRsharing.org prowadzi oddzielne rejestry dla:

- baz danych (w skład których wchodzi repozytoria),
- polityk ustanowionych dla uczelni, czasopism naukowych czy instytucji finansujących badania, związanych z zarządzaniem, przechowywaniem i udostępnianiem szeroko pojętych danych naukowych,
- wszelkiego rodzaju mniej lub bardziej formalnych norm i standardów dotyczących tej tematyki.

Chociaż informacje o repozytoriach przesyłane przez użytkowników są weryfikowane, to same rekordy zawierają jedynie podstawowe dane. Co jest zrozumiałe, skoro celem jest ewidencja, a nie ewaluacja repozytoriów. Jak w związku z tym wyłonić repozytoria, które są najbardziej wiarygodne? I co, tak naprawdę, ta wiarygodność oznacza?

## Potrzeba certyfikacji

Zaufane repozytorium cyfrowe (z angielskiego Trustworthy Data Repository, TDR) definiuje się jako „takie, którego misją jest zapewnienie niezawodnego, długoterminowego dostępu do zarządzanych zasobów cyfrowych swoim użytkownikom, teraz i w przyszłości” (RLG/OCLC Working Group on Digital Archive Attributes; 2002). Otwarte i wiarygodne usługi w zakresie zarządzania danymi naukowymi zapewniają:

- długoterminowe zarządzanie i dostarczanie międzynarodowej społeczności naukowej danych i/lub usług w zakresie danych o ocenie jakości,
- przechwytywanie i przechowywanie, kuratorstwo, długoterminowe przechowywanie, odkrywanie, dostęp, pobieranie, agregację, analizę i/lub wizualizację.

Chociaż problematyka oceny wiarygodności repozytoriów oraz jakości świadczonych przez nie usług stanowi jeden z najważniejszych obszarów działań organizacji związanych z inicjatywą otwartej nauki oraz otwartymi danymi badawczymi, dotychczas nie przyjęto jednego zestawu kryteriów i systemu ewaluacji europejskich repozytoriów danych badawczych. Warto jednak podkreślić, że certyfikat CoreTrustSeal wymieniany jest niezmiennie jako jeden z najważniejszych elementów mających stanowić podstawę takiego systemu.

---

<sup>1</sup> Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną z dn. 27.08.2024.

Certyfikat CoreTrustSeal ma stosunkowo krótką historię, która jest ściśle związana z rozwojem standardów i najlepszych praktyk w zarządzaniu danymi badawczymi. W miarę wzrostu ilości danych badawczych oraz ich roli w procesie naukowym inicjatywy, takie jak ISO 16363 i Data Seal of Approval (DSA) (wprowadzona w 2009 r. przez holenderskie Data Archiving and Networked Services (DANS)), zaczęły kłaść fundamenty dla certyfikacji repozytoriów danych (L'Hours i in., 2019). W 2008 r. Międzynarodowa Rada Nauki (International Council for Science, ICSU) utworzyła światowy system danych World Data System (WDS) jako globalną sieć repozytoriów danych i usług, które spełniają wysokie standardy jakości. W 2014 r. w odpowiedzi na rosnące zapotrzebowanie na spójne i jednolite standardy inicjatywy DSA i WDS rozpoczęły współpracę. W roku 2017 powstał CoreTrustSeal. Nowa organizacja i certyfikat mają na celu zapewnienie, że repozytoria danych spełniają zharmonizowane i ujednolicone kryteria jakości. CoreTrustSeal przejął najlepsze praktyki z obu wcześniejszych inicjatyw, tworząc 16 podstawowych wymagań, które muszą spełniać certyfikowane repozytoria.

CoreTrustSeal jest dynamicznie rozwijającą się inicjatywą. Organizacja regularnie aktualizuje swoje wymagania i procedury certyfikacyjne, aby nadążać za zmianami w technologii, przepisach prawnych i najlepszych praktykach w zarządzaniu danymi badawczymi. CoreTrustSeal jest „wstępem” do uzyskania kolejnych certyfikatów takich jak: Nestor Seal DIN 31644 oraz ISO 16363. Dodatkowo, we wrześniu 2020 r. przedstawiciele międzynarodowej organizacji FAIRsFAIR oraz rozwijanej przez Komisję Europejską inicjatywy European Open Science Cloud (EOSC) podpisali deklarację współpracy. Jednym z wymienionych obszarów kooperacji było wypracowanie jednolitych kryteriów oceny oraz działania mające na celu uzyskanie certyfikatu CoreTrustSeal przez jak największą liczbę repozytoriów związanych z EOSC.

## **Jak otrzymać CoreTrustSeal**

MOST Wiedzy Open Research Data Catalog otrzymał certyfikat CoreTrustSeal w październiku 2023 r. i stał się jedynym z nielicznych zaufanych repozytoriów otwartych danych badawczych w Polsce. Centrum Usług Informatycznych i Biblioteka Politechniki Gdańskiej dostosowały repozytorium do najwyższych standardów udostępniania danych. Repozytorium MOST Wiedzy, posiadając certyfikat CoreTrustSeal, może być rekomendowane teraz przez Komisję Europejską oraz agencje finansujące badania (np. NCN) jako miejsce deponowania danych badawczych z projektów. Logo CoreTrustSeal na stronie repozytorium gwarantuje naukowcom, że ich dane są zabezpieczone przed utratą, a zasady udostępniania są jasno określone i stabilne w czasie.

By aplikować o certyfikat CoreTrustSeal, konieczne jest spełnienie szeregu kryteriów opisujących, w jaki sposób repozytorium mierzy się z wyzwaniami organizacyjnymi, technicznymi, finansowymi czy też prawnymi. Postępy w realizacji wymogów na całym etapie aplikacji są drobiazgowo weryfikowane i oceniane przez zespół recenzentów. Repozytorium przygotowuje dokumentację aplikacyjną, która obejmuje szczegółowe informacje na temat jego działalności, polityk, procedur i infrastruktury. Dokumentacja ta musi wykazywać, że repozytorium spełnia 16 wymagań CoreTrustSeal (tabela poniżej), które obejmują aspekty, takie jak przechowywanie danych, bezpieczeństwo, polityki dostępu, zarządzanie i organizacja. Po złożeniu aplikacji zostaje ona wstępnie sprawdzona przez zespół CoreTrustSeal (Board) pod kątem kompletności i zgodności z podstawowymi wymaganiami. Aplikacja jest przypisywana dwóm niezależnym recenzentom, którzy mają doświadczenie w zarządzaniu danymi badawczymi. Recenzenci dokładnie analizują aplikację i dostarczone dokumenty i oceniają, w jakim stopniu repozytorium spełnia 16 kryteriów CoreTrustSeal. Każde kryterium jest

ewaluowane indywidualnie, a recenzenci sporządzają szczegółowe raporty z uwagami i sugestiami. Ten etap może być powtarzany maksymalnie pięć razy, aż do akceptacji przez recenzentów i wprowadzenie poprawek przez oceniane repozytorium. Na koniec zespół CoreTrustSeal podejmuje ostateczną decyzję o przyznaniu lub odmowie certyfikacji. Certyfikat CoreTrustSeal jest ważny przez trzy lata. Po tym okresie repozytorium musi przejść proces recertyfikacji, aby potwierdzić, że nadal spełnia wszystkie wymagania. Od 2024 r. opłata za certyfikację wynosi 3000 Euro.

Tab. 1. Wymagania stawiane repozytorium ubiegającemu się o certyfikat CoreTrustSeal

R1	Repozytorium ma wyraźną misję zapewnienia dostępu do danych i zachowania ich w swojej domenie.
R2	Repozytorium przechowuje wszystkie obowiązujące licencje obejmujące dostęp do danych i ich wykorzystanie oraz monitoruje zgodność.
R3	Repozytorium posiada plan ciągłości w celu zapewnienia stałego dostępu do swoich zbiorów i ich ochrony.
R4	Repozytorium zapewnia, w miarę możliwości, że dane są tworzone, nadzorowane, udostępniane i wykorzystywane zgodnie z normami dyscyplinarnymi i etycznymi.
R5	Repozytorium dysponuje odpowiednim finansowaniem i wystarczającą liczbą wykwalifikowanych pracowników zarządzanych w ramach przejrzystego systemu zarządzania, aby skutecznie realizować misję.
R6	W repozytorium przyjmuje się mechanizmy mające na celu zapewnienie wskazówek ekspertów (wewnętrznych lub zewnętrznych), w tym, w stosownych przypadkach, wytycznych naukowych.
R7	Repozytorium gwarantuje integralność i autentyczność danych.
R8	Repozytorium akceptuje dane i metadane w oparciu o zdefiniowane kryteria, aby zapewnić istotność i zrozumiałość dla użytkowników danych.
R9	Repozytorium stosuje udokumentowane procesy i procedury w zakresie zarządzania archiwalnym przechowywaniem danych.
R10	Repozytorium przejmuje odpowiedzialność za długoterminowe przechowywanie i zarządza tą funkcją w sposób zaplanowany i udokumentowany.
R11	Repozytorium dysponuje odpowiednią wiedzą fachową w zakresie jakości danych technicznych i metadanych oraz zapewnia użytkownikom końcowym dostęp do wystarczających informacji do dokonywania ocen związanych z jakością.
R12	Archiwizacja odbywa się zgodnie ze zdefiniowanymi przepływami pracy od pozyskania do rozpowszechnienia.
R13	Repozytorium umożliwia użytkownikom odkrywanie danych i odwoływanie się do nich w sposób trwały poprzez odpowiednie cytowanie.
R14	Repozytorium umożliwia ponowne wykorzystanie danych w czasie, zapewniając dostępność odpowiednich metadanych w celu wsparcia zrozumienia i wykorzystania danych.
R15	Repozytorium funkcjonuje w oparciu o dobrze obsługiwane systemy operacyjne i inne podstawowe oprogramowanie infrastrukturalne oraz korzysta z technologii sprzętowych i programowych odpowiednich dla usług, które świadczy na rzecz wyznaczonej społeczności.
R16	Infrastruktura techniczna repozytorium zapewnia ochronę obiektu i jego danych, produktów, usług i użytkowników

Źródło: CoreTrustSeal Standards and Certification Board, 2022.

## Podsumowanie

Repozytoria danych badawczych powinny się ubiegać o certyfikat CoreTrustSeal z kilku powodów. Przede wszystkim posiadanie tego certyfikatu zwiększa zaufanie użytkowników do repozytorium, świadcząc o jego zgodności z wysokimi standardami zarządzania danymi. Certyfikat ten jest uznawany na arenie międzynarodowej i potwierdza, że repozytorium spełnia wymagania dotyczące bezpieczeństwa, dostępności i integralności danych. Dodatkowo stanowi on czynnik prestiżowy dla uczelni, która może promować się certyfikowaną infrastrukturą. Kolejnym kluczowym aspektem jest długoterminowa dostępność danych. Posiadanie tego certyfikatu świadczy o wdrożeniu przez repozytorium odpowiednich mechanizmów i strategii, które gwarantują, że dane będą dostępne i użyteczne przez długi czas (np. zgodnie z wymaganiami agencji finansujących badania). Proces ubiegania się o certyfikat wymaga także dokładnej analizy i dokumentacji procedur zarządzania danymi. To pomaga repozytorium zidentyfikować obszary do poprawy i wdrożyć lepsze praktyki, co ostatecznie podnosi jakość oferowanych usług.

## Bibliografia:

1. CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. (2021). *CoreTrustSeal Foundation Statutes and Rules of Procedure (v02\_00-2021-06-22)* [online]. Zenodo. [Dostęp 30.07.2024]. Dostępny w: <https://doi.org/10.5281/zenodo.5025884>.
2. CORETRUSTSEAL STANDARDS AND CERTIFICATION BOARD. (2022). *CoreTrustSeal Requirements 2023–2025 (V01.00)* [online]. Zenodo. [Dostęp 30.07.2024]. Dostępny w: <https://doi.org/10.5281/zenodo.7051012>.
3. L'HOURS, H., KLEEMOLA, M., DE LEEUW, L. (2019). CoreTrustSeal: from academic collaboration to sustainable services. *IASSIST Quarterly* [online]. 43(1), s. 1–17. [Dostęp 30.07.2024]. Dostępny w: <https://doi.org/10.29173/iq936>.
4. DE LEEUW, L. (2021). *RDA Data Repository Attributes WG Case Statement CoreTrustSeal Board Feedback (1.0)* [online]. Zenodo. [Dostęp 30.07.2024]. Dostępny w: <https://doi.org/10.5281/zenodo.5801502>.
5. PRÍNCIPE, P., VIEIRA, A., KARIMOVA, Y., MOURA, P. (2023). *CoreTrustSeal Certification – much more than a certification: empowering DataRepositoriUM's processes towards FAIR data and skilled data curators* [online]. DataRepositoriUM. [Dostęp 30.07.2024]. Dostępny w: [https://repositorium.sdum.uminho.pt/bitstream/1822/87898/2/02\\_DataRepositoriUM\\_CoreTrustSeal.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/87898/2/02_DataRepositoriUM_CoreTrustSeal.pdf).
6. RLG/OCLC WORKING GROUP ON DIGITAL ARCHIVE ATTRIBUTES. (2002). *Trusted digital repositories: Attributes and responsibilities: An RLG-OCLC Report* [online]. Mountain View, CA: Research Libraries Group. [Dostęp 30.07.2024]. Dostępny w: <http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>.

---

KRAJEWSKI, P., SZUFLITA-ŻURAWSKA, M. (2024). Znaczenie i proces certyfikacji zaufanych repozytoriów danych badawczych – MOST WIEDZY Open Research Data Catalog i certyfikat CoreTrustSeal. *Biuletyn EBIB* [online]. Nr 3(214), Różnorodne odsłony otwartości w nauce, edukacji i kulturze. [Dostęp 12.09.2024]. ISSN 1507-7187. Dostępny w: <https://ebibojs.pl/index.php/ebib/article/view/923>.