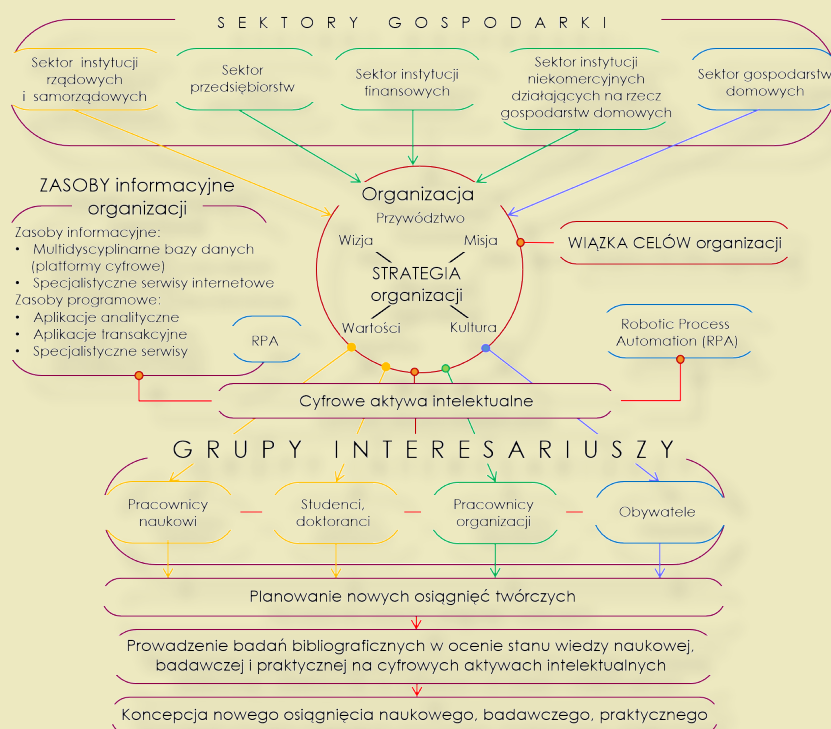




Bogdan Wit

Robotic Process Automation w zarządzaniu cyfrowymi aktywami intelektualnymi w organizacji



MONOGRAFIE

Robotic Process Automation
w zarządzaniu cyfrowymi aktywami
intelektualnymi w organizacji

Monografie – Politechnika Lubelska

Rada Naukowa Wydawnictwa Politechniki Lubelskiej

Przewodnicząca:
Agnieszka RZEPKA

Dyrektor CIN-T:
Dorota TKACZYK

Wydawnictwo Politechniki Lubelskiej:

Magdalena CHOŁOJCZYK
Jarosław GAJDA
Anna KOŁTUNOWSKA
Katarzyna PEŁKA-SMĘTEK

Przedstawiciele Dyscyplin Naukowych Politechniki Lubelskiej:

Marzenna DUDZIŃSKA
Małgorzata FRANUS
Arkadiusz GOLA
Paweł KARCZMAREK
Beata KOWALSKA
Anna KUCZMASZEWSKA
Jarosław LATAJSKI
Tomasz LIPECKI
Zbigniew ŁAGODOWSKI
Joanna PAWŁAT
Lucjan PAWŁOWSKI
Natalia PRZESMYCKA
Magdalena RZEMIENIAK
Mariusz ŚNIADKOWSKI

Przedstawiciele honorowi:

Zhihong CAO, Chiny
Miroslav GEJDOŠ, Słowacja
Karol HENSEL, Słowacja
Hrvoje KOZMAR, Chorwacja
Frantisek KRCMA, Czechy
Sergio Lujan MORA, Hiszpania
Dilbar MUKHAMEDOVA, Uzbekistan
Sirgii PAWŁOW, Ukraina
Natalia SAVINA, Ukraina
Natia SHENGELIA, Gruzja
Daniele ZULLI, Włochy

Bogdan Wit

Robotic Process Automation w zarządzaniu cyfrowymi aktywami intelektualnymi w organizacji



POLITECHNIKA
LUBELSKA
WYDAWNICTWO

Lublin 2022

Recenzenci:

dr hab. inż. Marcin Gąsior, prof. Politechniki Lubelskiej

dr hab. inż. Waldemar Jędrzejczyk, prof. Politechniki Częstochowskiej

Monografia naukowa wykonana w ramach prac badawczych Katedry Inżynierii Systemów Informacyjnych Wydziału Zarządzania w zakresie automatyzacji i wykorzystania programowych robotów cyfrowych w różnych obszarach życia społecznego i gospodarczego.

Publikacja wydana za zgodą Rektora Politechniki Lubelskiej

© Copyright by Politechnika Lubelska 2022

ISBN: 978-83-7947-529-2

Wydawca: Wydawnictwo Politechniki Lubelskiej

www.biblioteka.pollub.pl/wydawnictwa

ul. Nadbystrzycka 36C, 20-618 Lublin

tel. (81) 538-46-59

Druk: Soft Vision Mariusz Rajski

www.printone.pl

Elektroniczna wersja książki dostępna w Bibliotece Cyfrowej PL www.bc.pollub.pl

Książka udostępniona jest na licencji Creative Commons Uznanie autorstwa – na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0)

Nakład: 50 egz.

Spis treści

Wprowadzenie	9
1. Metodyka zautomatyzowania prowadzenia badań bibliograficznych	42
1.1. Proces opracowania zbioru założeń badań bibliograficznych	49
1.2. Proces pozyskania zasobów cyfrowych	58
1.3. Proces wykonania zapytania bibliograficznego i wyszukiwanie w zasobach RID.....	71
1.4. Proces analizy ilościowej i jakościowej danych.....	76
1.5. Proces poddania pełnotekstowej analizie treści dokumentów	77
1.6. Proces oceny wyników z konsumpcją zebranych informacji	78
1.7. Proces weryfikacji i walidacji badań bibliograficznych.....	79
2. Zarządcze i organizacyjne aspekty wdrażania RID	87
2.1. Cykl życia produktu cyfrowego RID.....	87
2.2. Zarządzanie zasobami RID.....	97
2.3. Zarządzanie ciągłością działania RID	105
2.4. Zarządzanie jakością produktów cyfrowych, danych i e-usługi.....	115
3. Rekomendowane działania w zarządzaniu zasobami ludzkimi.....	126
3.1. Rekomendacje dla użytkowników.....	126
3.2. Rekomendacje dla koordynatora	129
3.3. Rekomendacje dla operatora robota cyfrowego	132
Zakończenie	138
Załącznik A. Uogólniona metodyka badania bibliograficznego w cyklu życia badań bibliograficznych.....	143
Załącznik B. Metodyka zautomatyzowanego prowadzenia badań bibliograficznych z wykorzystaniem RPA	144
Załącznik C. Przykłady funkcjonalności aplikacji analitycznej w badaniach bibliograficznych na zasobach RID.....	147
Bibliografia.....	157
Spis rysunków	160

Streszczenie

Zarządzanie grupami lub zespołami monodyscyplinarnych lub interdyscyplinarnych specjalistów dla osiągnięcia większej sprawności działania jest procesem złożonym, gdyż dotyczy nie tylko interakcji społecznych, ale również odnosi się do cyfrowych aktywów intelektualnych. W planowym, celowym rozwoju dojrzałości organizacji istotne są prace badawcze – badania bibliograficzne, które mają na celu stwierdzić określony stan wiedzy naukowej, badawczej i praktycznej na podstawie cyfrowych aktywów intelektualnych. W oparciu o fakty ustala się poziom deficytu wiedzy, jej względnego nieuporządkowania co stanowi przyczynek do koncepcji nowego, poszerzonego lub pogłębionego twórczego osiągnięcia naukowego, badawczego, praktycznego, które jest odpowiedzią mającą wypełnić w sposób całościowy lub częściowy stwierdzoną lukę osiągnięć twórczych. Jednym z podejść do prowadzenia badań bibliograficznych jest archiwalne przechowywanie zbiorów danych źródłowych pozyskanych z badań ilościowych oraz analiza zebranych danych wtórnych przez użytkowników.

Tematyka monografii obejmuje aspekty metodyczne, przedsięwzięcia projektowe, aplikacyjne, zarządzania zasobami wiedzy oraz ludzkimi związane z automatyzacją prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej i praktycznej dla grup i zespołów specjalistów, wspomagana programowalnymi robotami cyfrowymi. Przedstawiony autorski sposób wykorzystania robotyki w badaniach bibliograficznych skutecznie łączy rekurencyjne pozyskiwanie zasobów w sposób zautomatyzowany z aktualizacją zgromadzonych zasobów oraz mapowaniem tematycznym zainteresowanych użytkowników w zespołach monodyscyplinarnych i interdyscyplinarnych. Opracowana metodyka w sposób znaczący umożliwia zmniejszenie lub zniwelowanie zidentyfikowanych siedem barier w badaniach bibliograficznych, w ten sposób przyczyniając się do wzrostu sprawności, efektywności w zarządzaniu zasobami ludzkimi w organizacjach.

Słowa kluczowe: Zarządzanie zespołami badawczymi, automatyzacja badań bibliograficznych, programowalne roboty cyfrowe, robotyzacja procesów biznesowych, sztuczna inteligencja, repozytoria instytucjonalne, duże zbiory danych wtórnych.

Robotic Process Automation in managing an organization's digital intellectual assets

Abstract

Managing groups or teams of monodisciplinary or interdisciplinary professionals for achieving greater agility is a complex process, as it not only concerns social interactions, but also relates to digital intellectual assets. In the planned, purposeful development of the organization's maturity, research work–bibliographic research, which aims to ascertain a certain state of scientific, research and practical knowledge on the basis of digital intellectual assets, is important. On the basis of facts, the level of knowledge deficit and its relative disorder is determined, which is a contribution to the concept of a new, extended or deepened creative scientific, research and practical achievement, which is an answer meant to fill, in whole or in part, the identified gap of creative achievements. One of the approaches to conducting bibliographic research is archival storage of source data sets acquired from quantitative research and analysis of collected secondary data by users.

The subject matter of the monograph includes methodological aspects, design, application, human resources management related to the automathematization of bibliographic research in the evaluation of scientific, research and practical knowledge for groups and teams of specialists, supported by programmable digital robots. The presented author's method of using robotics in bibliographic research effectively combines recursive retrieval of resources in an automated manner with the actualization of collected resources and thematic mapping of interested users in monodisciplinary and interdisciplinary teams. The developed methodology has significantly enabled the reduction or elimination of the identified seven barriers to bibliographic research, thus contributing to an increase in efficiency, effectiveness and effectiveness of human resource management in organizations.

Keywords: Research team management, bibliographic research automation, Robotic Process Automation (RPA), business process robotics, artificial intelligence, institutional repositories, large secondary data sets.

Akronimy

AI	– <i>Artificial Intelligence</i>
API	– <i>Application Programming Interface</i>
BCM	– <i>Business Continuity Management</i>
CAPTCHA	– <i>Completely Automated Public Turing Test to Tell Computers and Humans Apart</i>
GUI	– <i>Graphical User Interface</i>
HTML	– <i>HyperText Markup Language</i>
ICT	– <i>Information and Communication Technology</i>
OCR	– <i>Optical Character Recognition</i>
PDCA	– <i>Plan – Do – Check – Act</i>
MZPBB	– <i>Metodyka zautomatyzowanego prowadzenia badań bibliograficznych</i>
PDF	– <i>Portable Document Format</i>
RID	– <i>Repozytorium Instytucjonalne Danych</i>
RPA	– <i>Robotic Process Automation</i>
RSS	– <i>Really Simple Syndication</i>
VBA	– <i>Visual Basic for Applications</i>
W&W	– <i>Weryfikacja i Walidacja</i>
WWW	– <i>World Wide Web</i>
WoS	– <i>Web of Science</i>
XML	– <i>Extensible Markup Language</i>

Wprowadzenie

Funkcjonowanie przedsiębiorstw i rozwój gospodarczy determinowany jest w dużym stopniu od tworzenia i wykorzystania wiedzy w organizacjach. Wiedza indywidualna, zespołowa i instytucjonalna ewoluuje dzięki innowacjom w badaniach i nauce. W trakcie tego procesu wzrasta jej wartość merytoryczna a wykorzystana wiedza nabiera znaczenia praktycznego. W tworzeniu sformalizowanej nowej wiedzy istotne stają się procesy pozyskiwania, utrwalania, przechowywania, opracowywania, udostępnienia i wykorzystania wiedzy w organizacjach. Proces tworzenia nowej wiedzy w organizacjach w sposób najbardziej ogólny można przedstawić w dwóch etapach. Pierwszy to zbieranie faktów w celu zdobywania, zgłębiania wiedzy o stanie nauki i techniki dla określenia parametrów granicznych badań w tworzeniu nowej, rozszerzonej lub pogłębionej wiedzy. Drugi etap to proces tworzenia, kodyfikacji, przechowywania, weryfikacji i walidacji, upublicznienia oraz wykorzystania wiedzy. Wymienione etapy są niepowtarzalne, unikatowe i specyficzne dla określonej wiedzy w różnych organizacjach.

W niniejszej monografii przedstawiono proces dotyczący opisu automatyzacji pozyskania aktywów intelektualnych z rozproszonych źródeł danych w celu utworzenia scentralizowanego zasobu w organizacji, który jest określany jako Repozytorium Instytucjonalne Danych (RID). Repozytorium RID ulokowano w chmurze cyfrowej¹ w dostępie kontrolowanym, w trybie świadczenia e-usług specjalistom dla zapewnienia skutecznej realizacji procesów pozyskiwania, tworzenia i wykorzystania wiedzy. Całokształt opisanych działań jest wynikiem doświadczeń autora z wdrożenia RID podczas pracy z różnymi przedsięwzięciami realizowanymi w ramach pracy zawodowej i eksperckiej.

Zgromadzone zasoby intelektualne w utrwalonej postaci są określane cyfrowymi aktywami intelektualnymi i wykorzystane na etapie prowadzenia badań bibliograficznych przez zespoły, grupy specjalistów w ocenie stanu wiedzy naukowej, badawczej i praktycznej w pracach nad nowymi osiągnięciami twórczymi w oparciu o analizę zawartości zebranych zasobów w repozytorium. Głównym celem utworzenia RID było zwiększenie efektywności działania specjalistów zorganizowanych w zespoły lub grupy badawcze w organizacji. Ze względu na specyfikę działalności uczelni, wiedza i jej wykorzystanie podlega ocenie jakościowej wewnętrznej i zewnętrznej przez instytucje zewnętrzne związane z ewaluacją jakości działalności naukowej.

¹ Termin „chmura cyfrowa” jest równoważny terminowi „chmura obliczeniowa”.

Również osiągnięcia twórcze są weryfikowane, oceniane przez społeczność naukową, akademicką, różnych pracowników organizacji i obywateli². Z tego względu przedstawiony zakres merytoryczny monografii dotyczący tworzenia, wykorzystania produktu cyfrowego jakim jest RID jest specyficzny, gdyż obejmuje prowadzenie badań bibliograficznych na danych wtórnych w ocenie stanu wiedzy naukowej, badawczej i praktycznej uwarunkowane kontekstem wewnętrznym i zewnętrznym uczelni wyższej. Działania w zakresie tworzenia i utrzymania instytucjonalnego repozytorium dokumentów wpisują się zapisy ustawy o szkolnictwie wyższym w art. 11. „1. Podstawowymi zadaniami uczelni są: (...) 3) prowadzenie działalności naukowej, świadczenie usług badawczych oraz transfer wiedzy i technologii do gospodarki; 4) prowadzenie kształcenia doktorantów; 5) kształcenie i promowanie kadr uczelni; 6) stwarzanie osobom niepełnosprawnym warunków do pełnego udziału w: (...), c) prowadzeniu działalności naukowej; 9) upowszechnianie i pomnażanie osiągnięć nauki i kultury, w tym przez gromadzenie i udostępnianie zbiorów bibliotecznych, informacyjnych i archiwalnych; 10) działanie na rzecz społeczności lokalnych i regionalnych.”³. Z zasobów RID mogą korzystać w formie e-usługi różni specjaliści z różnych grup zawodowych, okreśłani jako użytkownicy RID, np. pracownicy, studenci, doktoranci, obywatele społeczności lokalnych i regionalnych, w tym także osoby z niepełnosprawnościami.

W założeniu zasoby zgromadzone w RID były, są i będą wykorzystane w badaniach bibliograficznych (przeglądach, studiach, badaniach literaturowych⁴) w uczelni wyższej^{5,6,7,8} i dotyczą oceny trzech, nieraz zależnych od siebie grup osiągnięć twórczych (wiedzy): naukowych, badawczych i praktycznych.

² Mikołajuk, L. (2015). *Repozytorium instytucjonalne jako istotny element promocji nauki*. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

³ Ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, (Dz.U. 2018 poz. 1668).

⁴ Zdonek, I., Hysa, B., & Zdonek, D. (2016). *Publikacje przeglądowe w naukach o zarządzaniu – istota i tendencje*. Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska.

⁵ Adam, U. A. ., & K., K. (2021). *Driving forces behind the management of Institutional Repositories: Qualitative evidences*. Malaysian Journal of Library & Information Science, 26(3), 33–56. Retrieved from <https://sare.um.edu.my/index.php/MJLIS/article/view/24444>

⁶ Ujournunna, J. C., & Agbawudikeizu, E. C. (2021). *Management of digital records for knowledge preservation in academic institutions for a sustainable development in education sector*. Law, Business and Sustainability Herald, 1(3), 5–9.

⁷ Haddow, G., & White, H. (Eds.). (2021). *Assessment as Information Practice: Evaluating Collections and Services*. Routledge.

⁸ Ahmed, J., Kataria, S., Dey, T., & Singh, S. (2021). *Digital Repository Services at Bennett University: A Case Study*. Library Philosophy and Practice, 1–12.

Ocena wiedzy naukowej w poznaniu naukowym jest niezbędna w celu określenia postępu tej wiedzy, deficytu wiedzy i luki naukowej, co stanowi wyzwanie twórcze dla specjalistów zajmujących się procesami naukowymi. W tym przypadku, poznanie naukowe dotyczy czynności, sposobów identyfikacji i ekstraktu określonej wiedzy z wytworów w postaci dokumentów cyfrowych. Nowa wiedza naukowa może mieć wymiar teoretyczny, badawczy lub praktyczny. Jednocześnie nie każda nowa wiedza ma charakter naukowy.

Ocena wiedzy badawczej dotyczy procesu prowadzenia badań, np. w ramach metodologii, metodyk, metod, narzędzi i technik badawczych. Prowadzenie badań ma szczególne znaczenie dla wiedzy naukowej, przy czym nie każde badania muszą mieć charakter naukowy. Wiele wyników badań są informacjami nowymi, przypominającymi, aktualizacyjnymi, weryfikującymi, mają charakter poznawczy, zwiększają liczbę przypadków przebadanych i przeanalizowanych, jednakże nie zwiększają bezpośrednio zasobów wiedzy naukowej⁹.

Ocena wiedzy praktycznej odnosi się do wiedzy inżynierskiej, technicznej lub dotyczącej zastosowań praktycznych wiedzy naukowej lub badawczej i jej wykorzystania w różnych aspektach. Najczęściej wymieniane są aspekty społeczne, ekonomiczne (gospodarcze) z uwzględnieniem oddziaływania na środowisko przyrodnicze.

Wynikiem badań bibliograficznych jest stwierdzenie określonego poziomu deficytu wiedzy, jej względnego nieuporządkowania co stanowi przyczynek do koncepcji nowego, poszerzonego lub pogłębionego twórczego osiągnięcia naukowego, badawczego, praktycznego, które jest odpowiedzią mającą wypełnić w sposób całościowy lub częściowy stwierdzoną lukę osiągnięć twórczych.

Badania bibliograficzne posiadają swój cykl życia, który tworzy zestaw zależnych od siebie procesów, działań, od inicjujących aż do zamykających. Proces prowadzenia badań bibliograficznych aktualnego stanu wiedzy w celu wychwycenia luk naukowych, badawczych, praktycznych nie jest prosty, napotyka na wiele barier o różnorodnym charakterze. Poniżej zostanie wymieniony niewyczerpujący wykaz najważniejszych barier o różnym znaczeniu i wpływie na odkrywaniu wiedzy z badań bibliograficznych przez grupy lub zespoły specjalistów.

⁹ Poskrobko. B. (2009). *Nauka i naukowość w ekonomii i zarządzaniu*. Optimum, Studia Ekonomiczne Nr 4 (44), s. 15.

BARIERA 1:

Bariera w zakresie znajomości, wykorzystania różnych, metod i technik badań bibliograficznych przez specjalistów

Specjaliści prowadzący badania bibliograficzne stosują wypracowane i sobie znane metody oraz techniki badawcze. Wykorzystują rozwiązania zaimplementowane w serwisach internetowych¹⁰, dedykowane oprogramowania lub pakiety komputerowe. Jednak użytkownicy mają świadomość problemów i ograniczeń (aspektów), które ujawniają się w trakcie wykonywania badań bibliograficznych z wykorzystaniem standardowych metod i technik osadzonych w oprogramowaniu.

W pierwszym aspekcie bariery w zakresie metod i technik analizy, specjaliści dostrzegają, że istnieje wiele alternatywnych rozwiązań, ich kombinacji oraz sposobów wyselekcjonowania, pozyskania, przechwytywania, przetwarzania zbioru dokumentów. Otrzymane wyniki badań bibliograficznych mogą się znacznie różnić między sobą. Najprostszą techniką utworzenia repozytorium jest wyszukiwanie dokumentów w serwisie Google Scholar i stworzenie osobistej kolekcji dokumentów w postaci biblioteki. Na bibliotecę można wykonywać proste operacje: dodawać, usuwać, oznaczać pozycje, cytować, grupować z wykorzystaniem etykiet oraz udostępniać bibliotekę publicznie¹¹.

W drugim aspekcie bariery, z pozyskanego zbioru dokumentów należy wyodrębnić informacje z nośnika oraz dokonać interpretacji. Nośnik jest elektroniczną reprezentacją, najczęściej zorientowaną na strony agregacją różnego typu danych: tekstowych, wartości liczbowych, obiektów graficznych, hiperłączy oraz ewentualnie zawiera indeks danych lub metadane przydatne do identyfikacji, zrozumienia i przedstawiania tych danych. Przetwarzaniu podlega zarówno zawartość dokumentu (treść), czyli różnego typu dane jak i nośnik danych (rys. 1a). W przypadku platform cyfrowych (multidyscyplinarnych baz danych) czy specjalistycznych serwisów internetowych stosowane są różne mechanizmy przetwarzania dokumentów i ich nośników. Najczęściej użytkownik serwisu przeszukuje zawartość zgromadzonych danych po tytułach,

¹⁰ Antasari, I. W. (2022). *Influence of Offline and Online Access Skills on Motivation to Access Institutional Repositories*. Library Philosophy and Practice, 2022(Feb).

¹¹ Biblioteka Google Scholar, <https://scholar.google.com/intl/pl/scholar/help.html#library>, data dostępu: 12.01.2022 r.

w streszczeniach, po słowach kluczowych lub co jest rzadkością w pełnych treściach całostronicowych dokumentów lub też eksportuje albo dokonuje zapisu dokumentu w oryginalnej postaci. Przetwarzanie nośników informacji jest również istotne, gdyż ustawione atrybuty decydują o przetwarzaniu nie tylko nośnika, ale i zawartości. Nośnik dokumentu można poddać archiwizacji, przetwarzaniu, sortowaniu, indeksowaniu oraz zmieniać jego atrybuty.

W trzecim aspekcie bariery istotne jest zidentyfikowanie i pozyskanie danych, które są umieszczone na nośniku i opisują dokument elektroniczny, tj. informację i jej nośnik w postaci indeksu dokumentu lub metadanych. Uogólniając, dane opisujące nośnik informacji oraz dane opisujące zawartość nośnika danych określane będą metadanymi. Metadane mogą być przechowywane w odpowiednich miejscach dokumentu lub w oddzielnych dokumentach będących z nimi w relacji najczęściej typu 1:1 (rys. 1b). Na przykład zbiór metadanych w dokumencie PDF¹² jest umieszczony wewnątrz i jest opcjonalny. Użytkownicy platform cyfrowych baz danych czy specjalistycznych serwisów internetowych przeszukują po metadanych oraz mogą je pozyskiwać lub zapisywać w postaci oddzielnych dokumentów. Dokumenty zawierające tylko metadane, które są w relacji z dokumentem źródłowym są plikami stowarzyszonymi z dokumentami źródłowymi. Pozyskanie metadanych cyfrowych do nośników oraz informacji jest niezbędne do oceny ilościowej, ale przede wszystkim służą one do oceny jakościowej informacji. Metadane podlegają przetwarzaniu maszynowemu¹³, umożliwiają cytowanie lub inne operacje przetwarzania na nośniku lub informacji¹⁴.

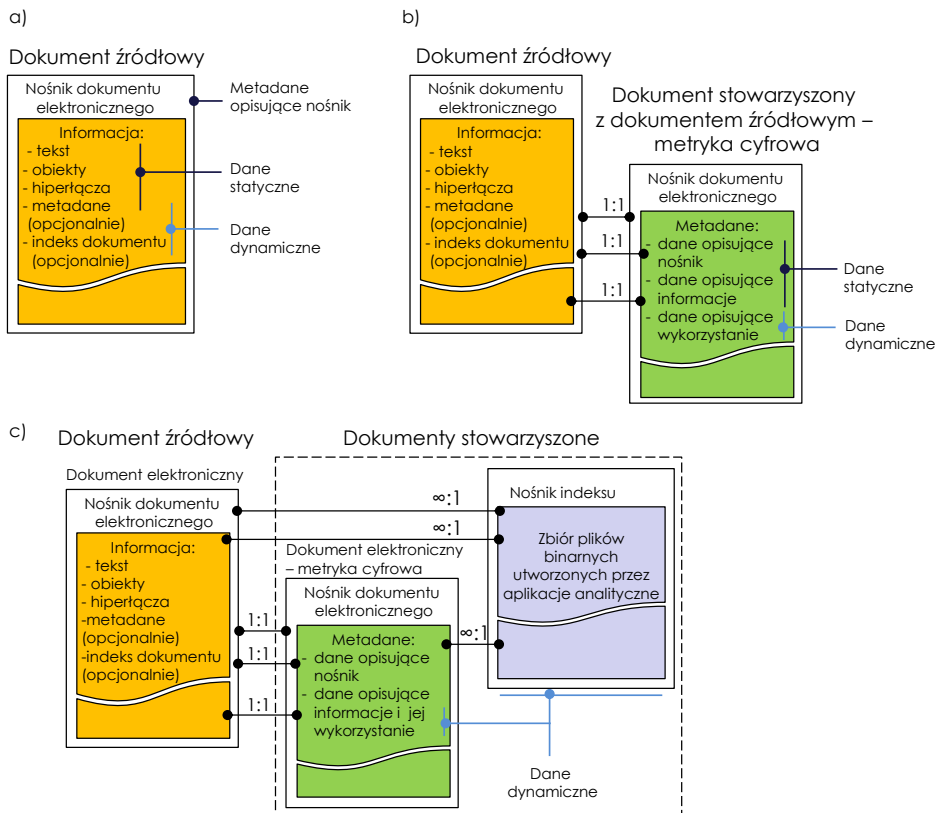
Czwarty aspekt bariery stanowi wykorzystywanie indeksów danych. Tak jak metadane indeks danych może być umieszczony opcjonalnie w dokumencie PDF lub zostać zapisany w postaci oddzielnego dokumentu (rys. 1c).

¹² *Portable Document Format (PDF)* umożliwia użytkownikom wymianę i przeglądanie dokumentów elektronicznych, niezależnie od środowiska, w którym zostały utworzone lub środowiska, w którym są przeglądane lub drukowane. Dokumenty PDF zawierają: 1) obiekty (adnotacje, hiperłącza, które nie są częścią samej zawartości strony), 2) strukturę pliku (struktura pliku określa, w jaki sposób obiekty są przechowywane w pliku), 3) strukturę dokumentu (strony, czcionki, adnotacje), 4) strumień treści i zasoby (strumień treści PDF zawiera sekwencję instrukcji opisujących wygląd strony lub innego elementu graficznego i związane z nimi zasoby). Źródło: ISO/DIS 32000 – 2:2016 *Document management – Portable document format – Part 2: PDF 2.0*

¹³ Lewandowski, T., Starczewski, M. (2014). *Po czym rozpoznać dobre repozytorium?* Biblioteka i Edukacja, 6, 1–16.

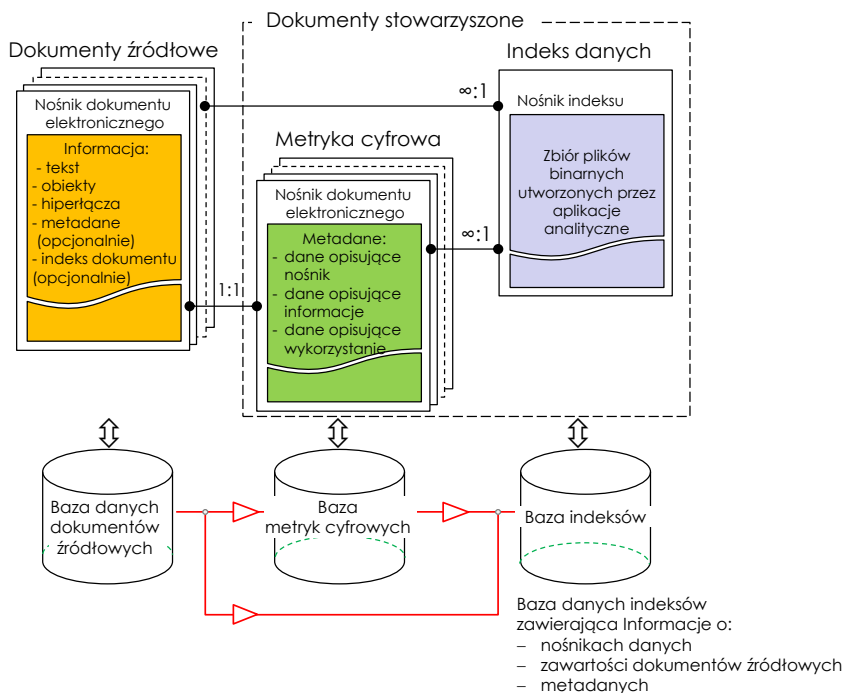
¹⁴ Nahotko, M. (2021). *Zastosowanie metadanych w cyklu życia danych badawczych*. In P. Korycińska (Ed.), *Horyzonty informacji* (Vol. 2, pp. 8–20). Uniwersytet Jagielloński, Biblioteka Jagiellońska.

Indeksy danych służą do szybkiego wyszukiwania informacji, zwłaszcza ma to znaczenia dla dużych liczebnie zbiorów danych. Różne możliwości umieszczenia metadanych i indeksów w dokumentach źródłowych i stowarzyszonych zostały zaprezentowane na rysunku 1.



Rys. 1. Nośniki dokumentów elektronicznych z zawartością danych: a) dokument źródłowy zawierający dane statyczne i dynamiczne, b) dokument źródłowy posiadający stowarzyszony dokument zawierający metrykę cyfrową, c) dokument źródłowy posiadający dokumenty stowarzyszone

Najczęściej nie tworzy się indeksu dla jednego dokumentu w relacji jeden do jednego, ale dla zbiorów dokumentów źródłowych w relacji typu wiele do jednego (relacja $\infty:1$). Zbiory dokumentów stowarzyszonych tworzą bazę danych. W celu doprecyzowania pojęć, przez *bazę danych* rozumie się bazę dokumentów bez oprogramowania umożliwiające operacje na zgromadzonych dokumentach. Będą to bazy: danych dokumentów źródłowych, metryk cyfrowych, indeksów (rys. 2).



Rys. 2. Bazy danych w badaniach bibliograficznych

BARIERA 2:

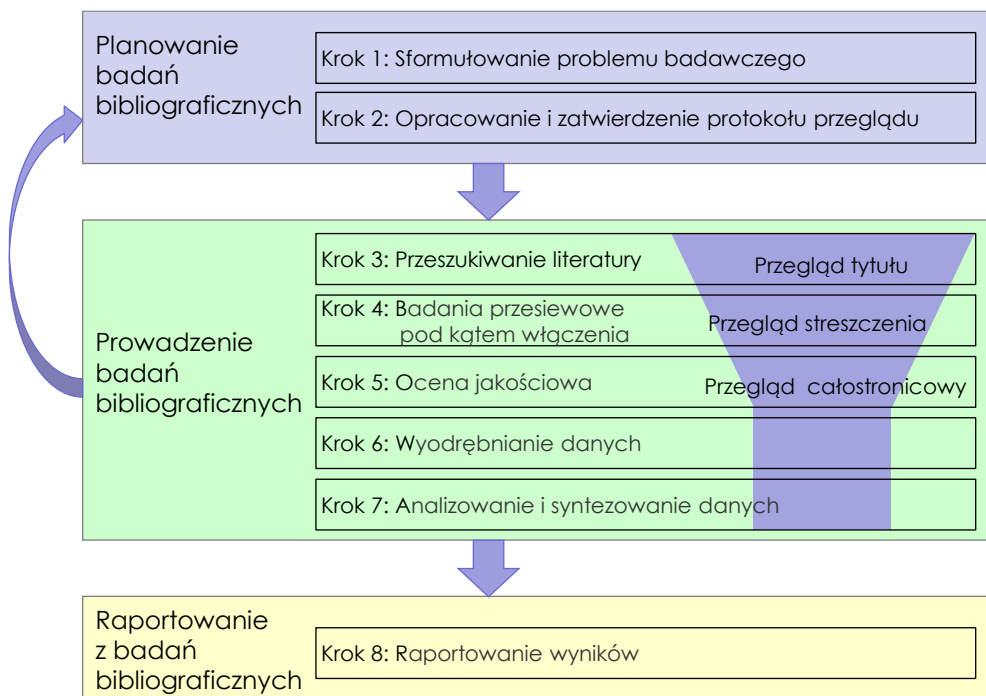
Bariera wykorzystania sprawdzonych metod wykorzystywanych w badaniach bibliograficznych przez grupy i zespoły specjalistów

W badaniach bibliograficznych wykorzystuje się różne niestandardowe (nietypowe, unikalne rozwiązanie) jak i standardowe metody^{15, 16} i techniki analiz zaimplementowane w różnego typu oprogramowaniu lub w specjalistycznych serwisach internetowych. Pomimo różnic w procedurach różnych

¹⁵ Tarrant, D., O'Steen, B., Brody, T., Hitchcock, S., Jefferies, N., & Carr, L. (2009). *Using OAI-ORE to Transform Digital Repositories into Interoperable Storage and Services Applications*. Code4Lib Journal.

¹⁶ Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), <http://www.prisma-statement.org>, data dostępu 15.01.2022 r.

metod¹⁷, technik¹⁸, najogólniej ujmując studia (przeglądy) literaturowe można przeprowadzić w trzech etapach, w ośmiu krokach: (1) sformułowanie problemu badawczego, (2) opracowanie i zatwierdzenie protokołu przeglądu (przegląd rozumiany jako formalny zapis ustaleń), (3) przeszukiwanie literatury, (4) badania przesiewowe pod kątem włączenia, (5) ocena jakościowa, (6) wyodrębnianie danych, (7) analizowanie i syntezywanie danych, (8) raportowanie wyników (rys. 3).



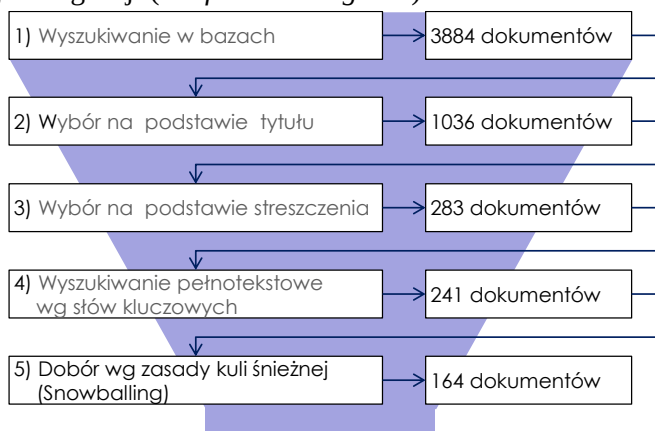
Rys. 3. Systematyczny przegląd literatury

Źródło: Xiao, Y., Watson, M. (2019). *Guidance on Conducting a Systematic Literature Review*. *Journal of Planning Education and Research*, 39(1), 93–112.

¹⁷ Wohlin, C., Mendes, E., Felizardo, K.R., Kalinowski, M. (2020). *Guidelines for the search strategy to update systematic literature reviews in software engineering*. *Inf. Softw. Technol.*, 127, 106366.

¹⁸ Zdonek, I., Hysa, B. (2017). *Analiza publikacji z obszaru nauk o zarządzaniu pod względem stosowanych metod badawczych*. *Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska*.

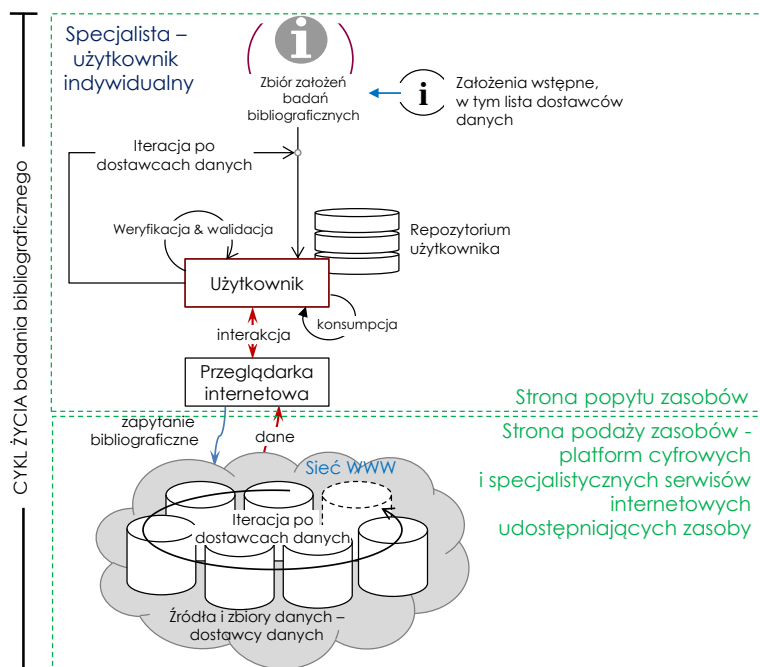
Na etapie prowadzenia badań bibliograficznych występuje selekcja (zawężenie) liczby dokumentów. Graficznie przedstawiony „lejek” wizualnie odzwierciedla ten proces. Na rysunku 4 zaprezentowano przykład sposobu selekcji dokumentów w celu wyodrębnienia dokumentów do dalszej analizy w zakresie sztucznej inteligencji (*Artificial Intelligence*).



Rys. 4. Przykład zastosowania metodyki doboru publikacji w artykule naukowym

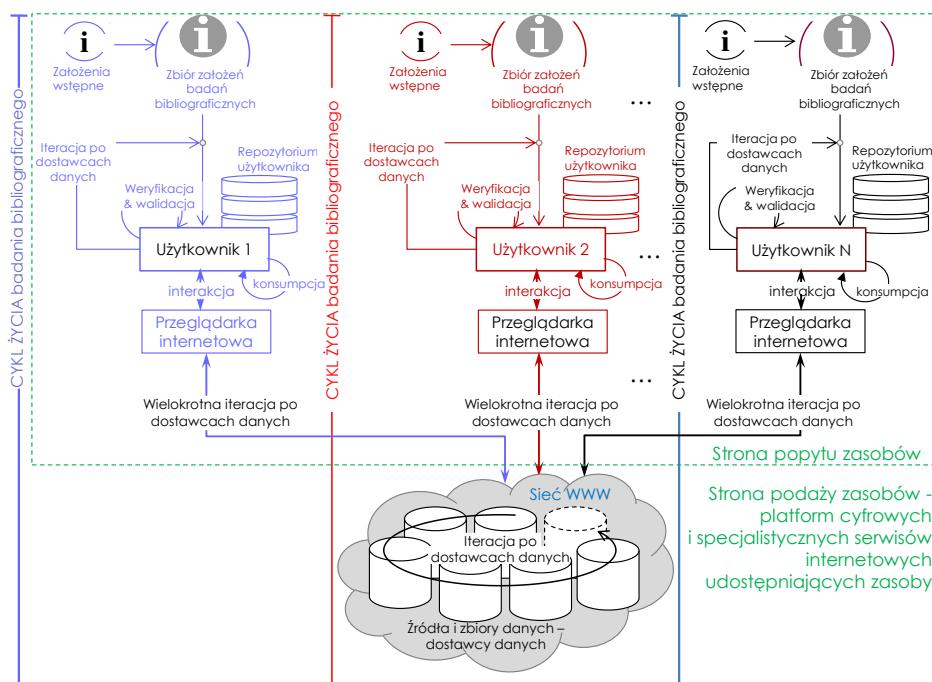
Źródło: Etapy i wartości liczbowe na podstawie Xie, Y., Cruz, L., Heck, P., Rellermeier, J. S. (2021). *Systematic Mapping Study on the Machine Learning Lifecycle*. In L. O'Conner (Ed.), *Proceedings – 2021 IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI, WAIN 2021* (pp. 70 – 73). [9474380] (*Proceedings – 2021 IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI, WAIN 2021*).

Podejście metodyczne w badaniach bibliograficznych lub podejście *ad hoc* albo działania szablonowe są działaniami praco- i czasochłonnymi wykonywanymi indywidualnie. Wysoka praco- i czasochłonność badań wynika z faktu, że istnieje wielu dostawców danych i realizowane działania są powtarzane na liście źródeł i zbiorach danych, dopóki nie zostaną przeprowadzone badania na wszystkich dostawcach przy tych samych założeniach wstępnych (rys. 5).



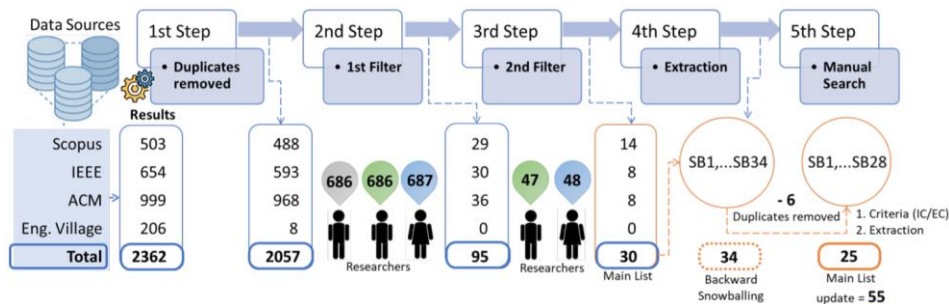
Rys. 5. Indywidualne i wielokrotne badania bibliograficzne na dostawcach danych – schemat poglądowy badań bibliograficznych

Ocena stanu wiedzy naukowej, badawczej i praktycznej ulega skomplikowaniu w grupach, zespołach monodyscyplinarnych lub interdyscyplinarnych specjalistów. Te same czynności są wielokrotnie wykonywane w ten lub bardzo analogiczny sposób. Każdy specjalista – użytkownik prowadzi badania w znany sobie sposób przy wykorzystaniu określonych metod i technik badawczych, w ten sposób wyznaczany jest jego indywidualny proces o różnej długości, wyznaczający cykl życia badań bibliograficznych. Nawet, uczestnicząc w zespole dyscyplinarnym, badania bibliograficzne wykonywane są w większości na tej samej liście źródeł i zbiorach, w tym samym lub różnym czasie. Opis indywidualnej pracy w grupie, zespole monodyscyplinarnych lub interdyscyplinarnych specjalistów został graficznie w sposób poglądowy zaprezentowany na rysunku 6.



Rys. 6. Badania bibliograficzne prowadzone indywidualnie w grupach lub zespołach specjalistów na dostawcach danych – schemat poglądowy badań bibliograficznych

Znaczenie zidentyfikowanej bariery jest znacznie mniejsze dla przedsięwzięć skutkującymi jednokrotnymi osiągnięciami, np. przy opracowaniu patentu, artykułu, monografii, gdyż badania bibliograficzne można rozdzielić pomiędzy specjalistów. Na rysunku 7 zaprezentowano przykładowy proces dekompozycji i podziału pracy w badaniach bibliograficznych dla zespołu monodyscyplinarnego, którego wynikiem była publikacja naukowa.



Rys. 7. Przykład publikacyjnego procesu selekcji w badaniach bibliograficznych

Źródło: Nascimento, E.D., Nguyen – Duc, A., Sundbø, I., & Conte, T.U. (2020). *Software engineering for artificial intelligence and machine learning software: A systematic literature review*. ArXiv, abs/2011.03751.

Natomiast w przypadku złożonych projektów badawczych w skali krajowej lub międzynarodowej zarówno w odniesieniu do zespołu monodyscyplinarnych jak też interdyscyplinarnych specjalistów z różnych organizacji zidentyfikowana bariera nabiera znaczenia.

Prowadzenie badań bibliograficznych przy tych samych lub podobnych założeniach bez znajomości sprawdzonych metod wykorzystywanych w badaniach bibliograficznych przez grupy i zespoły użytkowników uniemożliwia szybkie, wydajne, produktywnie ich przeprowadzenie. Nawet jeżeli czynności są realizowane w różnym czasie, to nadal się dublują. Takie działania są nieefektywne, wykluczające współdziałanie, współpracę i wspólne wykorzystanie zasobów. Dla wielu osiągnięć realizowanych w organizacji prowadzi to do nieefektywności zarządzania i wykorzystania zasobów ludzkich, informacyjnych, czasu, który mógłby być wykorzystany w innych procesach biznesowych, bardziej racjonalny, społeczny lub ekonomiczny sposób. Opracowanie metodyki dla grup i zespołów specjalistów w badaniach bibliograficznych w sposób znaczący umożliwiły zmniejszenie lub zniwelowanie zidentyfikowanej bariery, w ten sposób przyczyniając się do wzrostu efektywności w zarządzaniu zasobami ludzkimi w organizacjach.

BARIERA 3:

Bariera w zapewnieniu aktualności repozytoriów specjalistów przy masowości i ciągłości pozyskiwanych zasobów z rozproszonych źródeł danych

Konsekwencją rosnącej liczby źródeł, zasobów oraz przyrostu liczby i rozmiaru zasobów jest ich częstsza aktualizacja w platformach i specjalistycznych serwisach udostępniających. Implikuje to z kolei problemy po stronie użytkowników tych serwisów. Po pierwsze, następuje zwiększenie pracochłonności przeszukiwania na większym zbiorze danych u dostawców danych i ich pozyskiwania. Po drugie otrzymane wyniki kwerend mogą się znacznie różnić, gdy zostaną one przeprowadzone w odstępach kilkugodzinnych na różnych źródłach i zasobach. Zapewnienie aktualności pozyskanych zasobów i przechowywanych w repozytoriach specjalistów, co przekłada się na aktualność badań bibliograficznych jest trudne do osiągnięcia. Proces przyrostowej aktualizacji zasobów jest trudny w realizacji w sposób manualny wg schematu *wyszukanie – pozyskanie – analiza – ocena* lub *subskrypcja – pozyskanie – analiza – ocena*

albo w dostępie do danych przez API (*Application Programming Interface*): *budowanie kwerendy – pozyskanie – przetworzenie – ocena*, gdyż są to działania wieloetapowe, powtarzające się i monotonne. Dodatkowo podczas aktualizacji zasobów w repozytoriach należy sprawdzać na bieżąco czy zasoby się nie dublują lub wersjonują. W sposób ciągły nie można prowadzić studiów literaturowych dotychczasowymi metodami, gdyż ich częsta aktualizacja zakłócałaby strukturę i treść już wykonanych analiz. W tym sensie każda wykonana ocena stanu wiedzy staje się nieaktualna już w momencie zebrania danych i dokonania jej analizy, gdyż w tym czasie dodano i upubliczniono wiele zasobów. Sposobem na zmniejszenie znaczenia tej bariery może być wykorzystanie w badaniach bibliograficznych mechanizmu zarządzania ciągłością działania (*Business Continuity Management*, BCM) znanego w naukach o zarządzaniu i jakości (rozdz. 2.3). W ten sposób, można byłoby zrealizować przyrostową aktualizację zasobów w sposób dyskretny, ale zapewniając ich ciągłość. Czyli pozyskiwanie, aktualizacja zgromadzonych zasobów z rozproszonych źródeł danych następuje w ściśle zaplanowanym momencie czasowym zapewniając ciągłość pozyskiwania zasobów i ich aktualność. Realizacja techniczna BCM dla zapewnienia aktualności badań bibliograficznych przy masowości i ciągłości upublicznienia nowych osiągnięć twórczych może polegać na zastosowaniu automatyzacji w aktualizacji zasobów. Działania rutynowe realizowane w sposób manualny można przerzucić do działań i zadań, które mogą być wykonywane przez programowalne roboty cyfrowe (RPA)^{19,20}. Zalety włączenia programowanego robota cyfrowego²¹ do automatycznej aktualizacji zasobów dotyczą trzech zagadnień: 1) robot cyfrowy nie wymaga złożonego kodowania, 2) jest skalowalny i elastyczny, aby uwzględnić różnorodne przypadki użycia, 3) może być wykorzystany praktycznie na każdym etapie badań bibliograficznych, również w sposób wybiórczy i wielokrotny.

¹⁹ NICE RPA team with Steve Kaelble: *Robotic Process Automation for Dummies®*, NICE Special Edition, John Wiley & Sons, 2018, s. 6.

²⁰ Desai, P. (2020). *Robotic process automation: RPA Pre-requisite and pivotal points*: Special issue: IAISCT(SS4). 2020 *International Conference on Smart Technologies in Computing, Electrical and Electronics* (ICSTCEE), 446–451.

²¹ Termin „programowany robot cyfrowy” w niniejszej monografii będzie stosowany zamiennie z następującymi terminami: „programowym robotem cyfrowym”, „zrobotyzowaną automatyzacją procesów”, „robotyzacją procesów biznesowych”.

BARIERA 4:

Bariera wynikająca z masowości dokumentowania, upubliczniania osiągnięć twórczych oraz z masowości pozyskiwania i uzyskania zduplikowanych zasobów w repozytoriach użytkowników

Proces prowadzenia działań naukowych, badawczych i praktycznych realizowany jest przez wiele osób na całym świecie, nawet w tym samym czasie lub z przesunięciem wynikających ze stref czasowych, z różną częstotliwością i są one przeprowadzane w sposób masowy. Dodatkowo złożoność prowadzenia badań i dokumentowania ich wyników odbywa się w wielu dziedzinach naukowych, w różnej skali aktywności i w różnych stadiach czy stopniach ich zaawansowania. Jest to więc proces ciągły z momentami dyskretnymi niezbędnymi dla zapisania, udokumentowania i przesłania do platform cyfrowych udostępniających zasoby lub zaktualizowania zasobów w repozytoriach upubliczniionych. Czas pomiędzy przekazaniem zasobu a jego dostępnością w platformach cyfrowych jest ciągle skracany. Upublicznienie zasobów odbywa się minimalnym opóźnieniem. W ten sposób tworzy się podaż zasobów intelektualnych, które będą składowane i przetwarzane przez platformy cyfrowe lub specjalistyczne serwisy internetowe udostępniające zasoby. Upublicznienie zasobów następuje w różnych godzinach, przy czym aktualizacja danych może odbywać się codzienne, o różnych porach dnia, w sposób cykliczny lub nieregularny. Od strony popytowej, użytkownicy dostrzegają wysoką dynamikę przyrostu upubliczniionych zasobów, które stale wzrastają pod względem liczby zasobów jak i ich wielkości, zarówno z sektora publicznego jak i prywatnego. To powoduje, że użytkownicy platform lub specjalistycznych serwisów doświadczają nadmiarowości zasobów, co wiąże się z masowością ich pobierania z wielu rozproszonych źródeł danych i zasobów. Pozyskane i zapisane dane od wielu dostawców w lokalnych zasobach skutkuje redundancją danych a to z kolei wymusza proces przeszukania i usuwania zduplikowanych zasobów. Zaniechanie działań w usuwaniu zduplikowanych danych prowadzi do niewspółmiernego rozrostu zasobów i zmniejszenia jakości danych.

BARIERA 5:

Bariera wynikająca z przyrostu liczby źródeł i zasobów, ich rozporoszonej lokalizacji oraz wzrostu liczby i wielkości osiągnięć twórczych w zasobach udostępnianych

Konsekwencją wzrastającej liczby źródeł, zasobów oraz przyrostu liczby i rozmiaru zasobów jest ich częstsza aktualizacja w platformach udostępniających. Implikuje to z kolei problemy po stronie użytkowników.

Pierwszym aspektem tego problemu jest fakt, że wysoka dynamika przyrostu zasobów i praca na dużych pozyskanych zbiorach danych jest problemem metodycznym, organizacyjnym i technicznym. Skutkuje to zwiększeniem nakładu pracy w przeszukiwaniu, pozyskiwaniu i ocenie pozyskanych danych oraz jest znaczącym wyzwaniem intelektualnym w radzeniu sobie z wielością i złożonością wielu, różnego typu danych, określanych *big data*.

Drugim aspektem nadmiarowości różnych źródeł danych oraz zasobów zlokalizowanych w różnych rozproszonych repozytoriach danych jest ich duplikowanie. Przechowywanie tych samych danych w różnych źródłach i ich pozyskanie prowadzi do nadmiarowości pozyskanych dokumentów. Redundancja danych prowadzi do nadmiernego obciążenia nie tylko usługodawców, którzy również sprawdzają unikatowość zasobów, ale generuje problem po stronie odbiorcy w identyfikacji i radzeniu sobie ze złożonością. Użytkownicy pozyskując dane z różnych repozytoriów muszą uwzględnić usuwanie duplikatów jako dodatkowe działanie dla uzyskania wiarygodności danych i zmniejszenia czasu na odpowiedź na zapytania podczas przeszukiwania własnych zasobów.

Bariera wynikająca z rosnącej liczby źródeł i zasobów oraz wielkości tych zasobów jest znacznym utrudnieniem w badaniach bibliograficznych²². Na dynamikę przyrostu liczby platform cyfrowych i specjalistycznych serwisów internetowych użytkownicy nie mają wpływu, gdyż odbywa się poza ich kontrolą. Zaostrza się wtedy konkurencja na rynku dostawców danych. Rywalizacja pomiędzy wydawcami powoduje, że stwarzane są mechanizmy lepkości

²² Szafrński, L. (2020). *Repozytoria instytucjonalne na świecie: próba charakterystyki*. In A. Podrazik, J. Stanek-Kapcia, M. Urbaniec (Eds.), *Biblioteki uczelniane wobec środowiska akademickiego: nowe obszary działania: materiały Jubileuszowej Konferencji Biblioteki Głównej AGH z okazji obchodów 100-lecia Akademii Górniczo-Hutniczej im. Stanisława Staszica*, Kraków 12–13 września 2019 r. (s. 35–46). Wydawnictwo AGH.

serwisu, która ma przyciągać i zatrzymywać użytkowników jak najdłużej w serwisie (*sticky web site*)²³ oraz realizowane są mechanizmy próbujące uzależnić ich od firmowego serwisu przez narzędzia ułatwiające im pracę. Są to pozytywne aspekty tej tendencji, ale powstawanie kolejnych źródeł i zbiorów rozproszonych danych zwiększa poziom redundancji danych.

BARIERA 6:

Bariera wynikająca z braku unifikacji w dostępie do zasobów i ich pozyskiwania z różnych źródeł, o różnej architekturze informacji oraz modelu prawnym ich udostępniania

Brak jednego scentralizowanego zasobu w udostępnianiu danych sprawia, że należy uwzględnić różną architekturę informacji dostępnych źródeł. Korzystanie z dużej liczby rozproszonych źródeł danych, o zróżnicowanej strukturze, sposobie dostępu, w tym autoryzacji, prawach własności (modelu udostępniania) czy sposobów odpłatności powoduje zindywidualizowane podejście do każdego źródła zasobu. Brak przyjętego standardu architektury przechowywanej i udostępniania informacji powoduje, że nie jest możliwe zastosowanie jednego zunifikowanego schematu wyszukiwania informacji. Każde źródło danych (repozytorium) cechuje się własną budową syntaktyczną, semantyczną zapytań oraz różnym formatem i strukturą otrzymywania wyników. Dlatego w przypadku formułowania zapytań bibliograficznych do różnych repozytoriów danych uzyskuje się różne, nieporównywalne wyniki. Skutkuje to tym, że użytkownik jest zobligowany do utworzenia zbioru założeń badania bibliograficznego dla każdego źródła i zbioru danych oraz określenia sposobu (metodyki) weryfikacji tych danych. W odpowiedzi na zapytania bibliograficzne uzyskuje wyniki, które podlegają analizie, ocenie ilościowej i jakościowej, weryfikacji i walidacji. Problematyka oceny i procesy weryfikacji i walidacji został zaprezentowany w opisie bariery nr 7. Dostawca danych kieruje się własną polityką przeszukiwania, pobierania, gromadzenia i udostępniania zgromadzonych zasobów, sposobami autoryzacji użytkowników oraz opracowuje modele płatności za dostęp do usług lub zasobów. Dla użytkowników sposób udostępniania

²³ Judy Chuan–Chuan Lin (2007) *Online stickiness: its antecedents and effect on purchasing intention*, Behaviour & Information Technology, 26:6, 507–516.

oraz korzystanie z zasobów opiera się na określonych prawach autorskich ustalonych przez dostawców. Licencje określają możliwości ingerencji i sposób oraz warunki udostępniania również innym użytkownikom. Działania podejmowane w stosunku do wykorzystania zasobów muszą respektować zarówno prawa autorskie jak i handlowe do utworu. Dla użytkowników jest to bariera w pozyskiwaniu zasobów, zapoznawania się z zasobami, ponieważ wymagana jest rejestracja w serwisach, co w prosty sposób uniemożliwia zastosowanie automatyzacji. Weryfikacja tożsamości wiąże się sposobami logowania się do serwisów, procesu weryfikacji użytkownika i w niektórych przypadkach przezwycięzenia techniki Captcha (*Completely Automated Public Turing Test to Tell Computers and Humans Apart*). Techniki Captcha utrudniają a nawet uniemożliwiają pozytywną autoryzację osobom z niepełnosprawnościami, gdyż Captcha może uniemożliwić korzystania z serwisu i zarazem z danych. Zastosowanie mechanizmów autoryzacji dla potwierdzenia subskrypcji danych to kolejny wygenerowany problem w dostępie do zasobów i ich pozyskiwania.

BARIERA 7:

Bariera w zapewnieniu jakości udostępnianych danych przez dostawców, oceny ilościowej i jakościowej pozyskanych zasobów oraz procesu przeprowadzenia weryfikacji walidacji w repozytoriach użytkowników

Kolejną barierą w badaniach bibliograficznych jest ocena ilościowa i jakościowa wyników zapytań dokonanych przez użytkowników w serwisach internetowych dostawców danych. Ocena ilościowa jest wynikiem działań algorytmów przeszukiwań w bazach danych, mechanizmów indeksujących zaimplementowanych w systemach wyszukiwawczych dostawców danych oraz zależna od wielkości zbiorów danych. Dlatego użytkownicy uzyskują różne wyniki zapytań bibliograficznych u różnych dostawców danych. Sytuacja jest bardziej złożona przy ocenie jakościowej. Jakość danych dotyczy dwóch stron rynku, jedna strona związane jest ze źródłami i zbiorami danych, czyli po stronie dostawcy danych, druga strona dotyczy użytkowników wykonujących badania.

W zakresie strony związanej ze źródłami i zbiorami danych, profesjonalne serwisy udostępniające dane/metadane przechowują dla każdego zasobu trzy elementy danych²⁴:

1. unikatowy identyfikator zasobu UID (*unique item identifier*),
2. dane statyczne, które są metadanymi i są w założeniu niezmiennie²⁵, dotyczą treści dokumentów źródłowych oraz elementów bibliograficznych, np. autora/autorów, tytułu, streszczenia, słów kluczowych, w przypadku patentów są to dane wynalazcy i właściciela, data zgłoszenia i udzielenia, numery klasyfikacji patentowej itp.,
3. dane dynamiczne, które są metadanymi przyporządkowane do zasobu, zarówno nośnika danych jak i treści w dokumencie, np. liczba odczytu, pobrań, liczba publikacji autora w bazie oraz dane generowane przy przetwarzaniu i integracji bazy danych.

Metadane statyczne i dynamiczne oraz zawartość dokumentów są wykorzystywane w ocenie jakościowej danych po stronie dostawcy i użytkownika. Po stronie odbiorcy, użytkownik danych powinien mieć dostęp zarówno do metadanych jak i dokumentów źródłowych^{26,27}. Liczba serwisów udostępniających dokumenty źródłowe jest mniejsza od liczby serwisów udostępniających metadane, gdyż kluczowym zagadnieniem jest model prawny udostępniania zasobów, np. prawa autorskie. Po stronie użytkownika, jakość danych można ocenić po zakończeniu badań bibliograficznych oraz wykonaniu kompletnej bibliografii dla danego nowego osiągnięcia twórczego. Sporządzenie listy pozycji bibliograficznych, pomijając aspekt techniczny jej wykonania, ma istotny wpływ na ocenę jakościową wykonanej bibliografii w stosunku do powstałego utworu. Najczęściej ocena dotyczy liczby pozycji cytowanych autorów w bibliografii oraz zbioru różnych subiektywnych mierników lub wskaźników, np. sposób doboru listy źródeł i zasobów, ważność pozycji bibliograficznej w stosunku do zawartości, aktualność, adekwatność, prestiż autora/autorów lub czasopisma. Kwestia jakości odniesień do źródeł i sporządzenia bibliografii

²⁴ *WEB of Knowledge XML User Guide*. Thomson Reuters, May 9, 2013.

²⁵ W szczególnych przypadkach dostawcy danych dysponującymi prawami do zarządzania zasobem posiadają procedury zgłaszania błędów przez użytkowników do zmiany lub aktualizacji danych.

²⁶ Knoth, P., Anastasiou, L., & Pearce, S. (2014). *My repository is being aggregated: a blessing or a curse?* In: Open Repositories 2014 (OR2014), 9–13 Jun 2014, Helsinki, Finland.

²⁷ Crow, R. (2002). *The case for institutional repositories: a SPARC position paper*. ARL Bimonthly Report 223.

ma poboczne znaczenie ze względu na tematykę badawczą monografii, dlatego zagadnienie jakości będzie zasygnalizowane dwoma najważniejszymi aspektami. Pierwszy aspekt to jakość zasobu oceniana przez wydawcę, rozumiana jako stopień, do którego pozyskany zasób odpowiada jego formalnej specyfikacji (technicznej, edytorskiej, merytorycznej). Drugi aspekt dotyczy jakości zasobu w bibliografii ocenianej przez recenzenta i rozumianej jako stopień, w jakim zbiór inherentnych właściwości pozycji bibliograficznych spełnia założone wymagania i jest w zestawieniu w bibliografii.

Ostatnim zagadnieniem jest zasygnalizowanie procesu weryfikacji i walidacji badań bibliograficznych. Procesy weryfikacji i walidacji są częścią procesu zarządzania jakością w badaniach i również składową jakości w zarządzaniu wiedzą w organizacji. Procesy weryfikacji i walidacji (W&W) służą do ustalenia, czy wykorzystane aplikacje oraz otrzymane wyniki dla danego działania są zgodne z wymogami tego działania i czy przetworzone dane spełniają oczekiwania i potrzeby użytkownika w cyklu życia badania bibliograficznego. Procesy W&W powinny zapewnić obiektywną ocenę aplikacji, danych i procesów. Ogólnie, w metodyce badania bibliograficznego weryfikacja i walidacja to procesy polegające na zweryfikowaniu otrzymanych danych oraz rozstrzygnięciu, czy informacje pozyskane na podstawie danych są adekwatne, aktualne, poprawne, kompletne, spójne, spełniają założenia oraz są zgodne z potrzebami i oczekiwaniami użytkownika. W tym procesie może być zastosowany klasyczny cykl doskonalenia W. E. Deminga “planuj – wykonaj – sprawdź – działaj”²⁸ (*Plan – Do – Check – Act*, PDCA) lub podejście zwinne.

Ocena ilościowa i jakościowa pozyskanych zasobów oraz proces przeprowadzenia weryfikacji i walidacji w repozytoriach użytkowników jest zagadnieniem, sprawiającym problem użytkownikom, ponieważ ocena tych zagadnień jest subiektywna. Dla zmniejszenia lub zniwelowania znaczenia zidentyfikowanej bariery należy zobiektywizować proces oceny jakościowej przez ich zautomatyzowanie. W niniejszej monografii proces W&W będzie dotyczył wszystkich elementów opracowanej metodyki zautomatyzowanego prowadzenia badań bibliograficznych na etapie badań bibliograficznych.

* * *

²⁸ *Wspólna Metoda Oceny. Doskonalenie organizacji poprzez samoocenę*. Kancelaria Prezesa Rady Ministrów Warszawa, 2008, s. 9.

Tematyka monografii dotyczy automatyzacji prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej, praktycznej dla grup i zespołów specjalistów, wspomaganą programowalnymi robotami cyfrowymi. Programowalne roboty cyfrowe to rozwiązania m.in. oparte na sztucznej inteligencji w zakresie wykorzystania technologii zorientowanych na działania (RPA) wspierające decyzje użytkownika (AI). Przyczynkiem do podjęcia tematu było zidentyfikowanie i analiza niewyczerpujących siedmiu barier w prowadzeniu badań bibliograficznych. Bariery utrudniają użytkownikom bezpieczne, satysfakcjonujące korzystanie z platform cyfrowych oraz specjalistycznych serwisów w prowadzeniu badań bibliograficznych i tym samym uniemożliwiają skuteczne wychwycenie luk naukowych, badawczych, praktycznych. **Zwłaszcza, że zdolność percepcji, analiza zawartości, selekcji informacji każdej osoby posiada granicę, która nie może być przekroczona przy zastosowaniu znanych sobie technik, narzędzi informatycznych i organizacji pracy.** Istnieje pewien moment graniczny, przy którym należy już podjąć działania optymalizacyjne i racjonalizacyjne oraz skoncentrować się na prowadzeniu badań bibliograficznych. Zakres zmian powinien dotyczyć innowacji procesowej, produktowej, organizacyjnej z wykorzystaniem cyfryzacji umożliwiających automatyzację działań i wspomagających użytkownika w procesie prowadzenia badań bibliograficznych, włączając również analizę ilościową i jakościową. Spośród wielu rozwiązań technicznych, narzędzi wykorzystywanych w nauce, rozwiązaniach praktycznych w zakresie komputerowych metod automatyzacji²⁹ do powtarzalnych, rutynowych procesów stosowane są programowalne roboty cyfrowe. Zrobotyzowanie procesów na etapie badań bibliograficznych w ulepszaniu skuteczności wykorzystania e-usług dla grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów jest przedmiotem niniejszej monografii.

Konkludując, **w praktyce naukowej, badawczej i gospodarczej oraz w piśmiennictwie wymienione bariery są istotnymi problemami dla specjalistów z różnych sektorów gospodarki, które nie zostały w sposób profesjonalny rozwiązane.** Oczekiwania interesariuszy, tj. osób, grup, zespołów, którzy realizują nowe osiągnięcia twórcze i dokonują oceny stanu wiedzy naukowej, badawczej, praktycznej na podstawie badań bibliograficznych

²⁹ Yatskiv, S., Voytyuk, I., Yatskiv, N., Kushnir, O., Trufanova, Y., & Panasyuk, V. (2019). *Improved Method of Software Automation Testing Based on the Robotic Process Automation Technology*. 2019 9th International Conference on Advanced Computer Information Technologies (ACIT), 293–296. DOI: 10.1109/ACITT.2019.8780038.

odnoszą się do skutecznych, efektywnych i użytecznych rozwiązań spełniających ich potrzeby informacyjne, oczekiwania, wymagania. Ze względu, na złożoność, pracochłonność i monotoność badań bibliograficznych a także wielość wyników z licznych źródeł i zbiorów danych rozwiązania o różnym charakterze powinny uwzględniać automatyzację procesów powtarzalnych z wykorzystaniem programowalnych robotów cyfrowych. W tym obszarze istnieje więc identyfikowalna luka naukowa, badawcza i techniczna o charakterze zarówno studialnym, koncepcyjnym jak i aplikacyjnym, która powinna zostać uzupełniona, a zaprezentowane wnioski oraz rekomendacje mogłyby stanowić unikatowy wkład w rozwój metodyki, przede wszystkim w nauce o zarządzaniu i jakości.

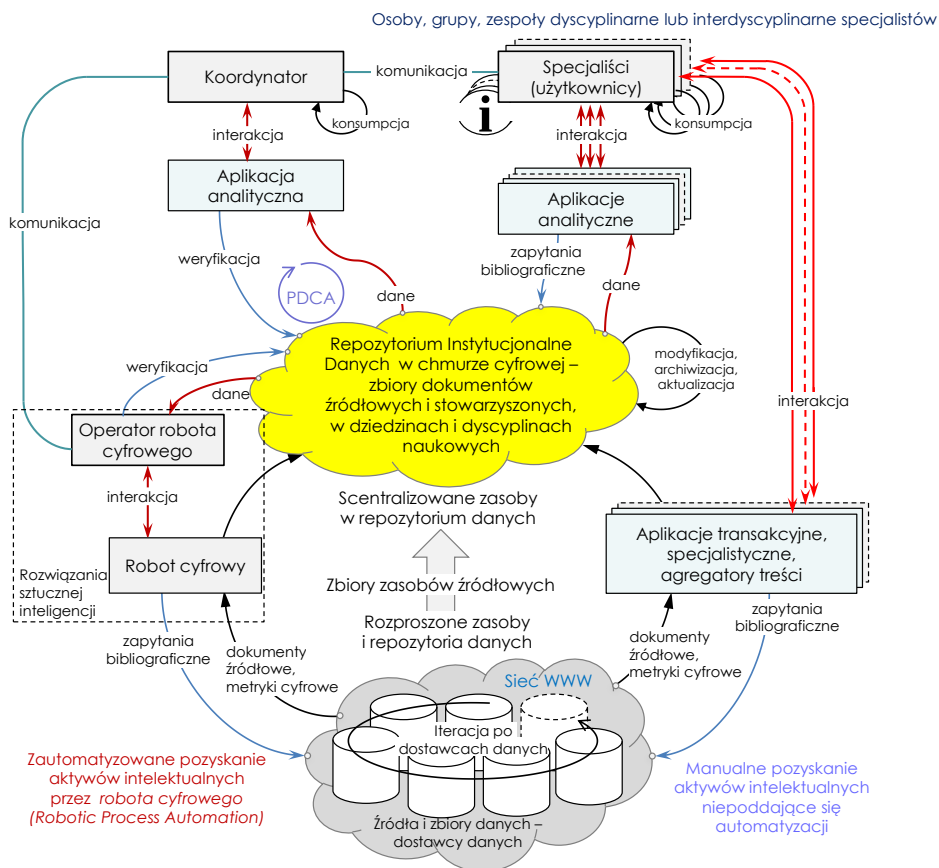
Stwierdzenie przez specjalistów określonego poziomu deficytu wiedzy, jej względnego nieuporządkowania i luki osiągnięć twórczych staje się tym samym przyczynkiem do podjęcia przez nich działań twórczych. Działania twórcze dotyczą koncepcji nowego, poszerzonego lub pogłębianego twórczego osiągnięcia naukowego, badawczego, praktycznego, która jest odpowiedzią mającą wypełnić w sposób całościowy lub częściowy stwierdzoną lukę osiągnięć twórczych. Ze względu na fakt masowości działań twórczych, w których każde nowe osiągnięcie twórcze musi być potwierdzone wynikami z badań bibliograficznych w pełni uzasadnione wydaje się niezwłoczne podjęcie koncepcyjnych ukierunkowanych na wzrost sprawności działań i skuteczności grup i zespołów specjalistów, wykorzystania dostępnych rozwiązań informacyjno-organizacyjnych i programistycznych. Właściwe użyte zasoby ludzkie, finansowe, rzeczowe, informacyjne oraz należyty dobór metod, narzędzi technik mogłyby się przyczynić do zauważalnego skrócenia cyklu życia badania bibliograficznego i stworzenia specjalistom komfortu psychicznego oraz komfortu pracy pobudzającego innowacyjność opartą na ocenie ilościowej i jakościowej zgromadzonych zasobów.

Tematyka badawcza obejmuje aspekty metodyczne, przedsięwzięcia projektowe, aplikacyjne, zarządzania zasobami ludzkimi związane z automatyzacją prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej i praktycznej dla grup i zespołów specjalistów, wspomaganą programowanymi robotami cyfrowymi. Jednym z podejść do prowadzenia badań bibliograficznych jest archiwalne przechowywanie zbiorów danych źródłowych pozyskanych z badań ilościowych oraz analiza zebranych danych wtórnych przez użytkowników. Jest to jedna z wielu koncepcji przetwarzania dużej ilości

danych^{30,31}, praktycznej realizacji badań bibliograficznych, która jest przedmiotem zainteresowań autora. Zbiór zasobów danych wtórnych jest przechowywany w repozytorium danych. Repozytorium powstaje przez wyszukanie i pozyskanie dokumentów źródłowych oraz metadanych opisujących zawartość dokumentu z wielu rozproszonych repozytoriów danych w trybie dostępu subskrypcji i otwartego dostępu przez programowalnego robota cyfrowego. W przypadku niektórych danych niemożliwych do pozyskania automatycznie, wyszukanie i zapisanie następuje manualnie przez użytkownika lub koordynatora. W niniejszej monografii przedstawiono koncepcję, która była weryfikowana empirycznie a dotyczy możliwości automatyzacji pozyskania zasobów (dokumentów źródłowych i stowarzyszonych dokumentów) z rozproszonych źródeł danych a następnie modyfikacji tych zbiorów i umieszczenie ich w jeden scentralizowany zbiór danych udostępnianych specjalistom. Weryfikowana koncepcja metodyki zautomatyzowanego prowadzenia badań bibliograficznych w postaci graficznej została zaprezentowana na rysunku 8.

³⁰ Davidson, E., Edwards, R., Jamieson, L., & Weller, S. (2019). *Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data*. *Quality & Quantity*, 53(1), 363–376.

³¹ Neale, B., Bishop, L. (2012). *The Timescapes Archive: a stakeholder approach to archiving qualitative longitudinal data*. *Qualitative Research*, 12, 53–65.



Rys. 8. Zautomatyzowane i manualne pozyskanie, aktualizacja zasobów oraz analiza danych wtórnych przez użytkowników z wykorzystaniem RID na etapie badań bibliograficznych

Pytania badawcze

W warstwie opisowej monografii w zaprezentowanych relacjach pomiędzy użytkownikami pełniącymi zdefiniowane role i rozwiązaniami projektowymi przyporządkowano weryfikowalne pytania badawcze odnoszące się do procesów w cyklu życia Repozytorium Instytucjonalnego Danych dla osób, grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów:

1. Jakie procesy mogą być zautomatyzowane i realizowane przez programowalne roboty cyfrowe jako rozwiązania sztucznej inteligencji na etapie prowadzenia badań bibliograficznych przez osoby, grupy, zespoły specjalistów w ocenie stanu wiedzy naukowej, badawczej i praktycznej?

2. Jak można ocenić, uzasadnić, zweryfikować przydatność programowalnych robotów cyfrowych w koncepcji nowych, poszerzonych lub pogłębionych twórczych osiągnięć naukowych, badawczych, praktycznych?
3. W jaki sposób bezpośredni i pośredni programowalne roboty cyfrowe oddziałują na interesariuszy wewnętrznych i zewnętrznych?

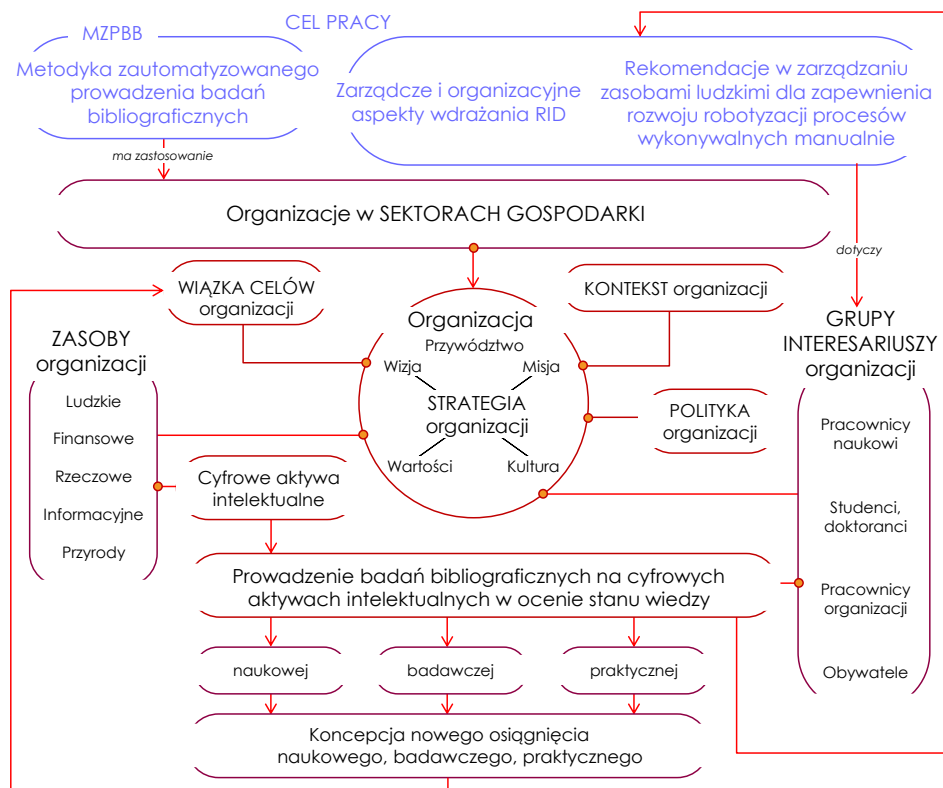
W warstwie aplikacyjnej weryfikacji podlegają zastosowane rozwiązania, które według autora są innowacyjne i mogą zostać sklasyfikowane do określonych typów innowacji: procesowej, produktowej, organizacyjnej. Wykazanie innowacyjności rozwiązań nastąpi opisowo w poszczególnych podrozdziałach monografii.

Mając na uwadze tematykę pracy, problematykę oraz powiązanie rozwiązań techniczno-organizacyjnych, programistycznych opracowany został cel, podmiot i przedmiot badań oraz zakres pracy wyznaczający granicę poznawczą.

Cel pracy

Celem pracy było opracowanie i udokumentowanie metodyki zautomatyzowanego prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej i praktycznej dla osób, grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów. **Instrumentalnym celem pracy** było wypracowanie zarządczych i organizacyjnych aspektów wdrażania RID oraz opracowanie rekomendacji w zarządzaniu zasobami ludzkimi dla zapewnienia rozwoju robotyzacji procesów wykonywalnych manualnie w organizacji.

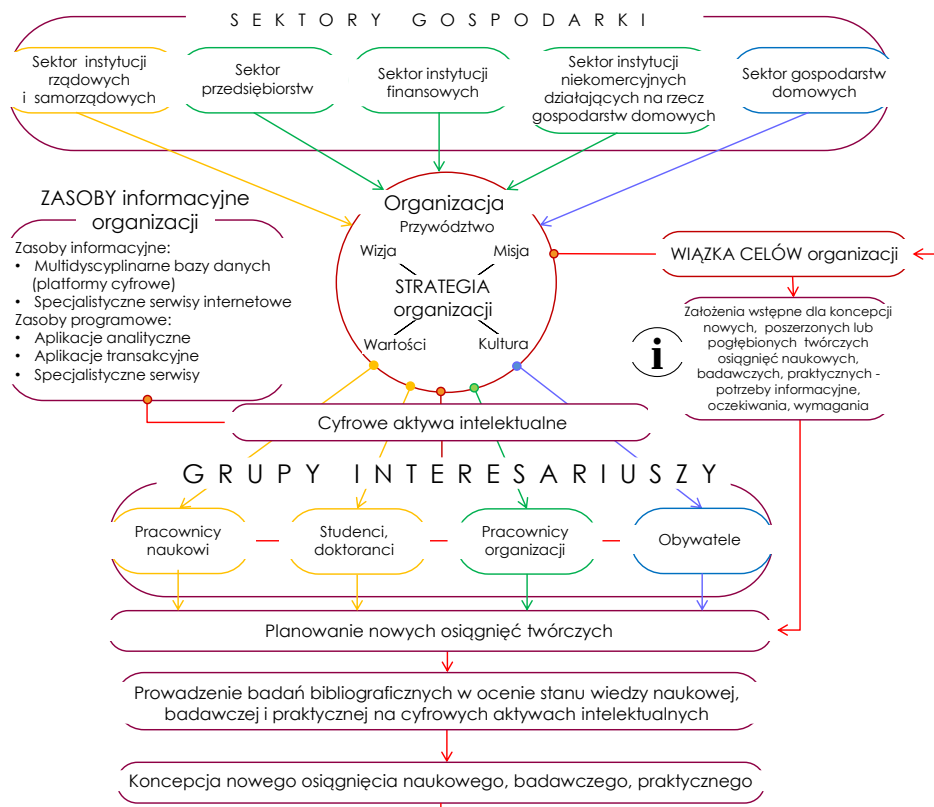
Każda organizacja wytycza kierunki działania zawarte w strategii organizacji poprzez sformułowanie jej wizji, misji, wartości, kultury organizacyjnej w ramach określonej polityki. W ramach zamierzeń strategicznych określanie i rozwijanie wizji organizacji, misji z określonymi wiązkami celów dokonywane jest przez interesariuszy. W kontekście wiązki celów organizacji, całokształt działań związanych z interesariuszami organizacji, którzy prowadzą badania bibliograficzne na cyfrowych aktywach intelektualnych w ocenie stanu wiedzy naukowej, badawczej, praktycznej w realizacji koncepcji nowego osiągnięcia naukowego, badawczego, praktycznego jest systemem podmiotowym i przedmiotowym monografii. W tym zakresie, cel pracy z całokształtem zagadnień, kontekstem wewnętrznym i zewnętrznym został zaprezentowany na rysunku 9.



Rys. 9. Cele pracy w kontekście prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej, praktycznej

Podmiot badań

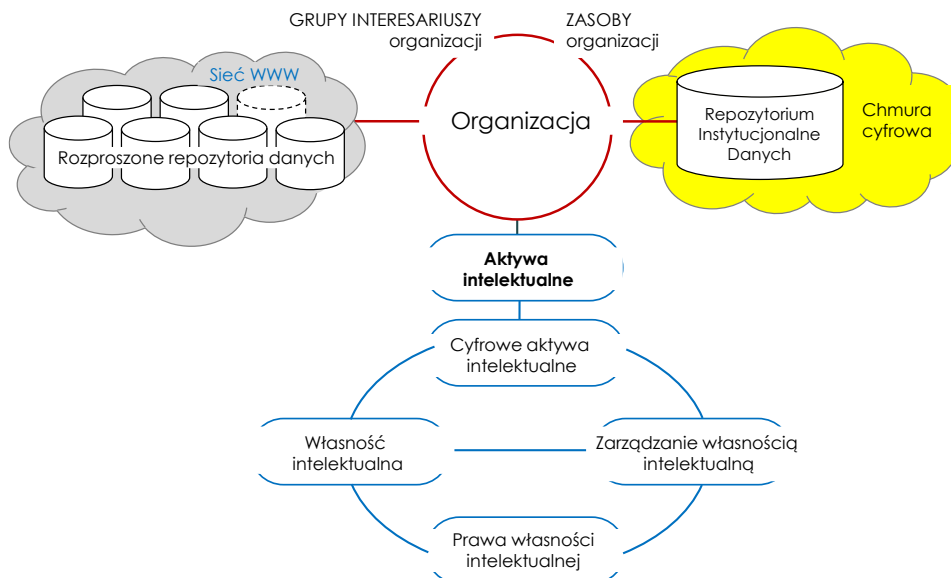
Podmiot badań dotyczy grup interesariuszy organizacji, którymi są osoby, grupy, zespoły dyscyplinarne lub interdyscyplinarne specjalistów pracujące w różnego typu organizacjach w różnych sektorach gospodarki prowadzące badania bibliograficzne na cyfrowych aktywach intelektualnych i zainteresowane oceną stanu wiedzy naukowej, badawczej i praktycznej. Abstrahując od rozwiązań szczególnych a odnosząc się do zagadnień prowadzenia badań bibliograficznych w skali makro, analiza podmiotowa specjalistów umożliwiła zidentyfikowanie czterech grup interesariuszy w różnych organizacjach, w pięciu sektorach instytucjonalnych gospodarki (rys. 10).



Rys. 10. Uszczegółowiony kontekst wewnętrzny i zewnętrzny prowadzenia badań bibliograficznych w skali makro

Zidentyfikowane grupy interesariuszy prowadzą badania bibliograficzne na cyfrowych aktywach intelektualnych. Aktywa intelektualne to niematerialne zasoby będące w dyspozycji organizacji, które stanowią dla niej wartość³². W relacji pojęć z aktywami intelektualnymi, które zostały zmaterializowane do postaci cyfrowej związane są pojęcia własności intelektualnej, prawa własności intelektualnej, zarządzanie własnością intelektualną. Wymienione terminy zostały uwzględnione w pracach nad budową, uzupełnieniem, aktualizacją i wykorzystaniem zgromadzonych zasobów cyfrowych w repozytorium instytucjonalnym danych. Graficznie przedstawienie relacji asocjacyjnej pomiędzy terminami zostały zaprezentowane na rysunku 11.

³² ISO 56000:2020 *Innovation management–Fundamentals and vocabulary*, s. 8.



Rys. 11. Relacje asocjacyjne pomiędzy terminami związanymi z cyfrowymi aktywami intelektualnymi w organizacji

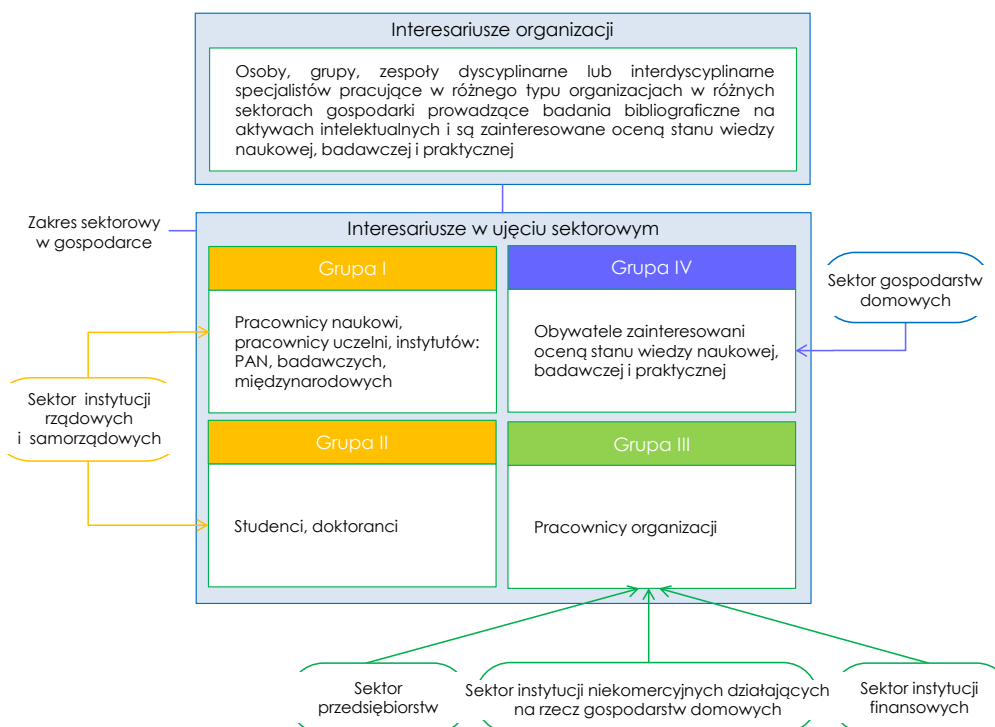
W przedmiotowej monografii aktywa intelektualne stanowią zasoby wiedzy skodyfikowanej do postaci cyfrowej będące w dyspozycji organizacji. Są to zasoby własne organizacji oraz zasoby cyfrowe zewnętrzne do których organizacja ma dostęp na określonych prawach autorskich. Zasoby zewnętrzne to rozproszone w sieci WWW repozytoria danych umiejscowione w multidyscyplinarnych bazach danych (platformach cyfrowych) i w specjalistycznych serwisach internetowych. Cyfrowe aktywa intelektualne dotyczą trzech kategorii osiągnięć twórczych: naukowe, badawcze i praktyczne. Uzupełnieniem aktywów intelektualnych są dane, które je opisują – metadane. Metadane są również zasobami organizacji będące w dyspozycji organizacji przechowywane w zasobach własnych lub w specjalistycznych serwisach, niekoniecznie w tych samych źródłach danych co aktywa intelektualne. Przy czym według zapisów w serwisie PlumX osiągnięcia twórcze określone zostały artefaktami jako „wszelkie wyniki badań dostępne online”. Artefakty zostały podzielone na 67 różnego typu, do których przypisano pięć kategorii metryk cyfrowych: cytaty, użycie, ujęcie, wzmianki, media społecznościowe³³. W niniejszej monografii, przechwycenie różnego rodzaju metadanych cyfrowych i zarchiwizowa-

³³ PlumX Metrics, <https://plumanalytics.com/learn/about-metrics>, data dostępu 12.10.2020 r.

nie ich w dokumentach źródłowych (wewnątrz dokumentu) lub dokumentach zewnętrznych są określane metrykami cyfrowymi.

Przegląd i analiza cyfrowych aktywów intelektualnych jest niezbędna dla tworzenia nowego osiągnięcia naukowego, badawczego, praktycznego w ramach działań strategicznych organizacji a zainteresowane tym są cztery grupy interesariuszy (rys. 12):

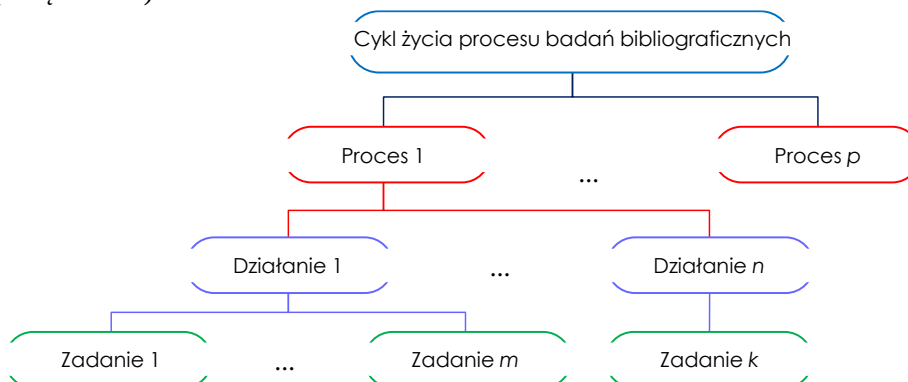
1. Pracownicy naukowci: pracownicy uczelni, instytutów PAN, instytutów badawczych, instytutów międzynarodowych. Są to grupy klientów prowadzących działalność naukową, badawczą, komercjalizację wyników działalności naukowej oraz *know – how* związanych z tymi wynikami.
2. Studenci, doktoranci różnych kierunków i specjalności.
3. Pracownicy organizacji, np. redaktorzy, dziennikarze, pracownicy B+R.
4. Obywatele, czyli osoby fizyczne, które zainteresowani są oceną stanu wiedzy naukowej, badawczej i praktycznej w celach osobistych, hobby-stycznych lub domowych na własny użytek lub w celach komercyjnych.



Rys. 12. Grupy interesariuszy z różnych sektorów gospodarki zainteresowane cyfrowymi aktywami intelektualnymi

Przedmiot badań

Przedmiotem badań w celu opracowania metodyki zautomatyzowanego prowadzenia badań bibliograficznych był zestaw procesów, działań i zadań realizowanych w cyklu życia procesu badań bibliograficznych (rys. 13) na podstawie uogólnionej metodyki prowadzenia badania bibliograficznego, jako procesu powtarzalnego, wielokrotnego dla wielu osiągnięć twórczych (Załącznik B).



Rys. 13. Przedmiot badań w ujęciu hierarchicznym w podziale cyklu życia na procesy, działania i zadania

Zakres badań przedmiotowych i podmiotowych

Zakres badań w opracowaniu metodyki zautomatyzowanego prowadzenia badań bibliograficznych obejmował analizę procesów, działań, zadań dla utworzenia repozytorium instytucjonalnego danych, ze szczególnym uwzględnieniem procesów, które mogą być zautomatyzowane przy wykorzystaniu programowego robota cyfrowego. W szczególności, zakres pracy obejmował całokształt problematyki metodyki zautomatyzowanego prowadzenia badań bibliograficznych (MZPBB) przedstawionej w różnym stopniu szczegółowości dla wymienionych działań:

1. Opracowanie cyklu życia badania bibliograficznego z wyszczególnionymi procesami, działaniami i zadaniami z przypisanymi rolami użytkowników. Wyznaczanie działań i zadań wykonywanych manualnie na etapie badania bibliograficznego było podstawą dla działań merytorycznych robotyzacji procesów. Wymieniony zakres prac dotyczył opracowania MZPBB z wyznaczeniem procesów do automatyzacji RPA.

2. Sporządzenie wymagań funkcjonalnych i нефункциональных w postaci jednolitego dokumentu elektronicznego. Pozyskanie wymagań funkcjonalnych i нефункциональных od specjalistów zainteresowanych zautomatyzowaniem działań był podstawą do dalszych działań merytorycznych w robotyzacji procesów (działanie niezaprezentowane w niniejszej monografii).
3. Opracowanie cyklu życia programowalnego robota cyfrowego RID na etapie badania bibliograficznego. W szczególności, opracowanie graficznych algorytmów w formie schematów blokowych umożliwiających zrozumienie złożoności problematyki oraz ułatwiających utworzenie graficznych algorytmów do zaprogramowania robota cyfrowego. Kolejnym działaniem było zaprogramowanie robota cyfrowego, testowanie, uruchomienie, weryfikacja i walidacja jego działania.
4. Opracowanie regulaminu dla użytkowników w korzystaniu zasobów RID (działanie niezaprezentowane w niniejszej monografii).
5. Wykonanie materiałów pomocy podnoszących lub rozszerzających kompetencje, tj. transmisje szkoleniowe on-line, filmy z rozwiązaniami najczęstszych problemów, e-kursy, dokumenty elektroniczne zawierające najczęściej zadawane pytania (FAQ) z kanałami dla świadczenia wsparcia synchronicznego i asynchronicznego użytkownikom (działanie niezaprezentowane w niniejszej monografii).
6. Opracowanie modelu zarządzania w ujęciu procesowym oraz analiza wykorzystania programowalnego robota w automatyzacji czynności na etapie badań bibliograficznych. Wymienione działania umożliwiły wypracowanie rekomendacji w zarządzaniu zasobami ludzkimi dla zapewnienia rozwoju robotyzacji procesów dotyczących prac manualnych.

Realizując cel naukowy, który został poprzedzony zdefiniowaną wiązką problemów w postaci barier oraz pytaniami badawczymi związanymi z automatyzacją prowadzenia badań bibliograficznych z wykorzystaniem programowalnego robota cyfrowego przyjęto strukturę trzech numerowanych rozdziałów, poprzedzonych obszernym wprowadzeniem, podsumowany rozdziałem nienumerowanym – zakończeniem. Całość uzupełniona została załącznikami, bibliografią, spisem rysunków.

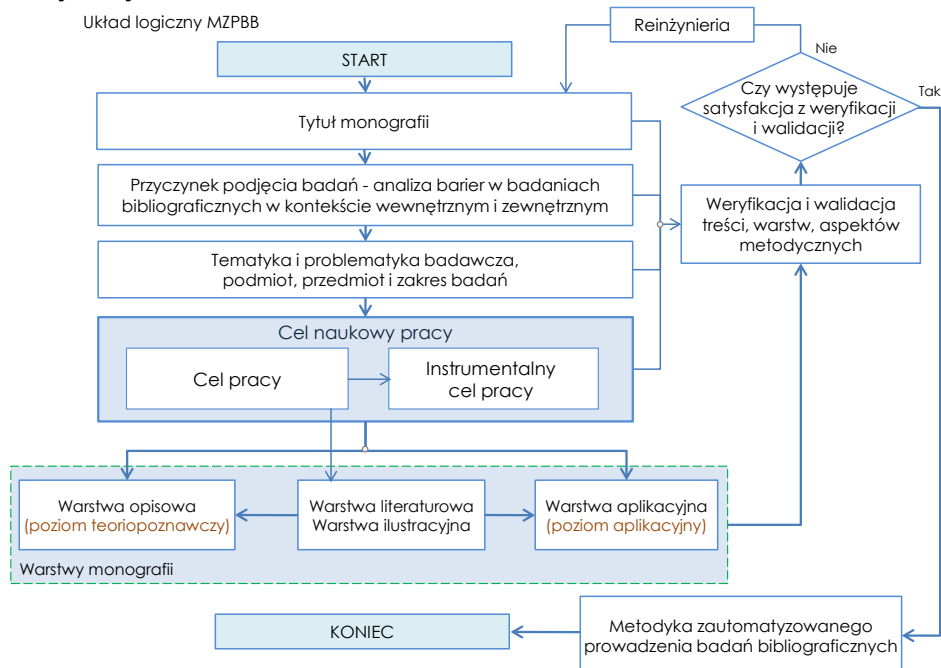
Całokształt podjętej tematyki dotyczył selektywnego zaprezentowania zakresu wykorzystania programowalnego robota cyfrowego na etapie badań bibliograficznych w następujący sposób:

- W rozdziale pierwszym przedstawiono autorską metodykę zautomatyzowanego prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej i praktycznej dla osób, grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów w cyklu życia RID. Opracowana metodyka dla określonego osiągnięcia twórczego zawiera zbiór powiązanych siedmiu zależnych procesów wyznaczając tym samym cykl życia badania bibliograficznego. Na poziomie teoriopoznawczym w warstwie opisowej tematyki i problematyki badawczej poszczególne podrozdziały prezentują kolejne procesy cyklu życia z algorytmami graficznymi, z podziałem na role użytkowników, również z niezbędnymi procedurami, które ustalają sposób przeprowadzenia procesu. Wykonana metodyka obejmowała powtarzalne procesy w cyklach życia badań bibliograficznych dla kolejnych, nowych osiągnięć twórczych zakreślając cykl życia badań bibliograficznych w cyklu życia RID.
- Na poziomie aplikacyjnym, z zastosowaniem treści z poziomu teoriopoznawczego przedstawiono w rozdziale drugim wybrane istotne aspekty zarządcze i organizacyjne procesu wdrażania Repozytorium Instytucjonalnego Danych. Proces prowadzenia badań bibliograficznych z wykorzystaniem RID został zagregowany do czterech zagadnień zaprezentowanych w poszczególnych podrozdziałach i dotyczy: określenia cyklu życia, zarządzanie zasobami, zarządzanie ciągłością działania, zarządzanie jakością produktów cyfrowych, danych i e-usług.
- W rozdziale trzecim sformułowano normatywne zalecenia dla trzech grup interesariuszy w postaci zbioru rekomendacji. Rekomendowane działania w zarządzaniu zasobami ludzkimi dla zapewnienia rozwoju e-usług RID mają odzwierciedlenie w trzech podrozdziałach i dotyczą trzech ról: użytkowników, koordynatora, operatora robota cyfrowego. Metodycznie, rozdział wpisuje się w warstwę aplikacyjną monografii.
- W zakończeniu pracy sformułowano najważniejsze wnioski wynikające z przeprowadzonych prac na poziomie teoriopoznawczym, jak i aplikacyjnym. Wyszczególniono dziesięć najważniejszych korzyści, które stanowią wartość dodaną i pożądaną przez użytkowników. Dokonano również oceny stopnia realizacji celu pracy oraz określono główne efekty na poziomie

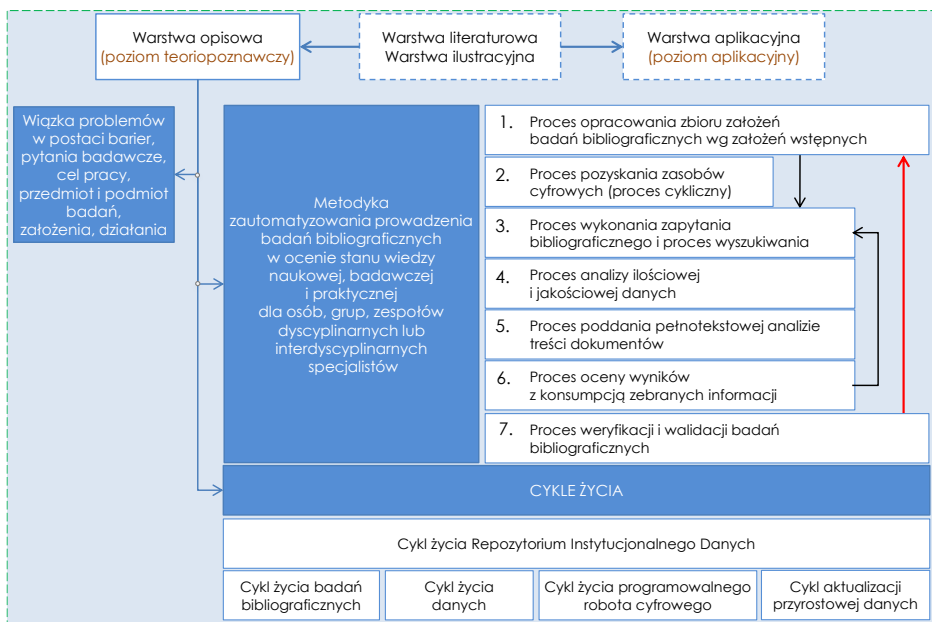
teoriopoznawczym, aplikacyjnym z wkładem pracy w rozwój dyscypliny Nauki o zarządzaniu i jakości.

Zaprezentowana w poszczególnych rozdziałach treść stanowi warstwę opisowej monografii. Aplikacyjność metodyki zostaje wyznaczona w praktyce inżynierskiej w działaniach konstrukcji architektonicznych struktur danych w chmurze cyfrowej, inżynierii wymagań, opracowania cyklu życia programowalnego robota cyfrowego oraz algorytmów z identyfikacją właścicieli procesów. Warstwa opisowa i aplikacyjna została wizualnie uzupełniona o warstwę ilustracyjną, tj. graficznego zaprezentowania modeli oraz algorytmów w postaci schematów blokowych, co stanowi szczególną wartość poznawczą, zarówno na poziomie teoriopoznawczym, jak i aplikacyjnym. Wykorzystanie różnych technik wizualnych pozwoliło przenieść tok rozumowania z płaszczyzny abstrakcyjnej na płaszczyznę gramatyki wizualnej, co umożliwia budowanie ogólnego obrazu i poznawanie zależności. Wszystkie rysunki oraz tabele w warstwie ilustracyjnej, pod którymi nie podano źródła są opracowaniami własnymi autora.

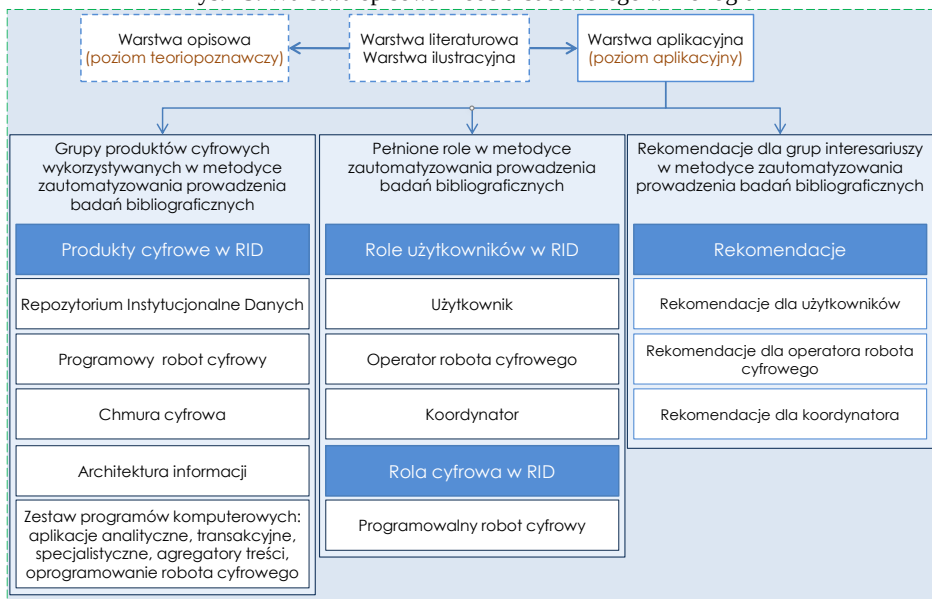
Graficznie, całokształt podjętej tematyki został algorytmicznie zaprezentowany na rysunkach 14–16.



Rys. 14. Algorytmiczne przedstawienie modelu badawczego w monografii



Rys. 15. Warstwa opisowa modelu badawczego w monografii

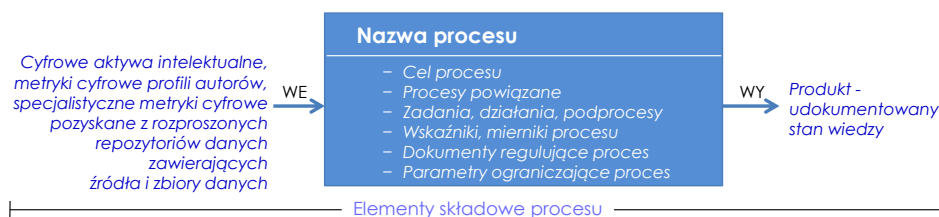


Rys. 16. Warstwa aplikacyjna modelu badawczego w monografii

Niniejsza monografia zapoczątkowała cykl prezentacji wyników badań nad sztuczną inteligencją w zakresie wspomagania decyzji – związaną z automatyzacją i wykorzystaniem programowych robotów cyfrowych w różnych obszarach życia społecznego i gospodarczego.

1. Metodyka zautomatyzowania prowadzenia badań bibliograficznych

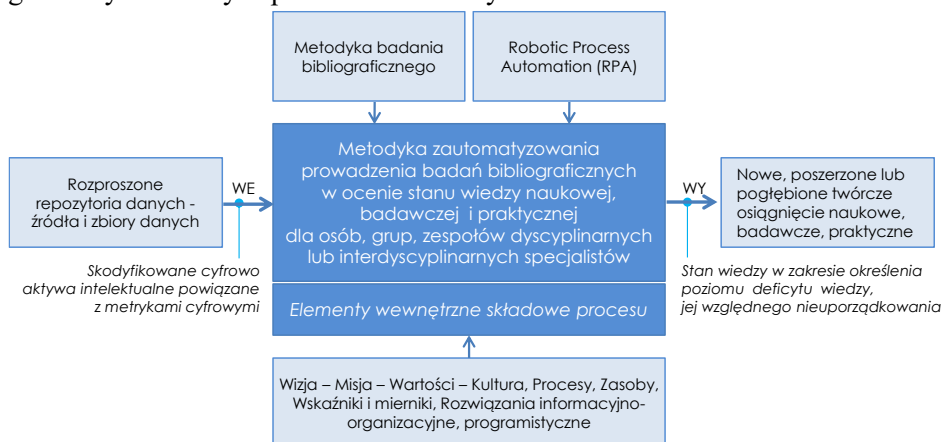
Jednym z kluczowych czynników sukcesu rzetelnej oceny stanu wiedzy skodyfikowanej jest profesjonalizacja i wieloaspektowe podejścia analityczne, wyrażające się kompleksowością, ciągłością, szybkością jej przeprowadzenia i zakończoną oceną ilościową oraz jakościową zebranych danych, adekwatną do planowanego nowego twórczego osiągnięcia. Przyczynkiem do przedstawienia metodyki badania bibliograficznego jest stwierdzony na podstawie badań deficyt metodyczny. W tym zakresie zostały ujęte wszystkie zależne procesy metodyczne, które były podstawą do opracowania autorskiej metodyki zautomatyzowanego prowadzenia badań bibliograficznych. W ujęciu procesowym MZPBB to proces zawierający zbiór powiązanych procesów mających na celu określenie stanu wiedzy naukowej, badawczej i praktycznej przez osoby, grupy, zespoły dyscyplinarne lub interdyscyplinarne specjalistów (rys. 1.1).



Rys. 1.1. Kluczowe elementy składowe procesu w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Wejście procesu stanowią skodyfikowane cyfrowo aktywa intelektualne, metryki cyfrowe profili autorów, specjalistyczne metryki cyfrowe dla aktywów intelektualnych pozyskane z rozproszonych repozytoriów danych zawierających źródła i zbiory danych. **Wyjście** procesu stanowi udokumentowany stan wiedzy dla określenia parametrów granicznych badań dla tworzenia nowej, rozszerzonej lub pogłębionej wiedzy. **Celem procesu** jest stwierdzenie stanu wiedzy, rozumianego jako określony poziom deficytu wiedzy, jej względnego nieuporządkowania. **Procesami powiązanymi** są procesy przynależne do platform robotyzacji procesów biznesowych oraz metodyki badania bibliograficznego. **Zadania, działania, podprocesy** umożliwiają osiągnięcie celu procesu. **Wskaźnikami, miernikami** metodyki zautomatyzowanego prowadzenia badań bibliograficznych była następująca, niewyczerpująca lista: stopień wykorzystanych zasobów, częstość aktualizacji, liczba dokumentów PDF zgromadzonych w RID, liczba dokumentów zduplikowanych. **Parametrami ograniczającymi**

proces, które ustanawiają granicę dla metodyki zautomatyzowanego prowadzenia badań bibliograficznych była następująca lista: wizja – misja – wartości – kultura, procesy, zasoby, wskaźniki i mierniki, rozwiązania informacyjno-organizacyjne i programistyczne. Kluczowe elementy składowe procesu (rys. 1.1) wykorzystane w metodyce zautomatyzowanego prowadzenia badań bibliograficznych zostały zaprezentowane na rysunku 1.2.

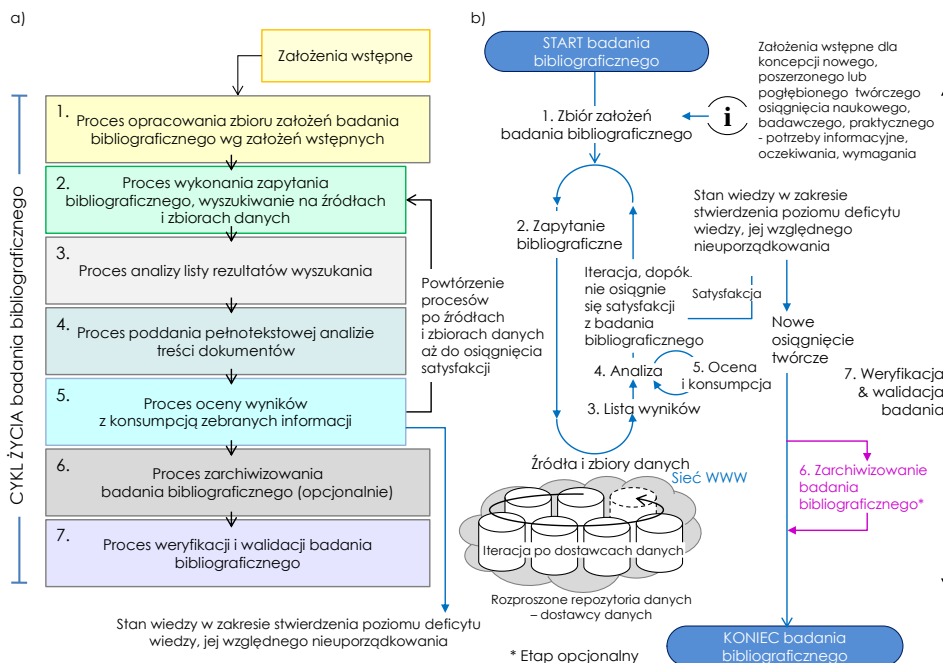


Rys. 1.2. Kluczowe elementy składowe procesu w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Podstawą do opracowania metodyki zautomatyzowanego prowadzenia badań bibliograficznych była metodyka badania bibliograficznego. Badania bibliograficzne posiadają swój cykl życia, tj. zestaw zależnych od siebie procesów, działań od inicjujących aż do zamykających tworząc cykl życia badania bibliograficznego. Uogólniając różne koncepcje, podejścia można zidentyfikować siedem procesów w metodyce:

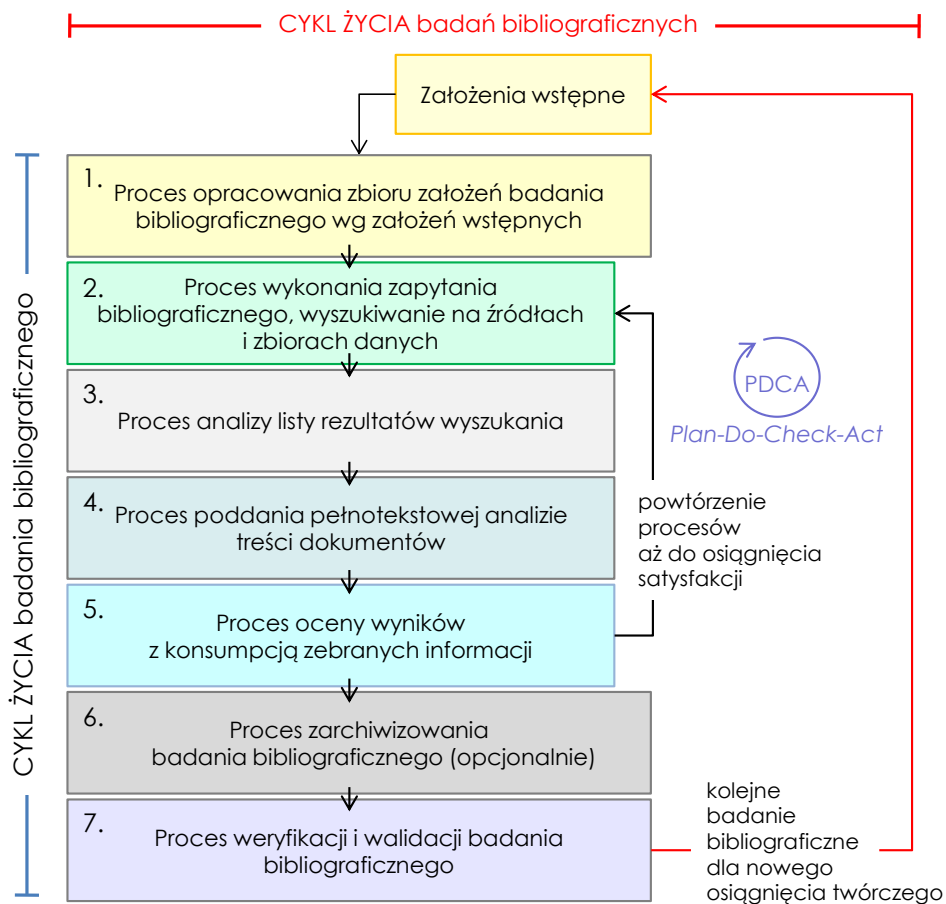
1. Proces zdefiniowania zbioru założeń badania bibliograficznego wg założeń wstępnych.
2. Proces wykonania zapytania bibliograficznego, wyszukiwanie na źródłach i zbiorach danych.
3. Proces analizy listy rezultatów wyszukiwania.
4. Proces poddania pełnotekstowej analizie treści dokumentów źródłowych i dokumentów stowarzyszonych.
5. Proces oceny wyników z konsumpcją zebranych informacji.
Decyzja o powtórzeniu procesów 2–5 aż do osiągnięcia satysfakcji z badania bibliograficznego.
6. Proces zarchiwizowania badania bibliograficznego (opcjonalnie).
7. Proces weryfikacji i walidacji badania bibliograficznego.

Uogólniona metodyka badania bibliograficznego w określeniu stanu wiedzy w zakresie stwierdzenia poziomu deficytu wiedzy, jej względnego nieuporządkowania dla nowego osiągnięcia twórczego została zaprezentowana na rysunku 1.3a i w sposób poglądowy na rysunku 1.3b (załącznik A).



Rys. 1.3. Uogólniona metodyka badania bibliograficznego: a) procesy w metodyce badania bibliograficznego w cyklu życia, b) schemat poglądowy badań bibliograficznych

Dla kolejnego osiągnięcia twórczego, następuje powtórzenie cyklu życia badania bibliograficznego. Powtarzalny cykl życia badania bibliograficznego dla wielu osiągnięć twórczych wyznacza cykl życia badań bibliograficznych specjalisty na tych samych źródłach i zbiorach lub uaktualnionych albo zmodyfikowanych (rys. 1.4). Stosowanie przez specjalistę wielokrotne metodyki badania bibliograficznego powinno skłaniać do refleksji, dążyć metodycznie uporządkowania procesów, dokonywać standaryzacji procesów i prowadzić do doskonalenia procesów w kolejnych podejściach. W tym procesie był zastosowany klasyczny cykl doskonalenia W. E. Deminga "planuj – wykonaj – sprawdź – działaj".



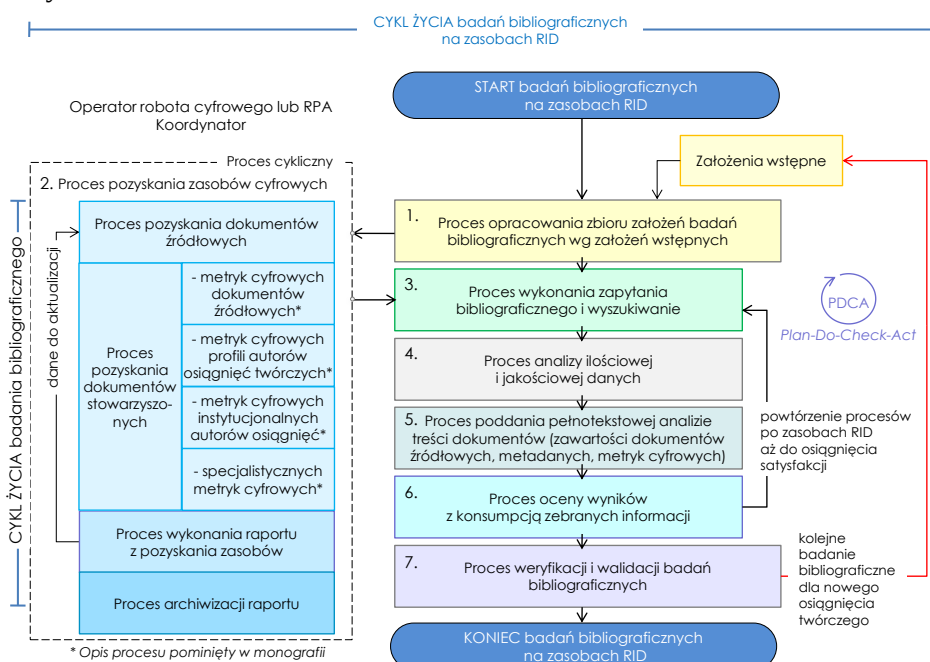
Rys. 1.4. Uogólniona metodyka badań bibliograficznych specjalisty dla wielu osiągnięć twórczych

Wymienione etapy stanowią ramę organizacyjną do opracowania autorskiej metodyki zautomatyzowanego prowadzenia badań bibliograficznych na zgromadzonych zasobach RID. Cykl życia metodyki zautomatyzowanego prowadzenia badań bibliograficznych rozpoczyna się momentem rozpoczęcia tworzenia zbioru założeń badań bibliograficznych wg określonych założeń i preferencji użytkownika a kończy procesem weryfikacji i walidacji badań bibliograficznych. Pełny cykl życia badania bibliograficznego na zasobach RID obejmuje siedem procesów:

1. Proces opracowania zbioru założeń badań bibliograficznych wg założeń wstępnych.

2. Proces cykliczny pozyskania zasobów: dokumentów źródłowych, metryk cyfrowych: dokumentów źródłowych, profili autorów osiągnięć twórczych, instytucjonalnych autorów osiągnięć twórczych, specjalistycznych metryk cyfrowych. Proces wykonania raportu z pozyskania zasobów. Proces archiwizacji raportu.
3. Proces wykonania zapytania bibliograficznego i wyszukiwanie w RID.
4. Proces analizy ilościowej i jakościowej uzyskanych danych.
5. Proces poddania pełnotekstowej analizie treści dokumentów (zawartości dokumentów źródłowych i dokumentów stowarzyszonych).
6. Proces oceny wyników z konsumpcją zebranych informacji.
Decyzja o powtórzeniu procesów 2–6 aż do osiągnięcia satysfakcji z badania bibliograficznego.
7. Proces weryfikacji i walidacji badań bibliograficznych.

Weryfikowana koncepcja metodyki zautomatyzowanego prowadzenia badań bibliograficznych na zasobach RID w postaci graficznej została zaprezentowana na rysunku 1.5.

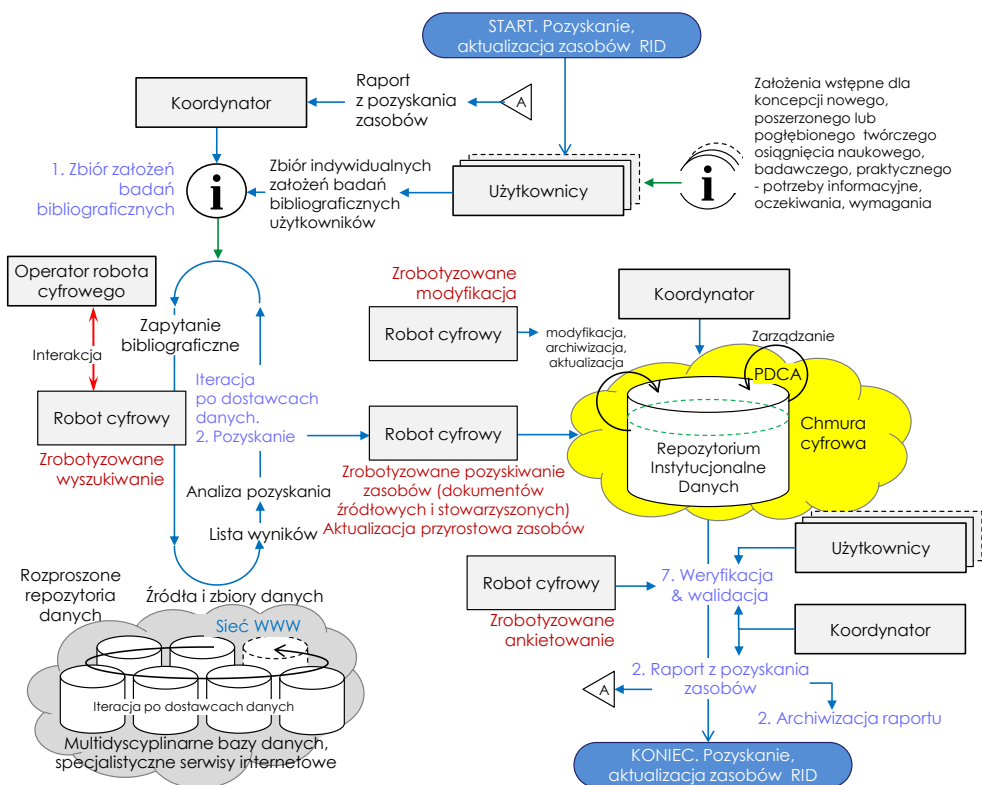


Rys. 1.5. Metodyka zautomatyzowanego prowadzenia badań bibliograficznych na zasobach Repozytorium Instytucjonalnego Danych

W niemniejszej monografii został pominięty opis procesów pozyskania metryk cyfrowych: profili autorów osiągnięć twórczych (metryk cyfrowych autorów), instytucjonalnych dla autorów osiągnięć twórczych oraz specjalistycznych metryk cyfrowych. Pominięte w opisie procesy były realizowane na podstawie danych bibliograficznych – metadanych statycznych, np. autorów osiągnięć twórczych pozyskanych z aktywów intelektualnych. Wynikiem procesu pozyskania zasobów był zbiór metadanych dynamicznych, który uzupełniał analizę ilościową i jakościową danych oraz ułatwił interpretację wyników przez specjalistów. Przetwarzanie metryk cyfrowych profili autorów osiągnięć twórczych, instytucjonalnych oraz specjalistycznych metryk cyfrowych był odrębnym tematem badawczym wymagającym modyfikacji cyklu życia robota cyfrowego, opracowania algorytmów dla automatyzacji, zbudowania oprogramowania robota cyfrowego, uruchomienia, walidacji uzyskanych wyników, włączenia wyników do wyników z analizy aktywów intelektualnych. Wymienione działania były komplementarne i wykonane na zasadzie analogii do opisywanej metodyki zautomatyzowanego prowadzenia badań bibliograficznych. Pominięte procesy będą przedmiotem dalszych prac analitycznych i programistycznych stanowiących integralną częścią metodyki automatyzacji badań bibliograficznych. Na tym etapie użytkowania RID metryki autorów osiągnięć naukowych, badawczych, praktycznych oraz metryki instytucjonalne i specjalistyczne były pozyskiwane w sposób manualny przez użytkowników lub koordynatora i włączone do zasobów RID.

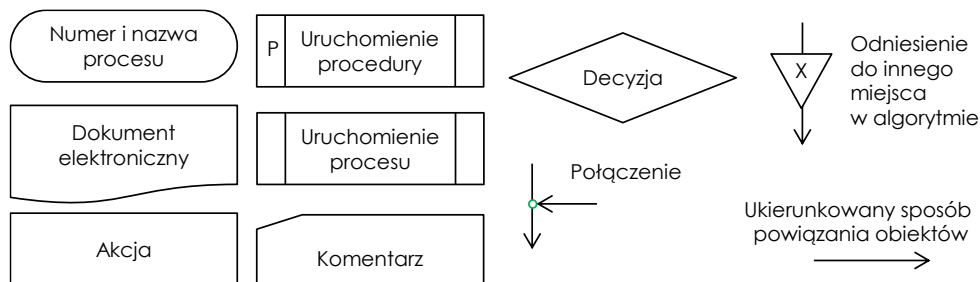
Kluczowym produktem cyfrowym wykorzystywanym w zautomatyzowanych badaniach bibliograficznych było utworzenie i zarządzanie Centralnym Repozytorium Danych umieszczonym w chmurze cyfrowej. Zaprezentowana metodyka zautomatyzowania prowadzenia badań bibliograficznych jest wynikiem doświadczeń autora z wdrożenia RID podczas pracy z różnymi przedsiębiorstwami realizowanymi w ramach pracy zawodowej, eksperckiej i została opisana z możliwościami rozszerzenia zakresu pozyskiwania zasobów (grup zbiorów danych) i skali użytkowania zasobów (grup odbiorców). Dlatego w warstwie opisowej monografii przeważający opis metodyki (rozdz. 1), zarządcze i organizacyjne aspekty wdrażania repozytorium (rozdz. 2) oraz rekomendacje (rozdz. 3) występuje gramatycznie w czasie teraźniejszym.

Model docelowy z wybranymi produktami cyfrowymi, interesariuszami na etapie pozyskania i aktualizacji zasobów RID został przedstawiony na rysunku 1.6.



Rys. 1.6. Proces pozyskania, aktualizacji zasobów RID w chmurze cyfrowej

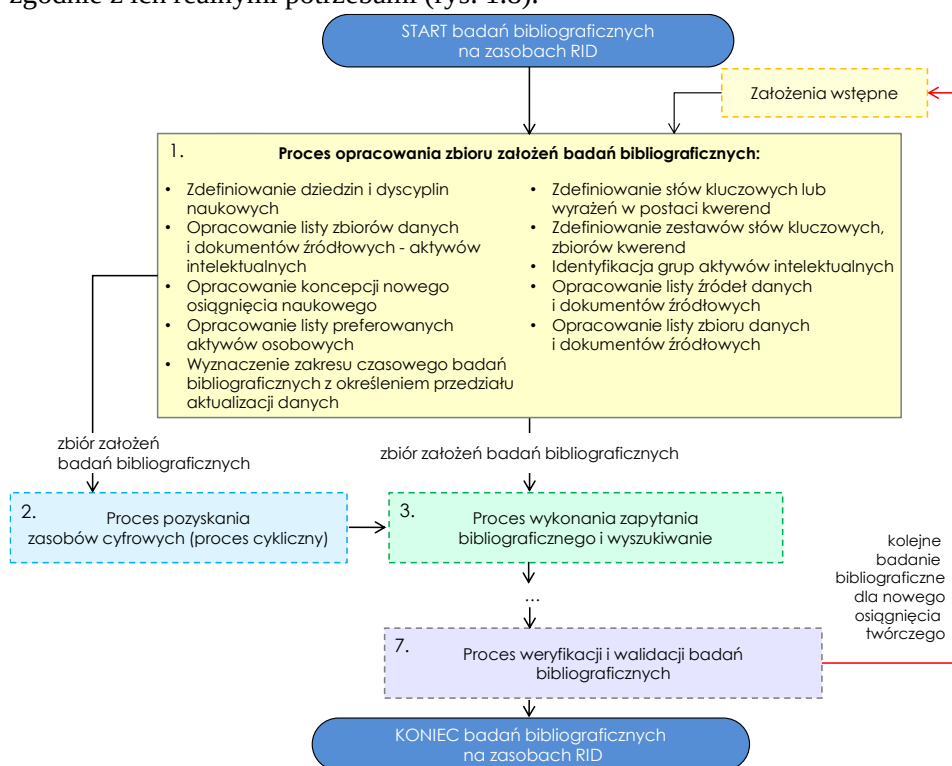
W następnych podrozdziałach zostaną zaprezentowane poszczególne procesy w cyklu życia badań bibliograficznych, w cyklu życia Repozytorium Instytucjonalnego Danych. Pod koniec każdego procesu badań bibliograficznych przedstawione zostały algorytmy graficzne z podziałem na role z niezbędnymi procedurami, które ustalają sposób przeprowadzenia procesu. W schematach algorytmów procesów wykorzystano 10 obiektów graficznych (rys. 1.7).



Rys. 1.7. Obiekty i elementy graficzne wykorzystane w schematach algorytmów procesów

1.1. Proces opracowania zbioru założeń badań bibliograficznych

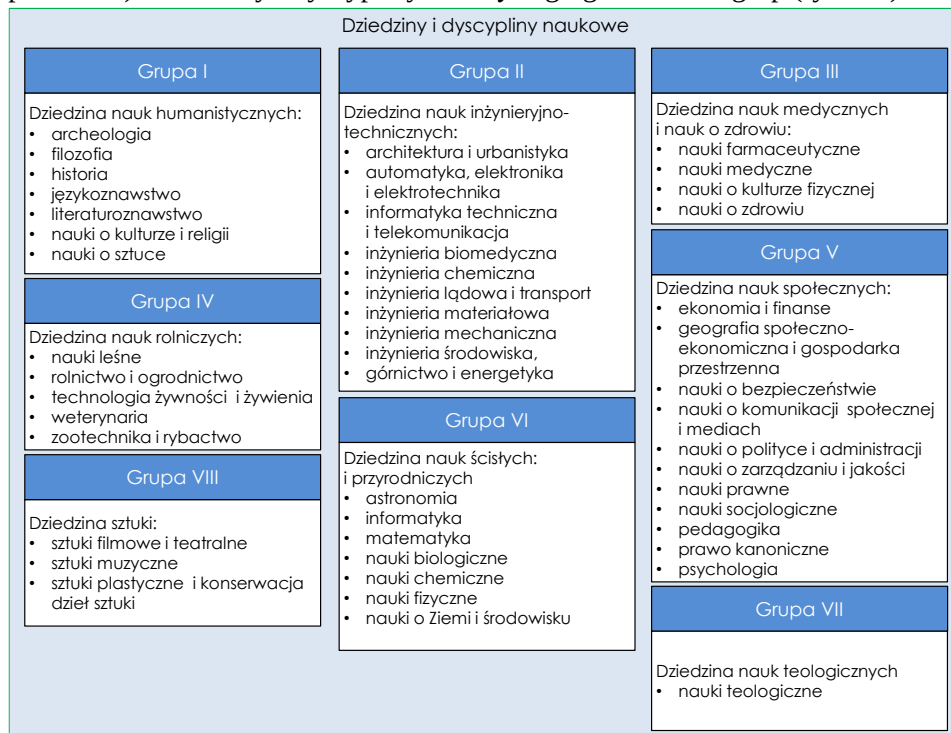
Pierwszym procesem w metodyce zautomatyzowanego prowadzenia badań bibliograficznych było opracowanie zbioru założeń bibliograficznych na podstawie indywidualnych założeń wstępnych opracowanych przez użytkowników. Z faktu wykorzystania metodyki przez osoby, grupy, zespoły monodyscyplinarnych lub interdyscyplinarnych specjalistów zbiór założeń uwzględniał wymagania specjalistów w dziedzinach i dyscyplinach naukowych zgodnie z ich realnymi potrzebami (rys. 1.8).



Rys. 1.8. Schemat pierwszego procesu dla opracowania zbioru założeń badań bibliograficznych

Proces opracowania zbioru założeń badań bibliograficznych składał się z kilkunastu działań wyznaczających parametry ograniczające badania lub dookreślające potrzeby informacyjne użytkownika. Większość wymagań użytkownika to wybór z zamkniętej listy opcji. Poniżej zaprezentowano listę zbioru założeń badań bibliograficznych. W algorytmie procesu nr 1 uwzględniono działania w opracowaniu założeń oddzielnie dla użytkownika i koordynatora.

W zakresie dziedzin i dyscyplin specjaliści określają swoje wymagania zamkniętej listy z klasyfikacji, która została określona w rozporządzeniu Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych (Dz.U. z 2018 r., poz. 1818). Dziedziny i dyscypliny zostały zagregowane w 8 grup (rys. 1.9).



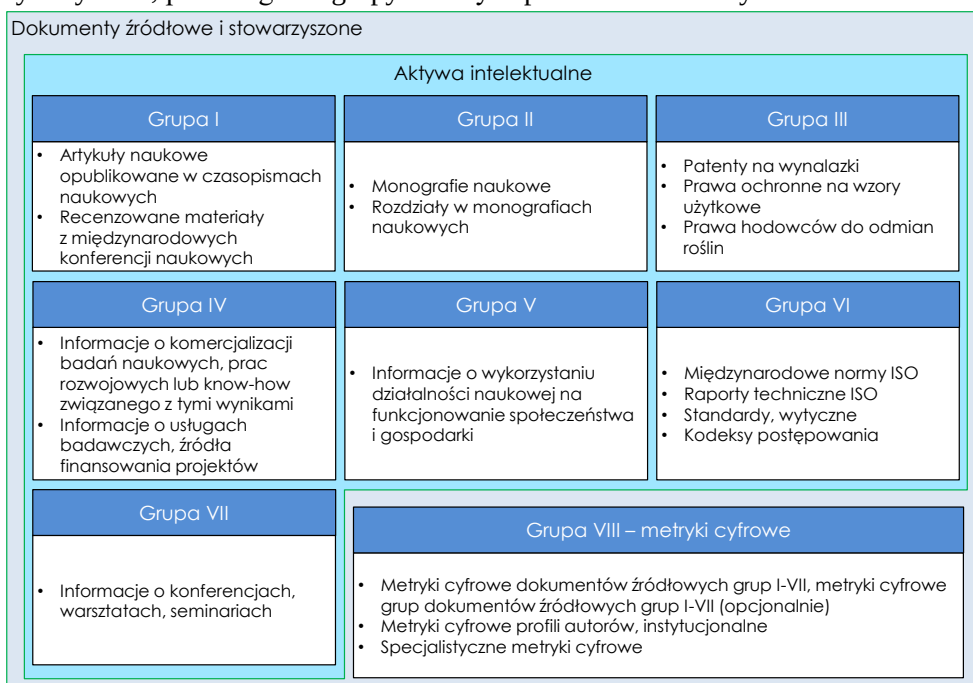
Rys. 1.9. Specyfikacja grup dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych
Źródło: Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych (Dz.U. z 2018 r., poz. 1818).

W zakresie zbiorów danych i dokumentów źródłowych zostały wyspecyfikowane siedem grup aktywów intelektualnych (grupa I–VII) i jedna grupa metryk cyfrowych (grupa VIII). Zbiory zbiorów danych i dokumentów źródłowych – aktywa intelektualne sklasyfikowane zostały w grupy:

- I. Artykuły naukowe opublikowane w czasopismach naukowych, recenzowane materiały z międzynarodowych konferencji naukowych.
- II. Monografie naukowe, rozdziały w monografiach naukowych.
- III. Patenty na wynalazki, prawa ochronne na wzory użytkowe, prawa hodowców do odmian roślin.

- IV. Informacje o komercjalizacji badań naukowych, prac rozwojowych lub *know – how* związanego z tymi wynikami, informacje o usługach badawczych, źródła finansowania projektów.
- V. Informacje o wykorzystaniu działalności naukowej na funkcjonowanie społeczeństwa i gospodarki.
- VI. Międzynarodowe normy ISO, raporty techniczne ISO, standardy, wytyczne, kodeksy postępowania.
- VII. Informacje o konferencjach, warsztatach, seminariach.
- VIII. Metryki cyfrowe dokumentów źródłowych grup I–VII, metryki cyfrowe grup dokumentów źródłowych grup I–VII (jeżeli istnieją i są dostępne), metryki cyfrowe autorskie, instytucjonalne, specjalistyczne metryki cyfrowe.

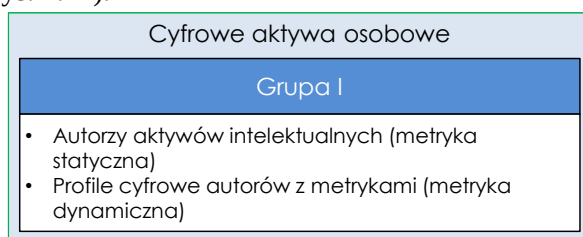
Syntetycznie, poszczególne grupy zostały zaprezentowane na rysunku 1.10



Rys. 1.10. Specyfikacja grup zbiorów danych, dokumentów źródłowych i stowarzyszonych zawierających aktywa intelektualne

Poszczególne pozycje w grupach aktywów intelektualnych nie są modyfikowalne przez użytkowników. Zaprezentowana klasyfikacja grup (liczba grup, pozycje aktywów) została opracowana arbitralnie i nie zawiera wszystkich, dostępnych aktywów intelektualnych możliwych do pozyskania^{34,35,36}. Na przykład w serwisie Scopus, tj. w bazie danych streszczeń i cytatów dla recenzowanych osiągnięć twórczych sklasyfikowano dane o 14 różnych typach publikacji: *Abstract report, Article, Book, Business Article, Book Chapter, Conference Paper, Conference Review, Editorial, Erratum, Article – in – Press, Letter, Note, Review, Short Survey*³⁷. W zakresie opracowania koncepcji nowego osiągnięcia naukowego każdy specjalista indywidualnie dokonuje wyboru, który jest zgodny z listą aktywów intelektualnych. Dodatkowo obligatoryjnie przygotowuje metrykę cyfrową dla swojego osiągnięcia twórczego.

W zakresie aktywów osobowych wyspecyfikowano jedną grupę, która obejmuje dane autorów aktywów intelektualnych oraz profile cyfrowe autorów z metrykami (rys. 1.11).



Rys. 1.11. Specyfikacja cyfrowych aktywów osobowych

W zakresie ustalenia przedziału czasowego badań bibliograficznych jest to zakres dwóch dat ustalany dla zestawu zbioru założeń bibliograficznych. Przedział aktualizacji przyrostowej danych jest określony w cyklu, np. miesięcznym, kwartalnym lub na żądanie użytkownika albo działaniem inicjowanym przez koordynatora.

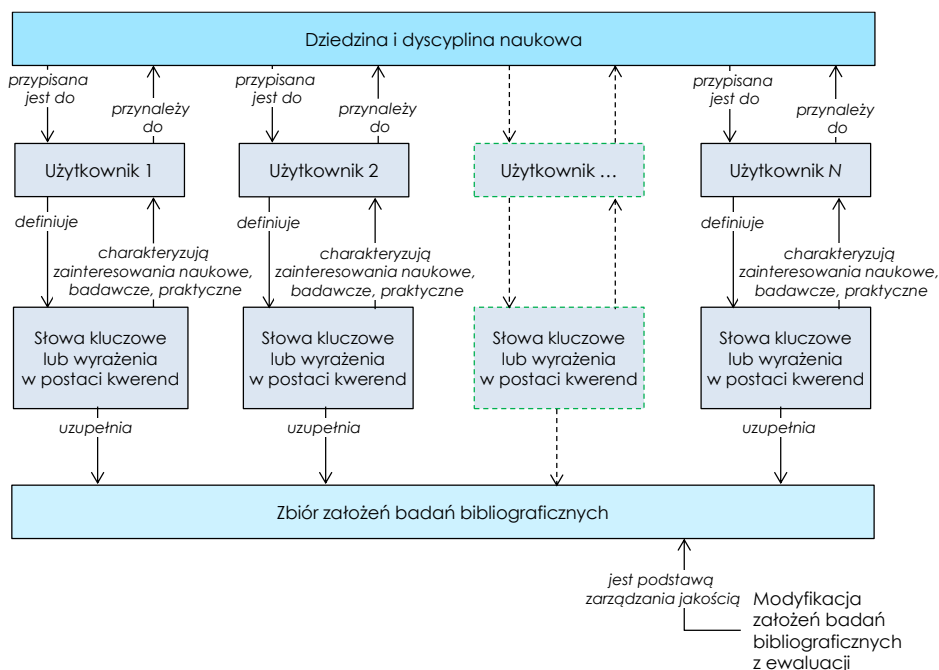
³⁴ Janiak, M., & Próchnicka, M. (2017). *Otwarte repozytorium Uniwersytetu Jagiellońskiego jako element systemu zarządzania informacją instytucjonalną: wyniki badań przeprowadzonych wśród przedstawicieli środowiska akademickiego uczelni*. Kraków: Biblioteka Jagiellońska.

³⁵ Mikołajuk L., *Nowoczesny system zarządzania wiedzą na przykładzie funkcjonowania repozytorium instytucjonalnego*, [w:] Piestrzyński T., Jerzyk-Wojtecka J. (red.), *Biblioteka w społeczeństwie wiedzy. Informacja, edukacja, profesjonalizm*. Konferencja naukowa Biblioteki Uniwersytetu Łódzkiego. Materiały konferencyjne, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015, s. 262–269, doi: 10.18778/8088–191–4.20.

³⁶ Gałuszka, E. (2017). *Repozytorium instytucjonalne jako narzędzie komunikacji naukowej i promocji nauki*. Nowa Biblioteka. Usługi, Technologie Informacyjne i Media, 27(4), 7–17.

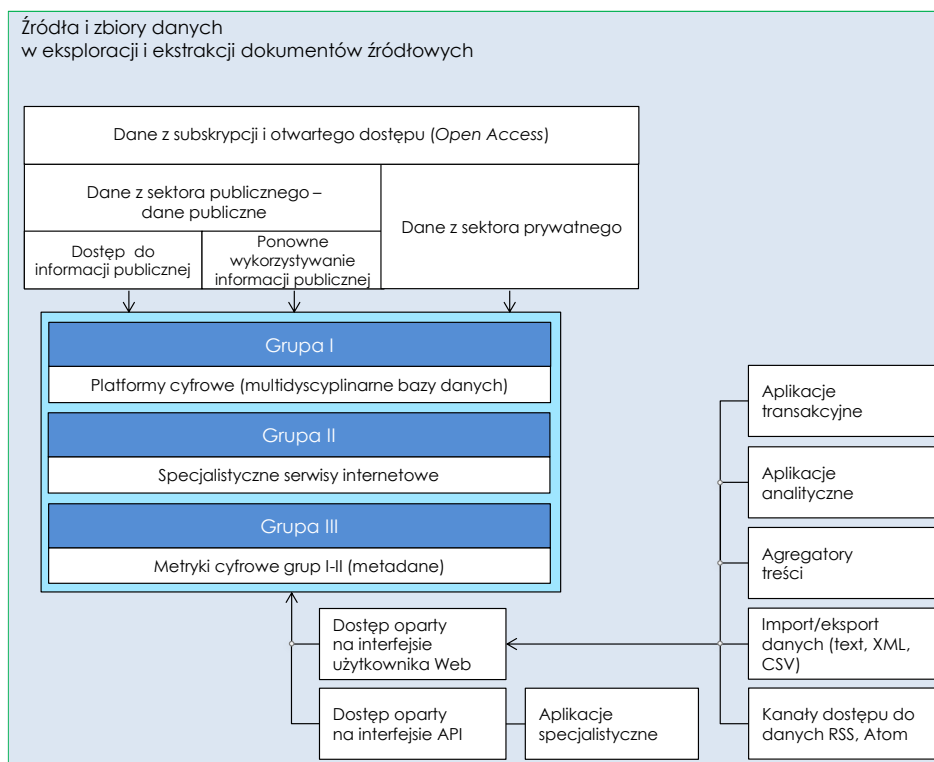
³⁷ *Research Intelligence. Research Metrics Guidebook*. Elsevier, March 2019, s.15.

W zakresie zestawu słów kluczowych, zbioru kwerend dane są pozyskiwane od użytkowników RID w dziedzinach i dyscyplinach naukowych (rys. 1.12).



Rys. 1.12. Schemat pozyskiwania i opracowania słów kluczowych lub wyrażeń w postaci kwerend

W zakresie grup źródeł danych i zbiorów danych w eksploracji, ekstrakcji dokumentów źródłowych wyspecyfikowano 2 grupy oraz grupę metryk cyfrowych opisujących zbiory (rys. 1.13).



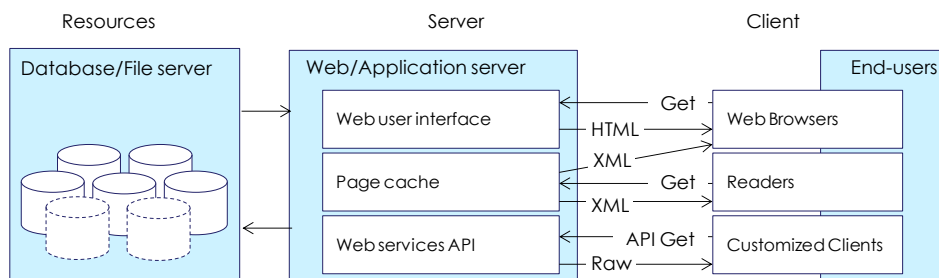
Rys. 1.13. Specyfikacja grup źródeł danych i zbiorów danych w eksploracji i ekstrakcji dokumentów źródłowych

Szczegółowa specyfikacja grup I–III nie jest jawna, gdyż stanowi *know – how* organizacji. Z punktu widzenia właściciela praw do źródeł i zbiorów danych dane mogą być pozyskiwane z sektora publicznego³⁸ lub z sektora prywatnego. Dostęp źródeł danych i zbiorów danych zależą od formuły prawnej jej udostępniania, może być odpłatny lub nieodpłatny. Podmioty wykonujące zadania publiczne, ze względu na ilość zgromadzonych informacji publicznych, są najważniejszymi źródłami wiedzy w państwie, które mogą być udostępniane innym podmiotom w zależności od ich indywidualnych potrzeb w sposób komercyjny lub niekomercyjny. Dostęp do informacji publicznej nie oznacza automatycznego prawa do jej ponownego wykorzystywania ze względu na ograniczenia prawne, tj. prawa własności intelektualnej do „utworu” lub „bazy danych”.

³⁸ Na przykład: Otwarte dane, <https://dane.gov.pl/institution/15,ministerstwo-nauki-i-szkolnictwa-wyzszego>, dostęp 22.12.2021 r.

Dostęp do zbiorów danych może odbywać się z wykorzystaniem interakcji pomiędzy człowiekiem a systemem komputerowym na dwa sposoby:

1. Graficznego interfejsu użytkownika (*Graphical User Interface*, GUI) przeglądarki internetowej lub innej aplikacji transakcyjnej, czytników treści lub agregatorów treści.
2. Interfejsu programowania aplikacji (*Application Programming Interface*, API) w bezpośrednim w sposobie komunikacji pomiędzy aplikacjami (rys. 1.14).



Rys. 1.14. Typowa architektura klient-serwer w dostępie klienta do danych na serwerze

Dla pierwszego przypadku w dostępie do zasobów, najczęściej wykorzystywane są przeglądarki internetowe (*Web Browsers*), za pomocą, których użytkownicy (*end-users*) mogą wyszukiwać, przeglądać i poddawać analizie informacje z wykorzystaniem rozwiązań w serwisie internetowym. W drugiej kolejności dostęp do zasobów może odbywać się z wykorzystaniem specjalistycznych aplikacji (botów). Jednak wiele serwisów w tym Google Scholar blokuje dostęp do zasobów, np. Twitterbot, Facebookexternalhit, PetalBot³⁹. W przypadku manualnej obsługi źródeł danych z wykorzystaniem interfejsu użytkownika istnieje możliwość połączeń z danymi zewnętrznymi z wykorzystaniem kanałów RSS (*Really Simple Syndication*) lub Atom. RSS/Atom jest kanałem dystrybucji informacji pomiędzy serwisem a aplikacjami klienckimi tzw. czytnikami kanału RSS (*Readers*) umożliwiającymi pobieranie, odczyt danych w formacie XML (*Extensible Markup Language*). Rolą czytnika kanału RSS jest pobranie oraz wyświetlenie nagłówków informacji. RSS daje możliwość jednoczesnego abonowania wielu źródeł informacji i przeglądania nagłówków oraz czytania wiadomości w jednym miejscu, bez konieczności odwiedzania poszczególnych stron, również subskrypcji wyników wyszukiwania przez kanały RSS. Serwisy informujące o możliwości dystrybucji informacji

³⁹ Robots.txt, <https://scholar.google.com/robots.txt>, dostęp 22.12.2021 r.

posługują się hiperłączami w postaci ikon kanałów RSS. Skopiowanie adresu kanału do czytnika kanału RSS umożliwia abonowanie kanału. Z technicznego punktu widzenia RSS jest plikiem dołączonym do strony internetowej w postaci hiperłącza z ikoną RSS. Jest to plik o odpowiedniej strukturze języka HTML (*HyperText Markup Language*) zawierający nagłówki oraz dane. Plik nie zawiera treści informacji a jedynie nagłówki. Oprócz dedykowanych programów, czytniki RSS-ów są wbudowane jako funkcja w popularnych programach pocztowych.

Multidyscyplinarne bazy danych lub specjalistyczne serwisy internetowe są systemami informatycznymi, które mogą zawierać rozwiązania programistyczne API jako formę komunikacji pomiędzy serwerem (*Server*) a klientem (*Client*). Zalety uwzględniania wymiany danych na poziomie programów z uwzględnieniem API w dostępie do źródeł danych są następujące:

- możliwość wyszukiwania, wyodrębnienia potrzebnych danych, pozyskiwania danych w czasie rzeczywistym,
- możliwość pozyskiwania danych na dużą skalę i możliwość zautomatyzowanego czynności pozyskiwania danych,
- możliwość interakcji z innymi systemami bez uwzględniania ich komplementarności i zgodności,
- niskie koszty: 1) implementacji API w oprogramowaniu po stronie klienta i serwera, 2) utrzymania, 3) udostępniania danych,
- elastyczność modyfikacji struktury danych i zawartości po stronie dostawcy API,
- skalowalność rozwiązania przez możliwość rozszerzenia zakresu udostępnianych danych,
- możliwość identyfikacji i autoryzacji użytkownika korzystającego z API.

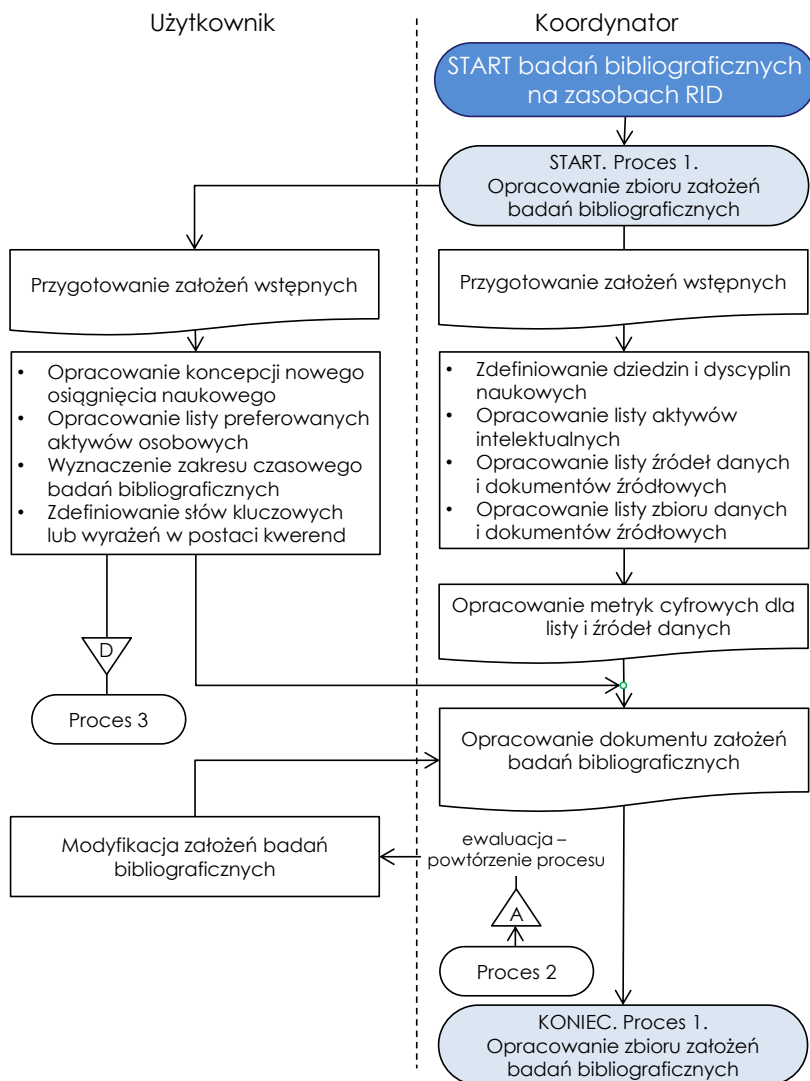
Wady implementacji API w oprogramowaniu w dostępie do źródeł danych są następujące:

- nie każdy serwis internetowy posiada możliwości komunikacji przez API, np. najobszerniejszy zasób on-line Google Scholar nie posiada API⁴⁰,
- zmiany w API po stronie dostawcy danych wymuszają zmiany po stronie odbiorcy dla zachowania interoperacyjności,

⁴⁰ Google Scholar, <https://scholar.google.com/intl/en/scholar/help.html#export>, dostęp 12.09.2021 r.

- dostęp do źródeł danych wymaga zgodności komunikacji dwóch systemów, po stronie dostawcy i odbiorcy API. Interfejs API powinien być precyzyjnie opisany i udokumentowany umożliwiając programistom implementację.

Wynikiem procesu *Proces opracowania zbioru założeń badań bibliograficznych wg założeń wstępnych* jest dokument zawierający zbiór założeń badań bibliograficznych. Algorytm procesu został zaprezentowany na rysunku 1.15.



Rys. 1.15. Algorytm procesu opracowania zbioru założeń badań bibliograficznych dla założeń wstępnych

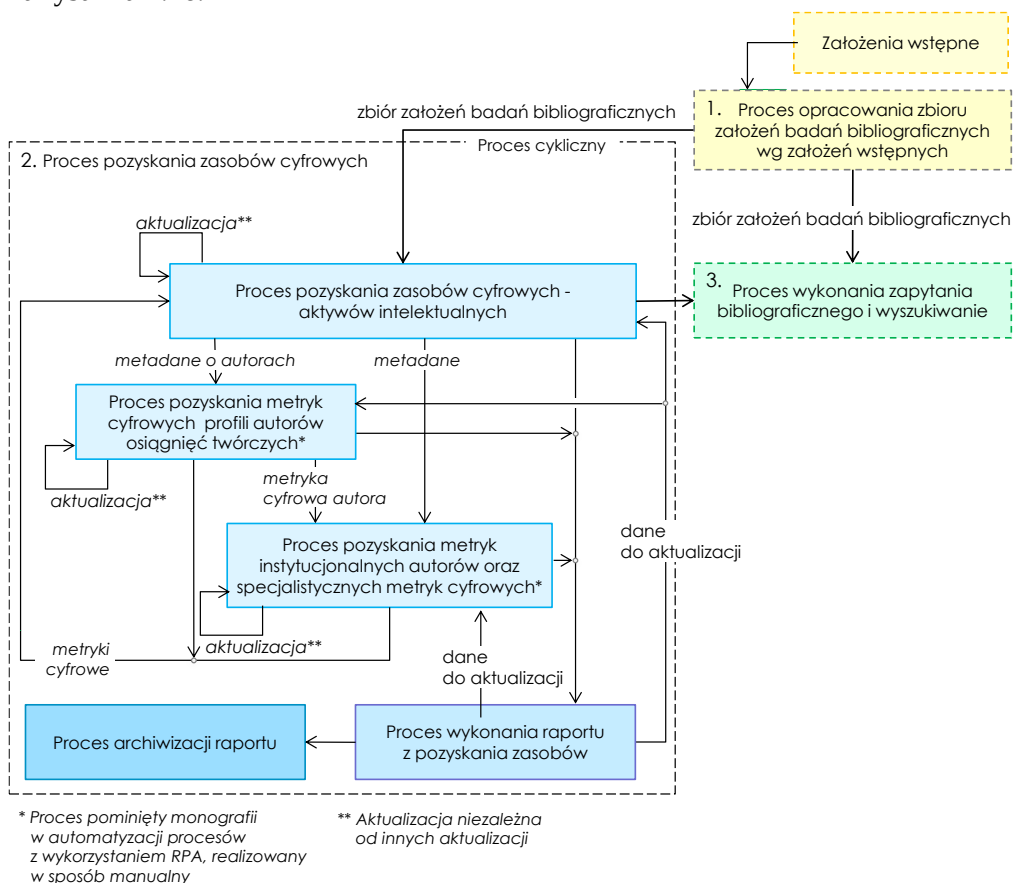
Kolejnym procesem jest proces pozyskania zasobów cyfrowych.

1.2. Proces pozyskania zasobów cyfrowych

Proces pozyskania zasobów cyfrowych obejmuje uzupełnienie lub aktualizację przyrostową zasobów RID i dotyczy trzech grup do pozyskania:

1. Aktywów intelektualnych.
2. Metryk cyfrowych profili autorów osiągnięć twórczych.
3. Specjalistycznych metryk cyfrowych.

W przypadku aktualizacji zasobów poszczególne grupy mogą być aktualizowane niezależnie od siebie, a więc częstość aktualizacji może być różna. Po wykonaniu procesu pozyskania dowolnej grupy zasobów następuje proces wykonania raportu z pozyskania zasobów oraz proces jego archiwizacji. Dane z raportu zostaną wykorzystane do procesu następnej aktualizacji. Ogólny schemat procesu pozyskania zasobów cyfrowych został zaprezentowany na rysunku 1.16.

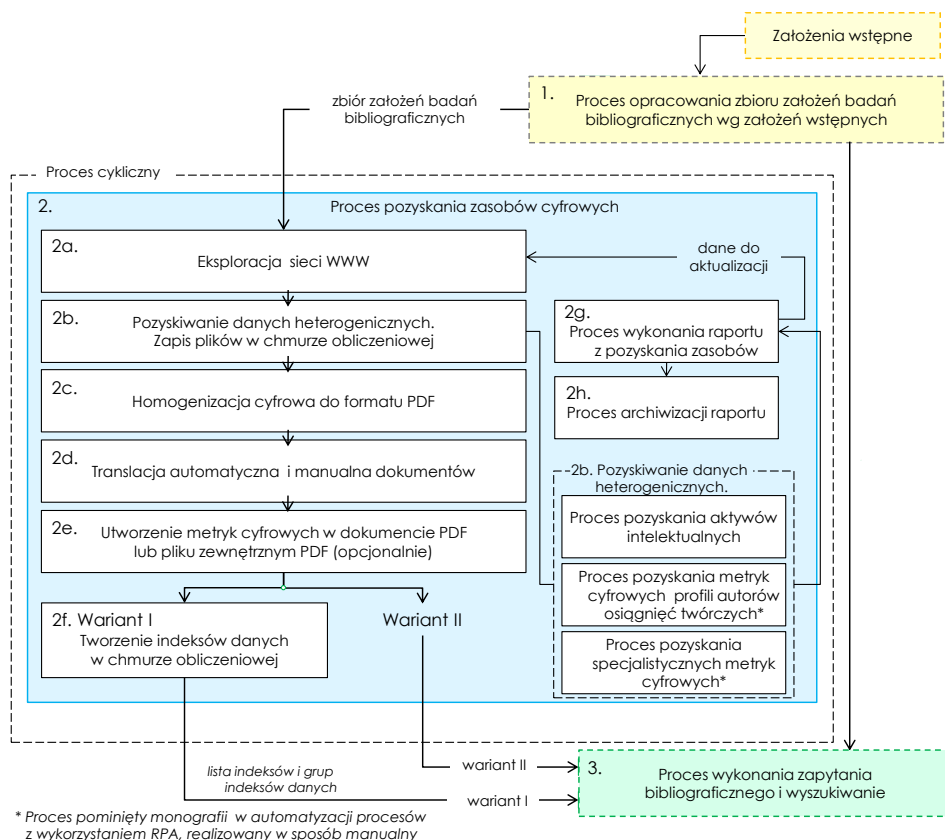


Rys. 1.16. Ogólny schemat procesu pozyskania zasobów cyfrowych

Proces pozyskania zasobów cyfrowych posiada osiem podprocesów:

1. Eksploracja sieci WWW.
2. Pozyskiwanie danych heterogenicznych: proces pozyskania aktywów intelektualnych, proces pozyskania metryk cyfrowych profili autorów osiągnięć twórczych, metryk instytucjonalnych autorów, proces pozyskania specjalistycznych metryk cyfrowych. Zapis plików w chmurze obliczeniowej.
3. Homogenizacja cyfrowa do formatu PDF.
4. Translacja automatyczna dokumentów i manualna dla zasobów niepodających się translacji automatycznej.
5. Utworzenie metryk cyfrowych w dokumencie PDF lub pliku zewnętrznym PDF (opcjonalnie).
6. Tworzenie indeksów danych w chmurze obliczeniowej tylko w wariancie I.
7. Proces wykonania raportu z pozyskania zasobów.
8. Proces archiwizacji raportu.

Powiązania pomiędzy procesami w procesie pozyskania zasobów cyfrowych zostały zaprezentowane na rysunku 1.17.

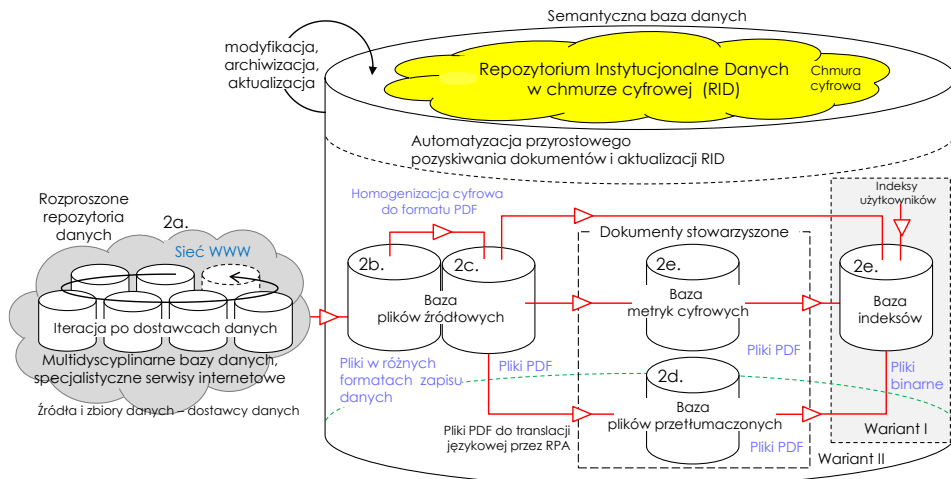


Rys. 1.17. Szczegółowy schemat procesu pozyskania zasobów cyfrowych

W procesie drugim zaplanowano dwa warianty uzupełniania zasobami RID. Wariant pierwszy uwzględnia utworzenie bazy indeksów danych w chmurze obliczeniowej. Baza indeksów obejmuje bazę dokumentów źródłowych w PDF, bazę metryk cyfrowych i bazę plików przetłumaczonych. Popularny program Adobe Acrobat DC i Acrobat Reader DC do grudnia 2018 roku umożliwiał osadzenie indeksu w pliku PDF do szybkiego wyszukiwania informacji. Program Acrobat Pro umożliwia utworzenie indeksu dla kolekcji dokumentów. Podczas tworzenia nowego indeksu program Acrobat tworzy plik binarny z rozszerzeniem PDX i nowym folderem obsługi, który zawiera jeden lub więcej plików z rozszerzeniem IDX. Pliki IDX zawierają wpisy indeksu.

Wszystkie pliki binarne muszą być dostępne dla użytkowników, którzy chcą przeszukiwać indeks⁴¹.

W wariancie drugim w zasobach RID nie tworzy się bazy indeksów. Użytkownicy korzystają z czytników dokumentów PDF (1.18).



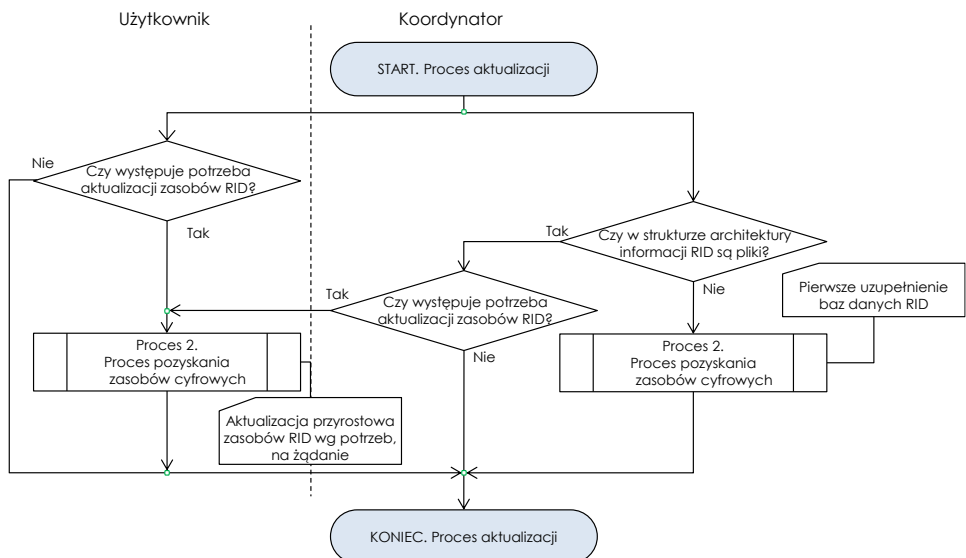
Legenda:

2a. – 2e. Nazwy podprocesów w procesie 2

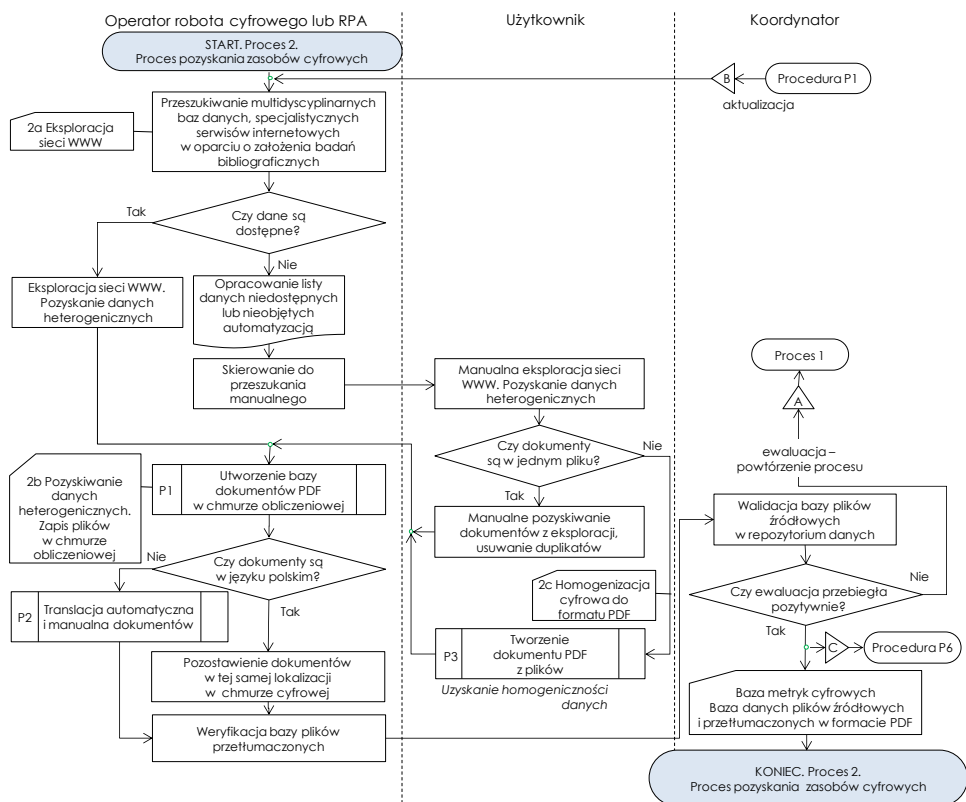
Rys. 1.18. Proces pozyskania zasobów cyfrowych w metodzie badań bibliograficznych na zasobach RID

Z punktu widzenia inicjacji procesu pozyskania zasobów cyfrowych może wystąpić w dwóch momentach: 1. Pierwszego uzupełnienia zasobami struktury RID, 2. Przyrostowej aktualizacji zasobów RID jako procesu cyklicznego lub proces inicjowany przez koordynatora w dowolnym momencie czasowym albo na żądanie użytkownika. Algorytm procesu drugiego uwzględniający dwa momenty wystąpienia aktualizacji zaprezentowano na rysunku 1.19. Algorytm procesu pozyskania zasobów cyfrowych z podziałem na pełnione role w procesie został zaprezentowany na rysunku 1.20.

⁴¹ Tworzenie indeksów PDF, https://helpx.adobe.com/acrobat/using/creating-pdf-indexes.html#create_an_index_for_a_collection_acrobat_pro

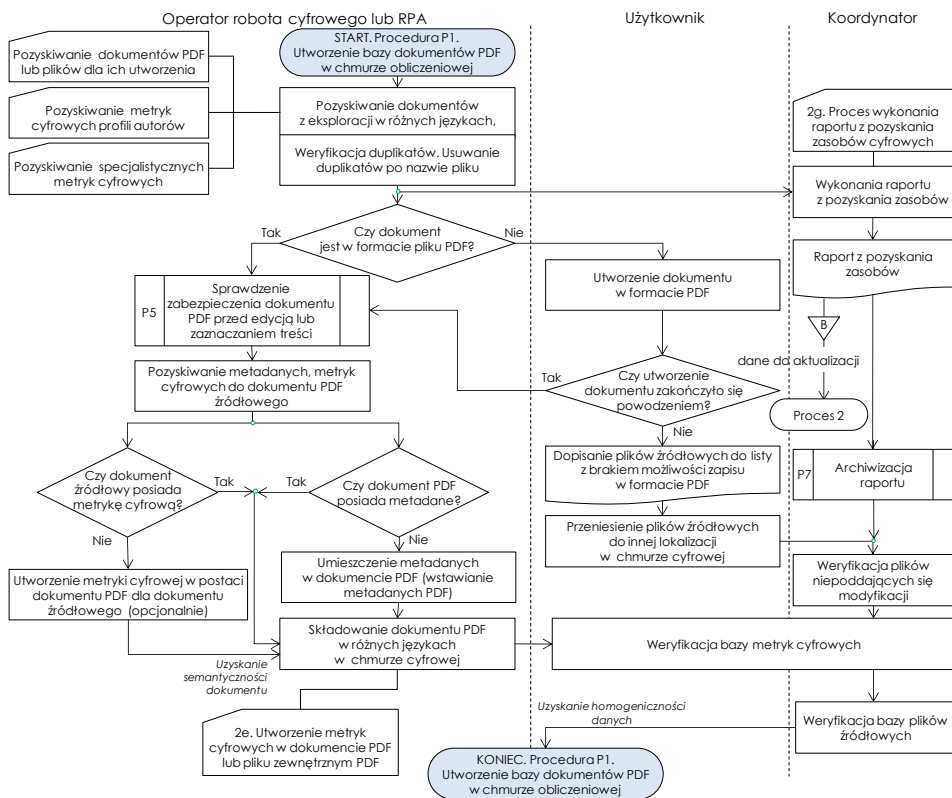


Rys. 1.19. Algorytm procesu aktualizacji w procesie pozyskania zasobów cyfrowych

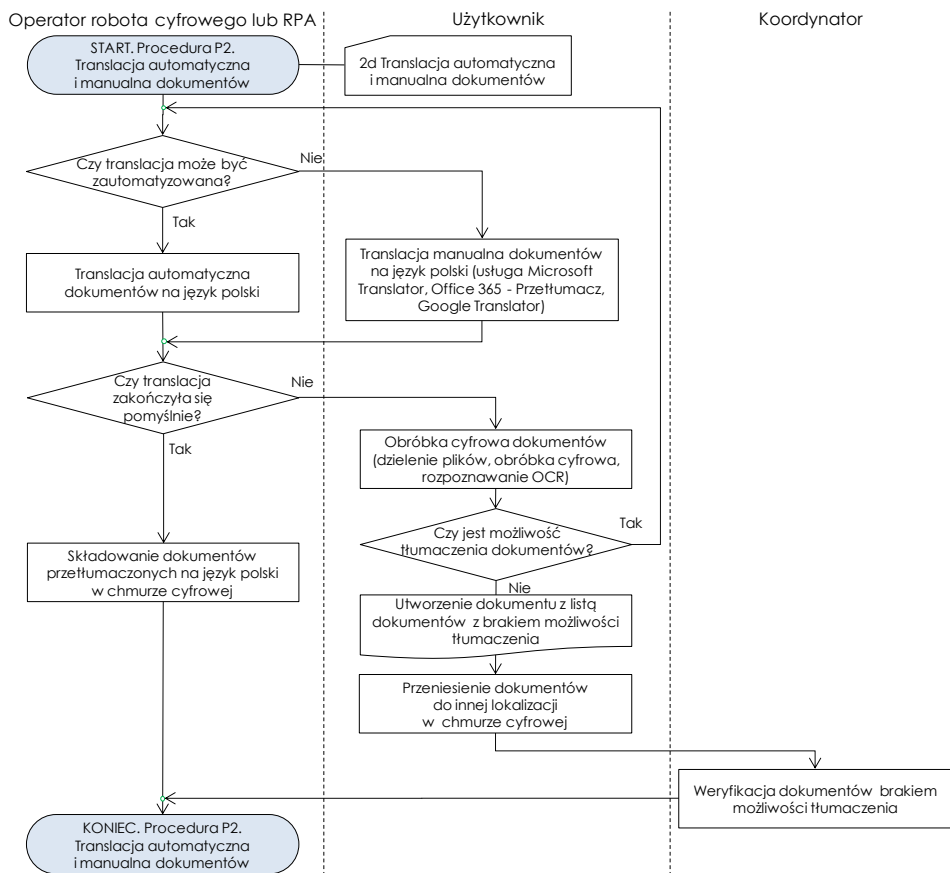


Rys. 1.20. Algorytm procesu pozyskania zasobów cyfrowych

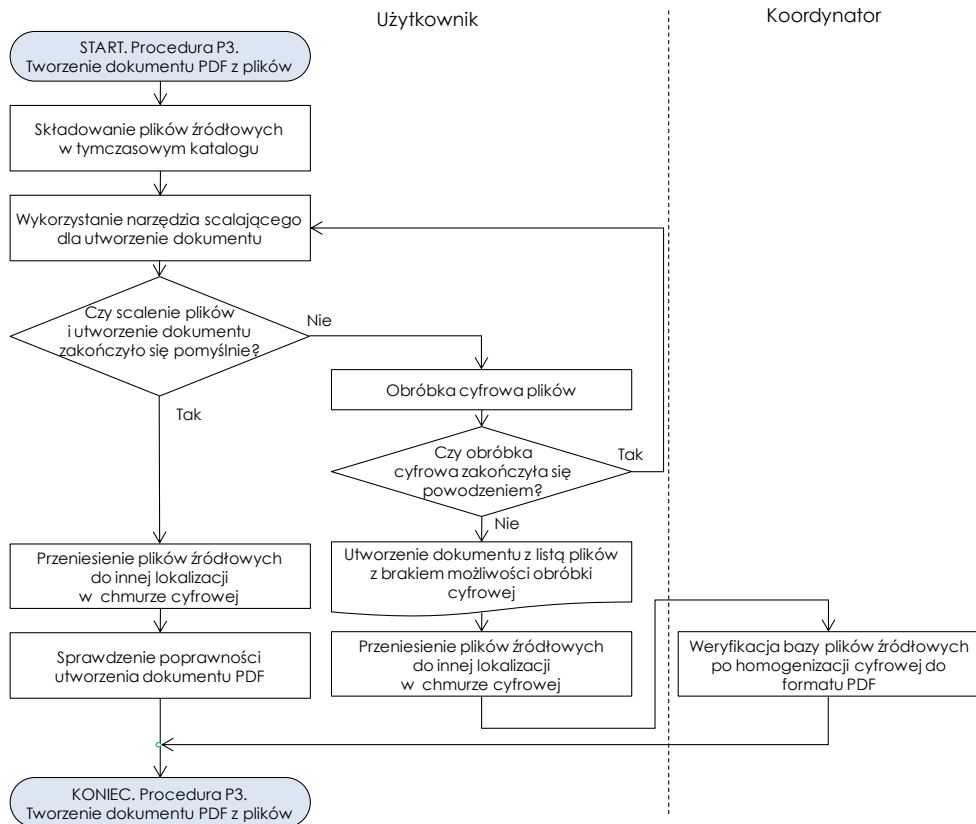
W algorytmie procesu drugiego wykorzystano siedem procedur: P1. *Utworzenie bazy dokumentów PDF w chmurze obliczeniowej* (rys. 1.21), P2. *Translacja automatyczna i manualna dokumentów* (rys. 1.22), P3. *Tworzenie dokumentu PDF z plików* (rys. 1.23), P4. *Tworzenie indeksów* (rys. 1.24), P5. *Sprawdzenie zabezpieczenia dokumentu PDF przed edycją lub zaznaczaniem treści* (rys. 1.25), P6. *Indeksowanie* (rys. 1.26), P7. *Archiwizacja raportu* (rys. 2.28). Redundancja danych może być realizowana przez programowalnego robota (procedura P1) lub z wykorzystaniem specjalistycznych aplikacji przez użytkownika (Proces 2).



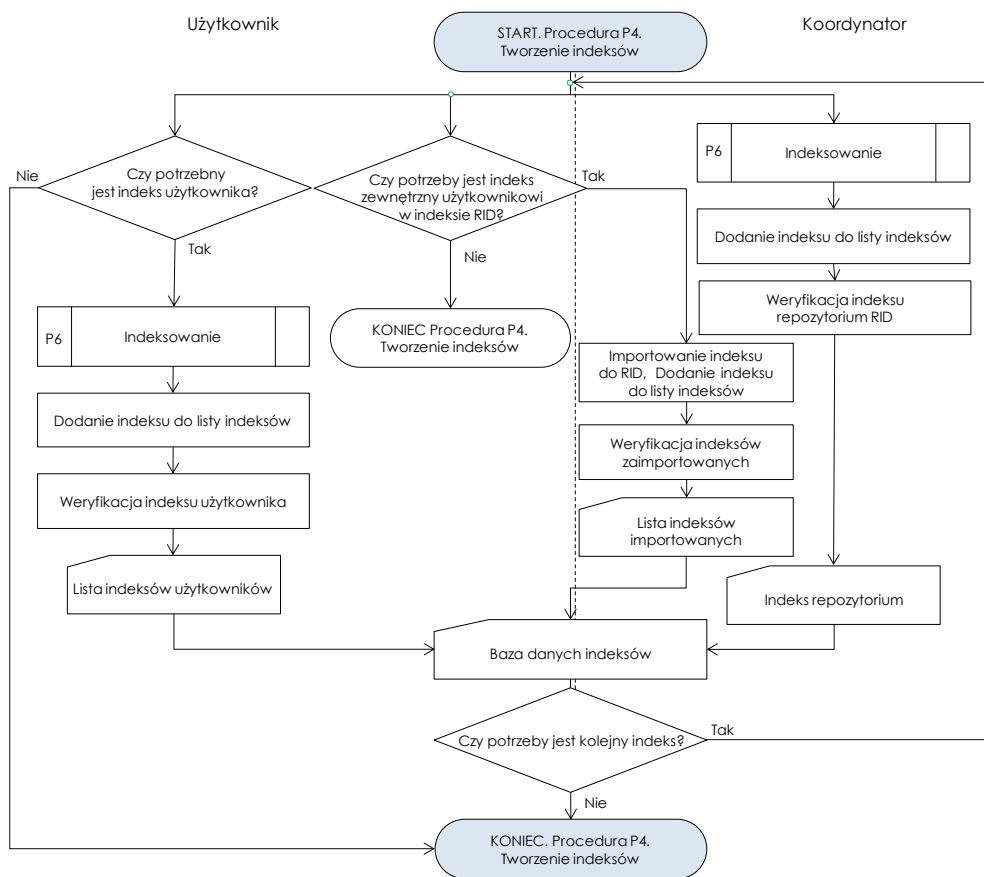
Rys. 1.21. Algorytm procedury P1. *Utworzenie bazy dokumentów PDF w chmurze obliczeniowej*



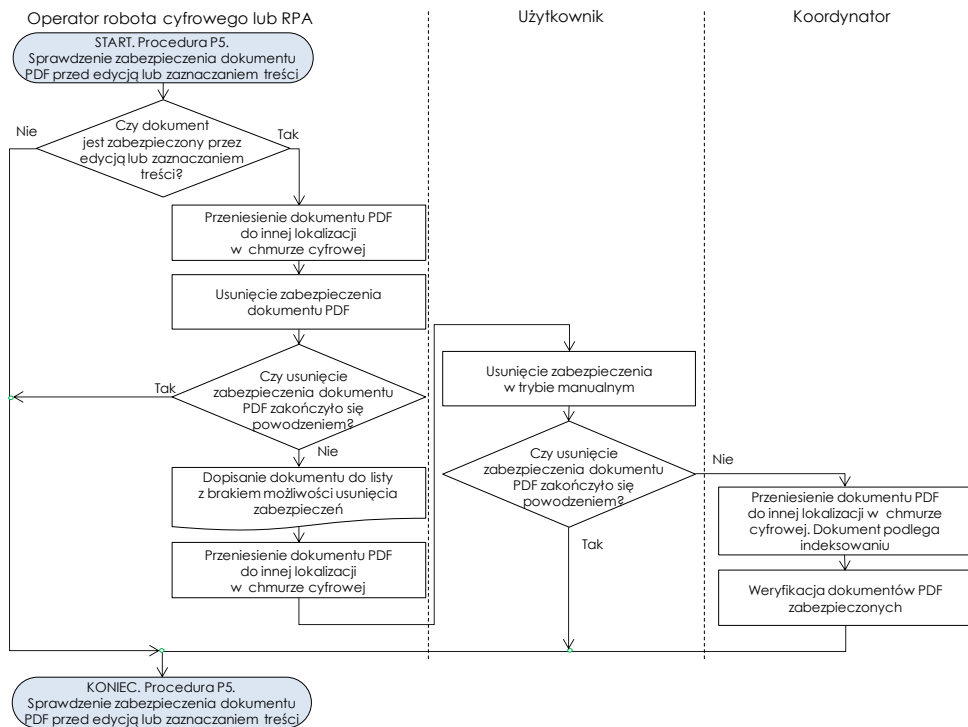
Rys. 1.22. Algorytm procedury P2. *Translacja automatyczna i manualna dokumentów*



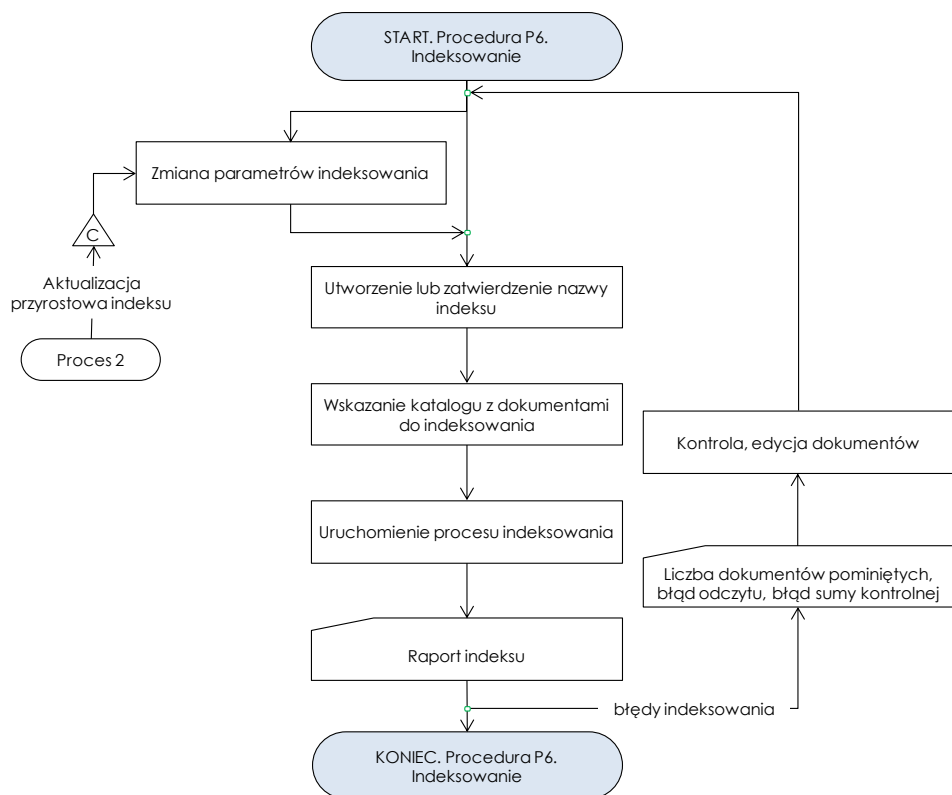
Rys. 1.23. Algorytm procedury P3. *Tworzenie dokumentu PDF z plików*



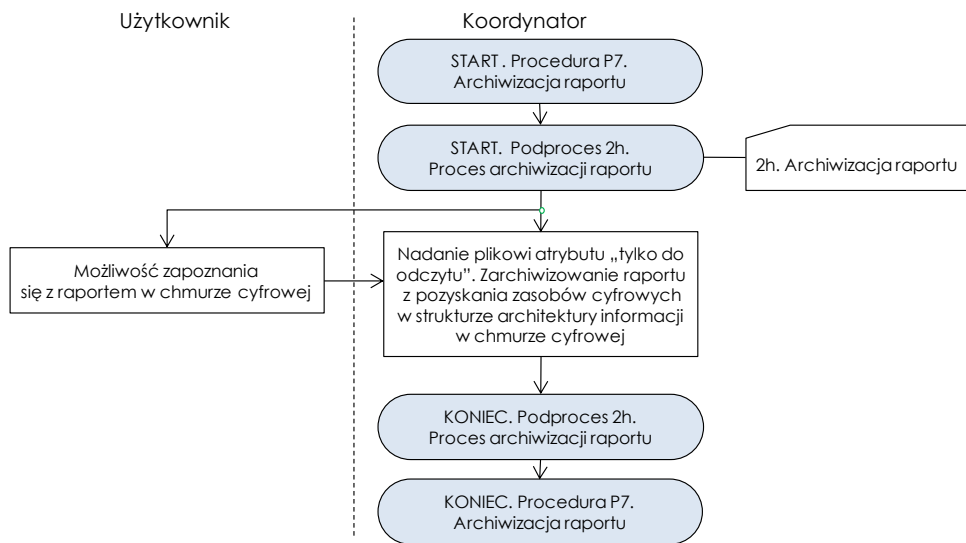
Rys. 1.24. Algorytm procedury P4. *Tworzenie indeksów*



Rys. 1.25. Algorytm procedury P5. Sprawdzenie zabezpieczenia dokumentu PDF przed edycją lub zaznaczaniem treści

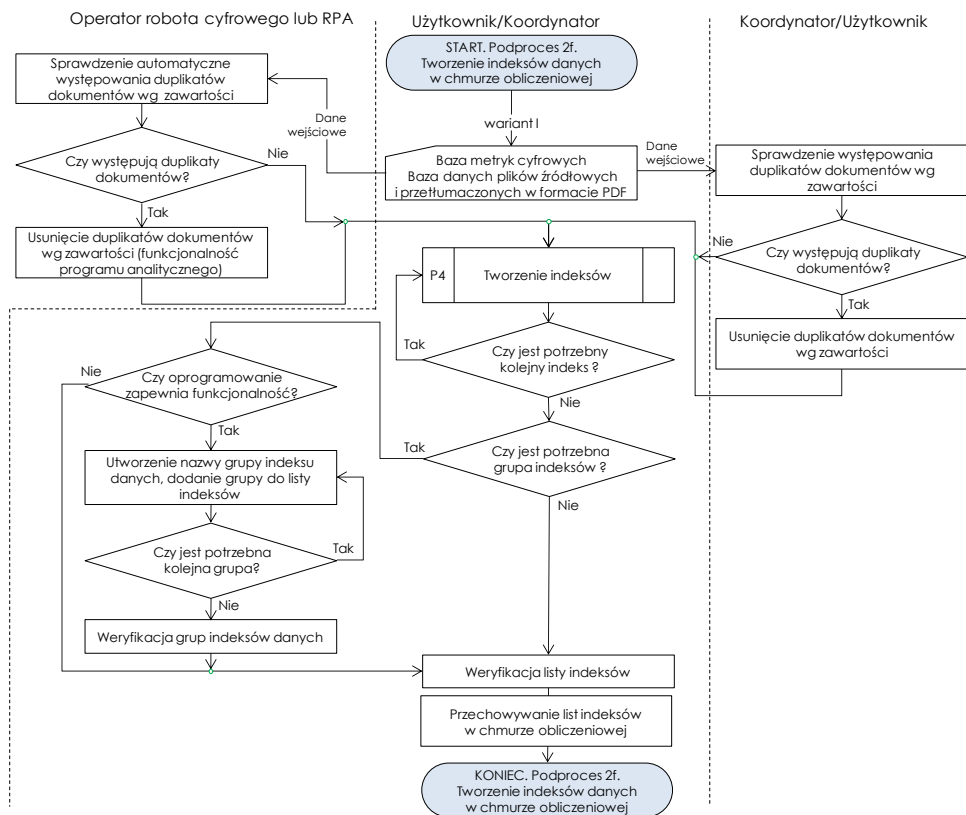


Rys. 1.26. Algorytm procedury P6. *Indeksowanie*



Rys. 1.27. Algorytm procedury P7. Archiwizacja raportu

Dla wariantu I opracowano algorytm podprocesu 2f. *Tworzenie indeksów danych w chmurze obliczeniowej* (rys.1.28). Wariant I jest tworzony obligatoryjnie. Alternatywnym rozwiązaniem dla użytkownika jest wyszukiwanie wyrazów lub wyrażeń z zaawansowanymi filtrami z wykorzystaniem czytników PDF.



Rys. 1.28. Algorytm podprocesu 2f. *Tworzenie indeksów danych w chmurze obliczeniowej*

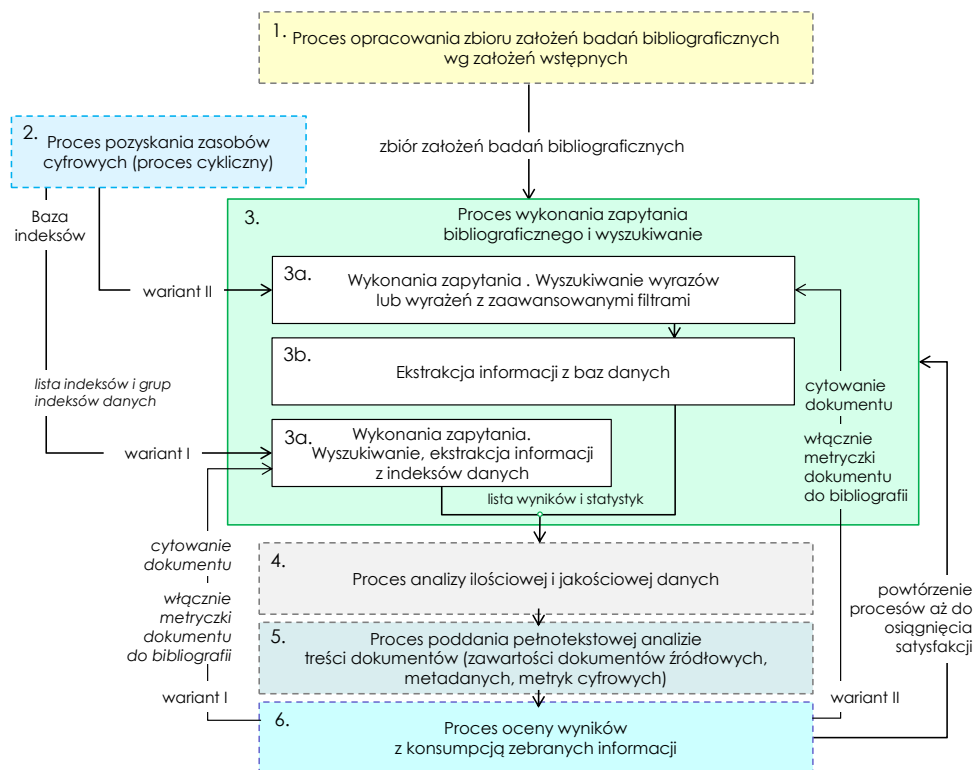
Wynikiem procesu pozyskania zasobów cyfrowych jest baza danych RID składająca się z dokumentów PDF zaindeksowanych z dokumentów źródłowych zhomogenizowanych cyfrowo do formatu PDF. W następnym procesie użytkownik będzie mógł skorzystać z przeszukiwania po indeksie głównym lub wyszukiwać po dokumentach źródłowych zhomogenizowanych cyfrowo w PDF.

1.3. Proces wykonania zapytania bibliograficznego i wyszukiwanie w zasobach RID

Kolejne etapy dotyczą użytkowników, którzy korzystają z bazy danych RID. Możliwe są dwa niezależne warianty wyszukiwania w zasobach RID.

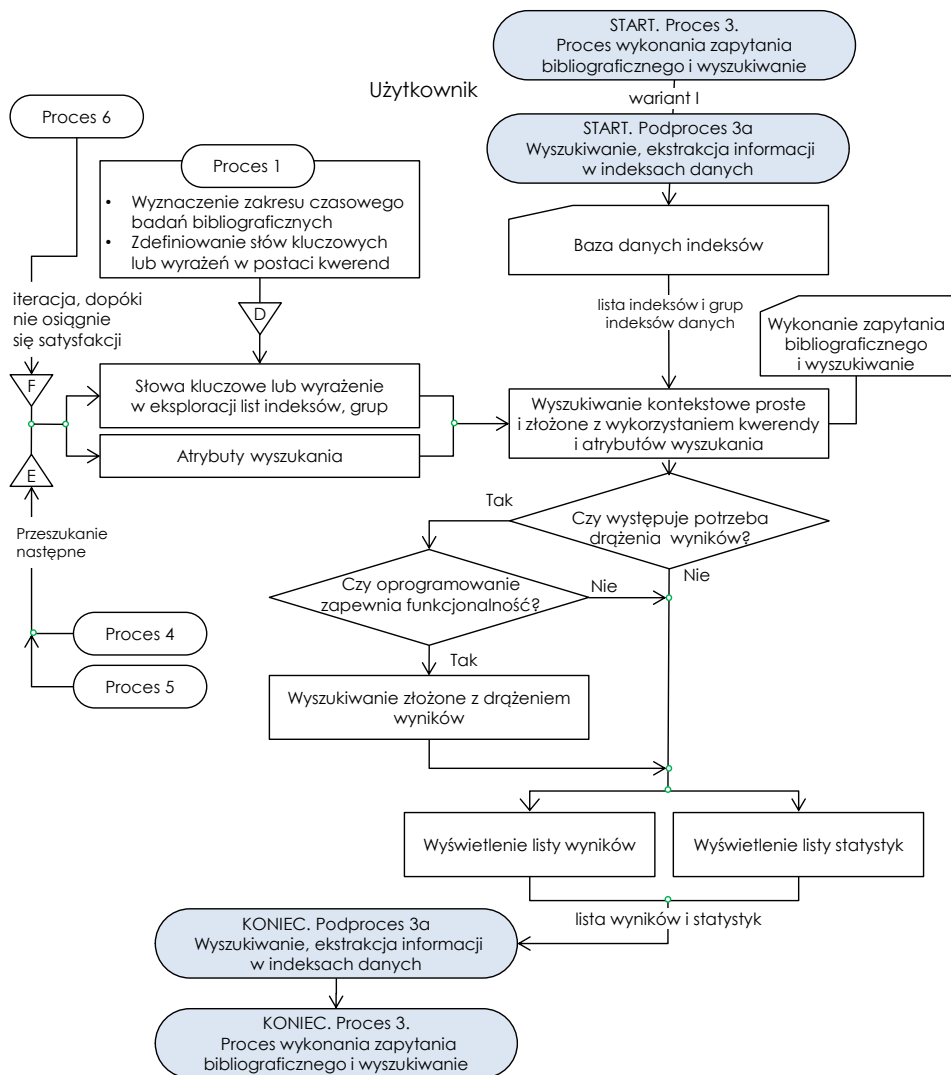
Wariant I dotyczy wykonania zapytania bibliograficznego i wyszukiwanie po wybranych przez użytkownika indeksach, dokonywać ekstrakcji informacji z indeksów danych (Załącznik C). Głównym indeksem dostępnym dla wszystkich użytkowników jest indeks repozytorium, który jest administrowany przez koordynatora. Użytkownik w momencie wyszukiwania może dodawać swoje indeksy (prywatne) z lokalnych zasobów, bez możliwości ich współdzielenia z innymi specjalistami, np. indeksując wiadomości i skrzynki pocztowe, dokumenty elektroniczne, bazy danych, dokumenty XML, archiwa WWW itp. Użytkownik może przekazać lub udostępnić indeks danych innym użytkownikom. Jednak rolą koordynatora jest dodawanie indeksów prywatnych lub importowanie indeksów z dokumentami źródłowymi do RID, które będą udostępniane wszystkim użytkownikom. Podczas przeszukiwania użytkownik ma możliwość zaznaczenia i łączenia wybranych indeksów do przeszukania z wykorzystaniem zbudowanej kwerendy. Swoboda wyboru indeksu do wyszukiwania oznacza, że użytkownik może nie korzystać z indeksu RID a wyłącznie z indeksu lub indeksów prywatnych lub udostępnionych indeksów przez innych użytkowników. W zależności od wykorzystanej aplikacji analitycznej język kwerend dla wykonania zapytania bibliograficznego może być zgodny z językiem zapytań internetowych wyszukiwarki Google a w zakresie funkcjonalności aplikacji uzyskuje się macierz odpowiedzi, którą można wykorzystać do analizy ilościowej oraz dalszego selekcjonowania zbiorów przez zaawansowane wyszukanie w zawężeniu wyników w celu dotarcia do relewantnych danych, np. drążenie wyników.

Wariant II dotyczy wykonania zapytania bibliograficznego i wyszukiwanie bezpośrednie na wszystkich dokumentach PDF. W zależności od użytego czytnika dokumentów PDF istnieje możliwość budowania wyrażeń z zaawansowanymi filtrami i wykonania informacji z baz danych. Szczegółowy schemat procesu wykonania zapytania bibliograficznego i wyszukiwania został przedstawiony na rysunku 1.29.

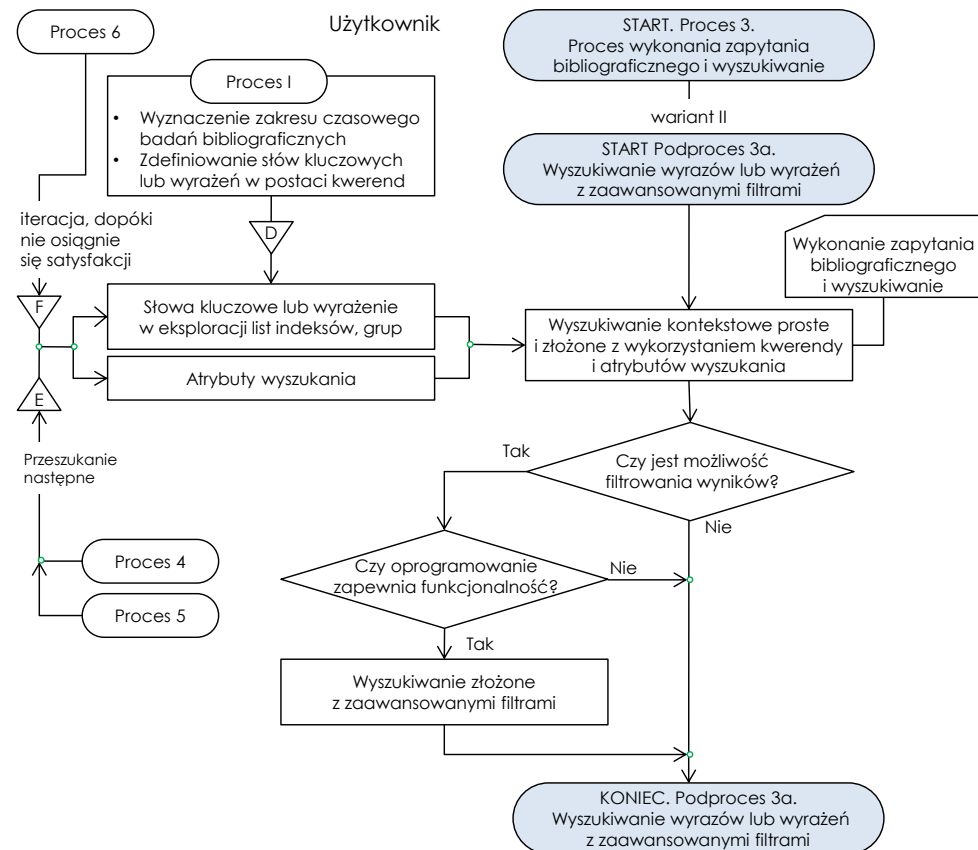


Rys. 1.29. Szczegółowy schemat procesu wykonania zapytania bibliograficznego i wyszukiwanie

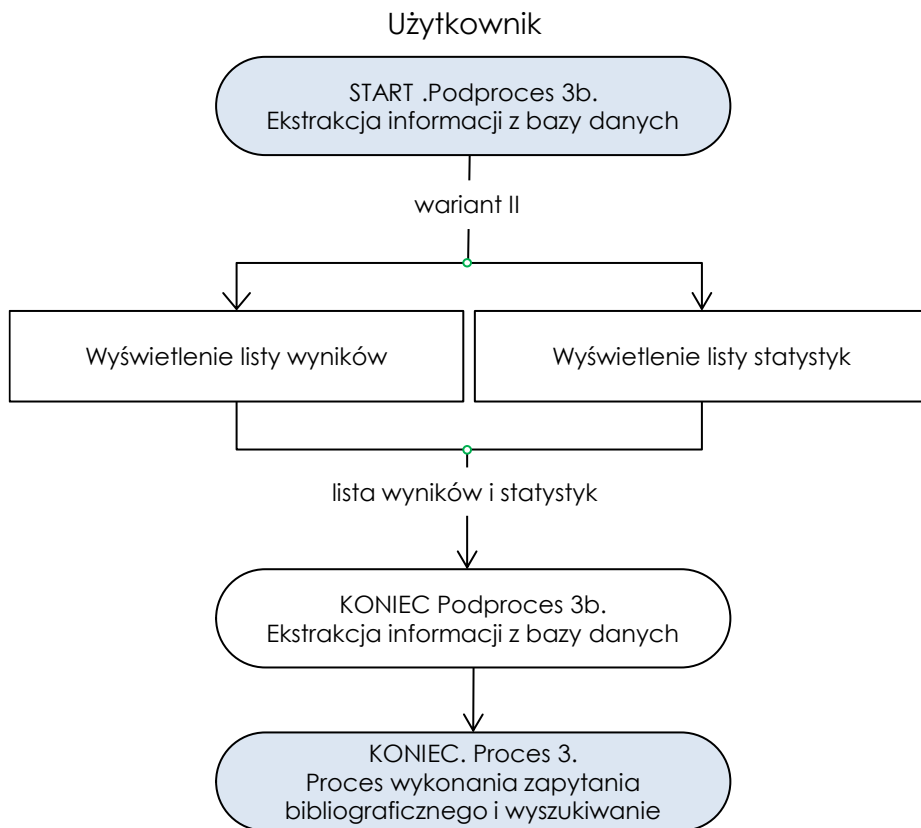
Algorytm procesu wykonania zapytania bibliograficznego i wyszukiwanie w wariancie I, po wybranych przez użytkownika indeksach został zaprezentowany na rysunku 1.30.



Rys. 1.30. Algorytm podprocesu 3a. Wyszukiwanie, ekstrakcja informacji w indeksach danych



Rys. 1.31. Algorytm podprocesu 3a. *Wyszukiwanie wyrazów lub wyrażeń z zaawansowanymi filtrami*

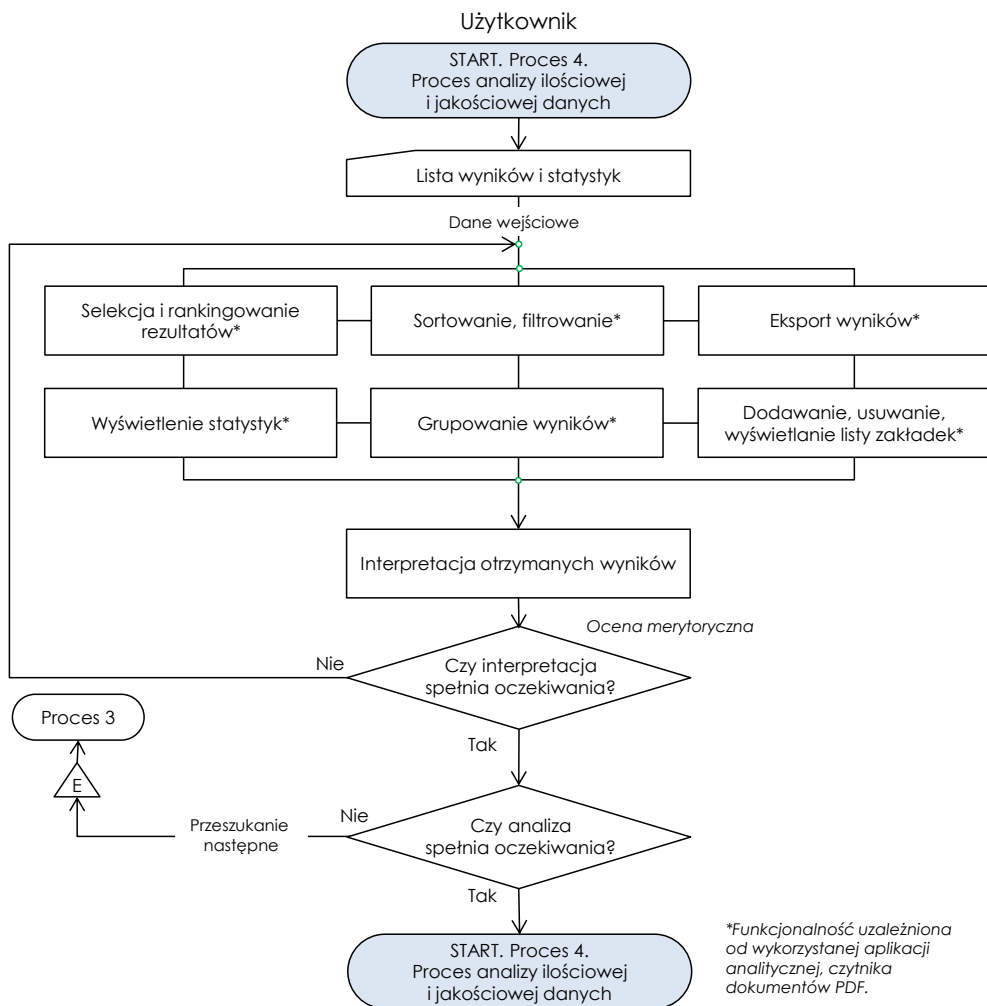


Rys. 1.32. Algorytm podprocesu 3a. *Ekstrakcja informacji z bazy danych*

Wynikiem procesu wykonania zapytania bibliograficznego i wyszukiwania jest lista otrzymanych wyników wyszukiwania oraz lista statystyk. Kolejny proces dotyczy wykonania analizy ilościowej i jakościowej danych.

1.4. Proces analizy ilościowej i jakościowej danych

Algorytm procesu analizy ilościowej i jakościowej danych został zaprezentowany na rysunkach 1.33.

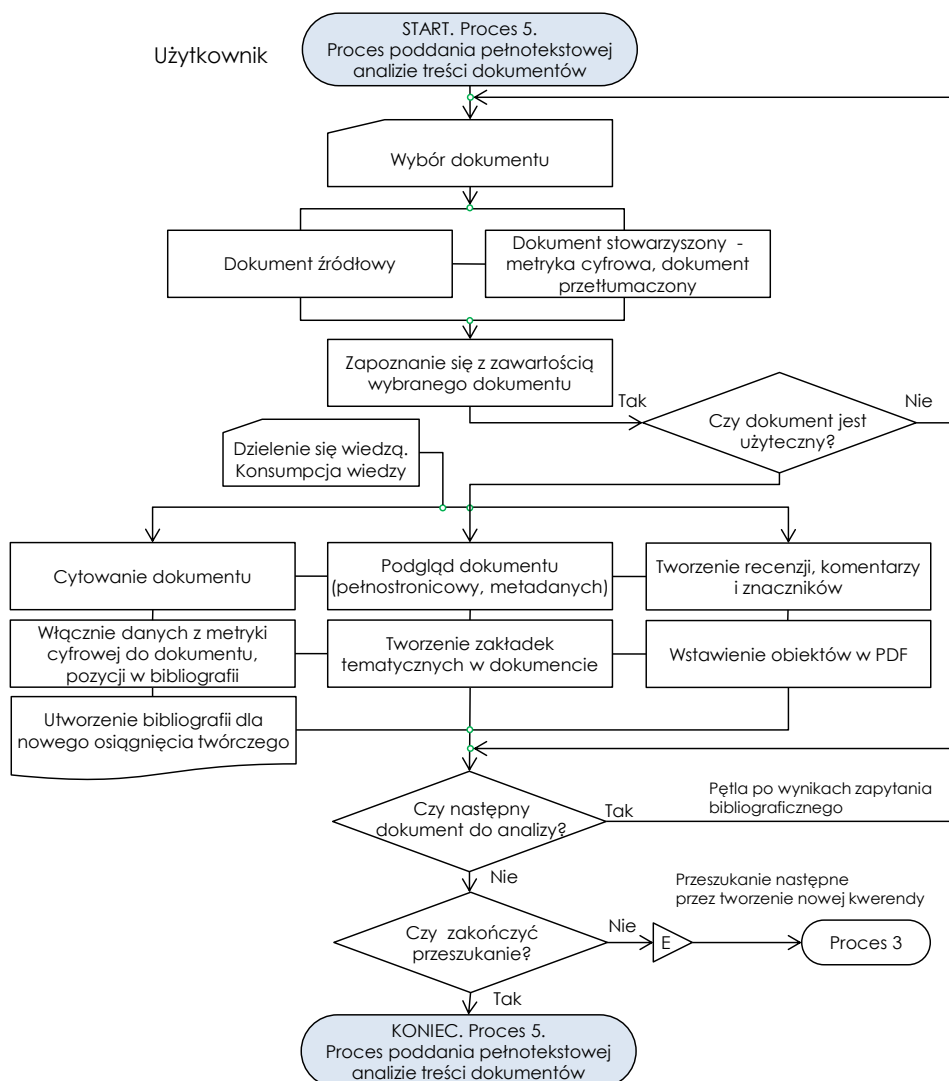


Rys. 1.33. Algorytm procesu analizy ilościowej i jakościowej danych

Wynikiem procesu analizy ilościowej i jakościowej danych jest interpretacja otrzymanych listy dokumentów i ich ocena merytoryczna. Kolejnym etapem jest proces poddania pełnotekstowej analizie treści dokumentów.

1.5. Proces poddania pełnotekstowej analizie treści dokumentów

Użytkownik dokonuje analizy treści dokumentów źródłowych oraz może zapoznać się z dokumentami stowarzyszonymi – metryką cyfrową lub dokumentem przetłumaczonym. Algorytm procesu poddania pełnotekstowej analizie treści dokumentów został zaprezentowany na rysunkach 1.34.



Rys. 1.34. Algorytm procesu poddania pełnotekstowej analizie treści dokumentów

1.6. Proces oceny wyników z konsumpcją zebranych informacji

```

graph TD
    Start([START. Proces 6.  
Proces oceny wyników  
z konsumpcją zebranych informacji]) --> Ocena[Ocena wyników]
    Ocena --> Satisfy{Czy występuje  
satisfakcja oceny  
wyniku?}
    
    Satisfy -- Tak --> End([KONIEC. Proces 6.  
Proces oceny wyników  
z konsumpcją zebranych informacji])
    
    Satisfy -- Nie --> Dzielnie[Dzielenie się wiedzą.  
Konsumpcja wiedzy]
    Dzielnie --> Cytowanie[Cytowanie dokumentu]
    Dzielnie --> Podglad[Podgląd dokumentu  
(pełnotronnicowy, metadanych)]
    Dzielnie --> Recenzje[Tworzenie recenzji, komentarzy  
i znaczników]
    
    Cytowanie --> Wlaczne[Włączenie danych z metryki  
cyfrowej do dokumentu,  
pozycji w bibliografii]
    Wlaczne --> Bibliografia[Utworzenie bibliografii dla  
nowego osiągnięcia twórczego]
    
    Podglad --> Zakladki[Tworzenie zakładzek  
tematycznych w dokumencie]
    Recenzje --> PDF[Wstawienie obiektów w PDF]
    
    Bibliografia --> Weryfikacja[Weryfikacja bibliografii]
    Zakladki --> Analiza{Czy następny  
dokument do analizy?}
    PDF --> Analiza
    
    Weryfikacja --> VerifPos{Czy weryfikacja  
bibliografii pozytywna?}
    VerifPos -- Tak --> End
    VerifPos -- Nie --> Korekta[Korekta]
    Korekta --> Bibliografia
    
    Analiza -- Tak --> Peta[Pętla po wynikach zapytania  
bibliograficznego]
    Peta --> Proc5([Proces 5])
    Proc5 --> Zakończ{Czy zakończyć  
przeszukiwanie?}
    Zakończ -- Tak --> End
    Zakończ -- Nie --> E([E])
    E --> Proc3([Proces 3])
    Proc3 --> Iteracja[iteracja, dopóki  
nie osiągnie  
się satysfakcji]
    Iteracja --> Satisfy
    
    Satisfy -- Nie --> G([G])
    G --> Proc7([Proces 7])
    Proc7 --> Weryfikacja

```

Rys. 1.35. Algorytm procesu oceny wyników z konsumpcją zebranych informacji

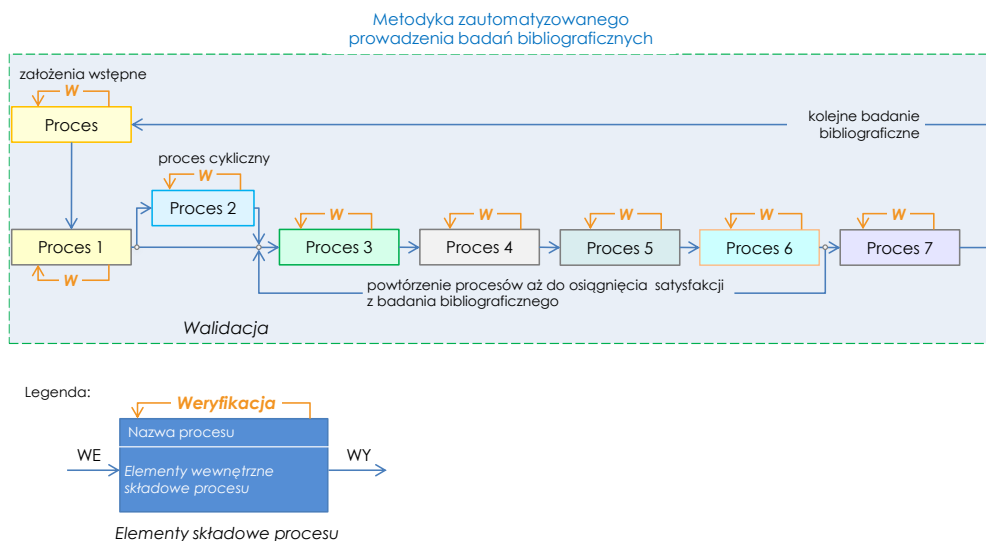
Rezultatem procesu oceny wyników z konsumpcją zebranych informacji jest: cytowanie dokumentu, włączenie danych z metryki cyfrowej do dokumentu, pozycji w bibliografii, utworzenie bibliografii dla nowego osiągnięcia twórczego. W tym procesie użytkownik na przemian dokonuje podglądu dokumentu z możliwością dzielenia się wiedzą z innymi specjalistami. Po konsumpcji wiedzy w bieżącym dokumencie użytkownik podejmuje decyzję czy przeszukuje następny dokument z listy wyników a po zakończeniu listy podejmuje decyzję o zakończeniu przeszukania. Kolejnym etapem jest proces weryfikacji i walidacji badań bibliograficznych.

1.7. Proces weryfikacji i walidacji badań bibliograficznych

Proces weryfikacji i walidacji (W&W) badań bibliograficznych jest końcowym procesem prowadzonym na etapie badań bibliograficznych. Procesy zostały przyporządkowane do jednego procesu, jednak metodycznie są to procesy dokonywane w różnych momentach czasowych. Pojęcia weryfikacji i walidacji można wyrazić poprzez następujące pytania:

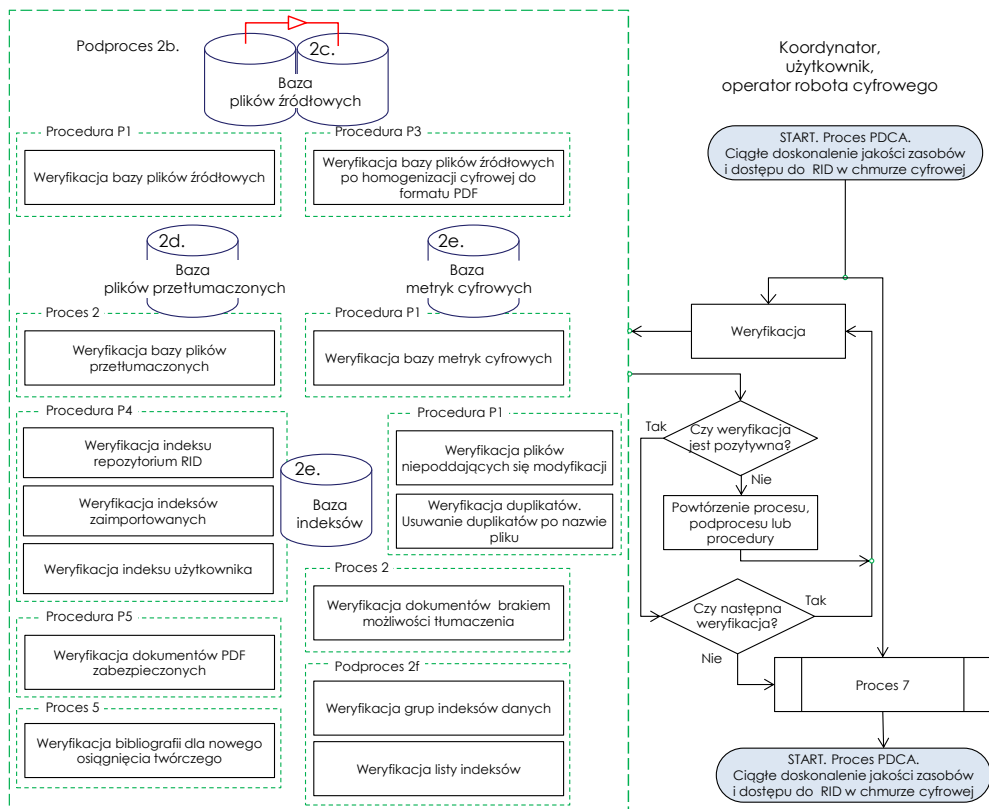
1. Weryfikacja – czy prawidłowo wykonywane są elementy składowe procesu i zależności pomiędzy procesami w metodyce zautomatyzowanego prowadzenia badań bibliograficznych?
2. Walidacja – czy prawidłowo została zrealizowana metodyka zautomatyzowanego prowadzenia badań bibliograficznych?

Na rysunku 1.36 przedstawiono w sposób graficzny procesy weryfikacji, które dotyczą konkretnych procesów oraz zależności pomiędzy procesami (dopasowania WY–WE). Proces walidacji obejmuje współzależność pomiędzy procesami (relacje), sposób zestawienia wszystkich procesów w sieć zależnych procesów (kolejność). Syntetycznie ujmując proces weryfikacji obejmuje pojedyncze procesy, natomiast walidacja obejmuje w sposób systemowy wszystkie procesy w przyjętej metodyce.



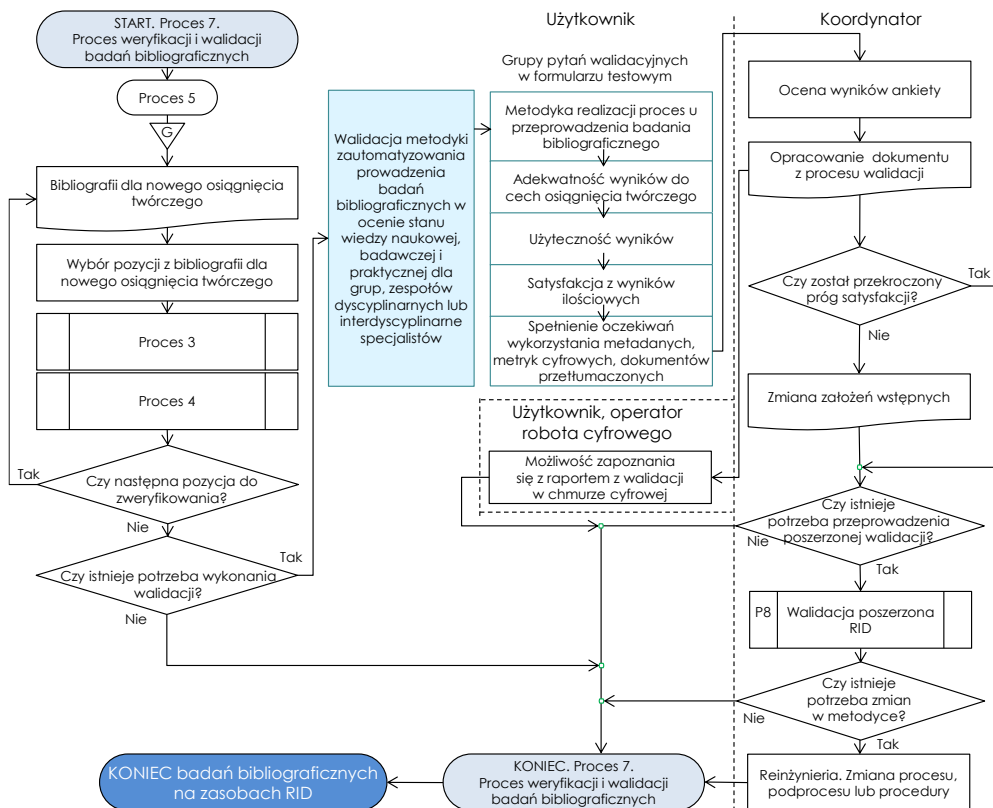
Rys. 1.36. Proces weryfikacji i walidacji w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Weryfikacja i walidacja są procesami przynależnymi do procesu zarządzania doskonalenia jakości zasobów i dostępu do RID w chmurze cyfrowej, jak również składową jakości w zarządzaniu wiedzą w organizacji. Procesy W&W służą do ustalenia, czy wykorzystane kluczowe elementy składowe procesów: nazwy procesów, wejścia i wyjścia, cele, procesy powiązane (zależne), zadania, działania, podprocesy, wskaźniki, mierniki procesów, dokumenty regulujące procesy, parametry ograniczające a także wykorzystywane zasoby, w tym aplikacje informatyczne są zgodne z wymaganiami użytkownika oraz czy produkty procesów spełniają przeznaczenie i potrzeby użytkownika. Procesy W&W powinny zapewnić obiektywną ocenę kluczowych elementów składowych i ich powiązania w ocenie – czy informacje pozyskane na podstawie danych są adekwatne, aktualne, poprawne, kompletne, spójne, spełniają założenia, są zgodne z potrzebami i oczekiwaniami użytkownika. W tym procesie może być zastosowany klasyczny cykl doskonalenia W. E. Deminga “planuj – wykonaj – sprawdź – działaj” (PDCA). Procesy weryfikacji są wykorzystywane w metodyce we wszystkich procesach i wytwarzanych produktach cyfrowych, w tym: bazy plików źródłowych, bazy plików przetłumaczonych, bazy metryk cyfrowych, bazy indeksów. Algorytm procesu W&W jest oznaczony jako proces klasycznego cyklu ciągłego doskonalenia PDCA (rys. 1.37).

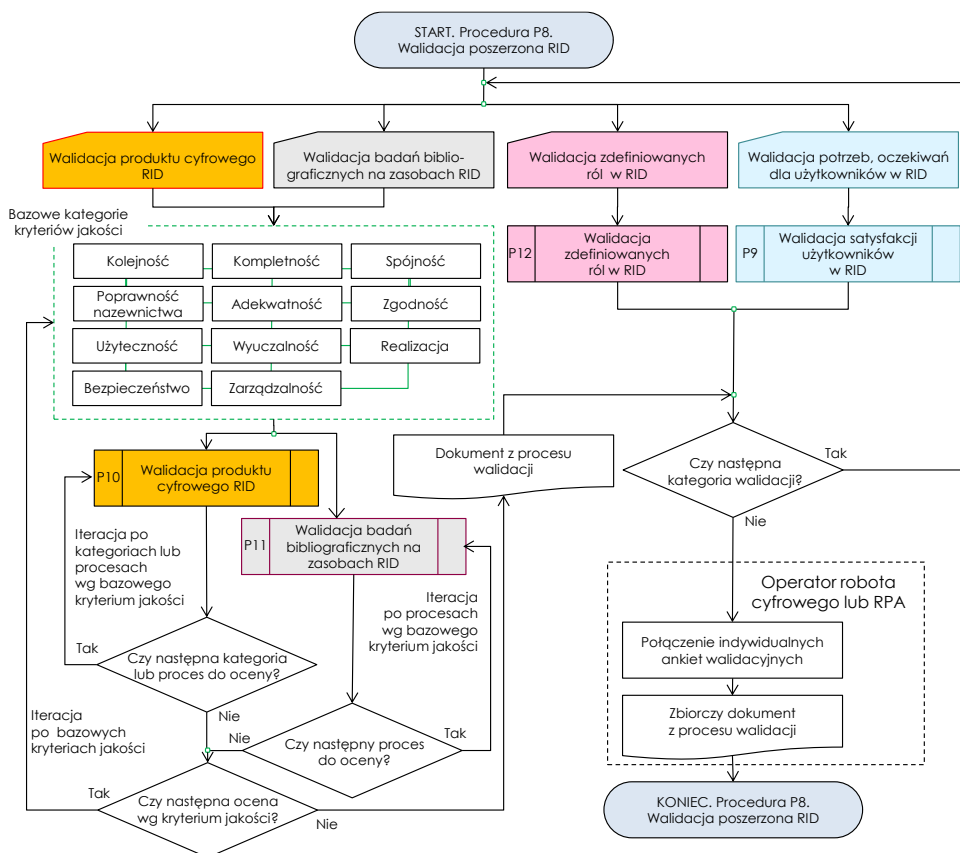


Rys. 1.37. Proces weryfikacji w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Algorytm procesu weryfikacji i walidacji badań bibliograficznych został zaprezentowany na rysunkach 1.38, w którym wykorzystano procedurę P8. *Walidacja poszerzona RID* (rys. 1.39).



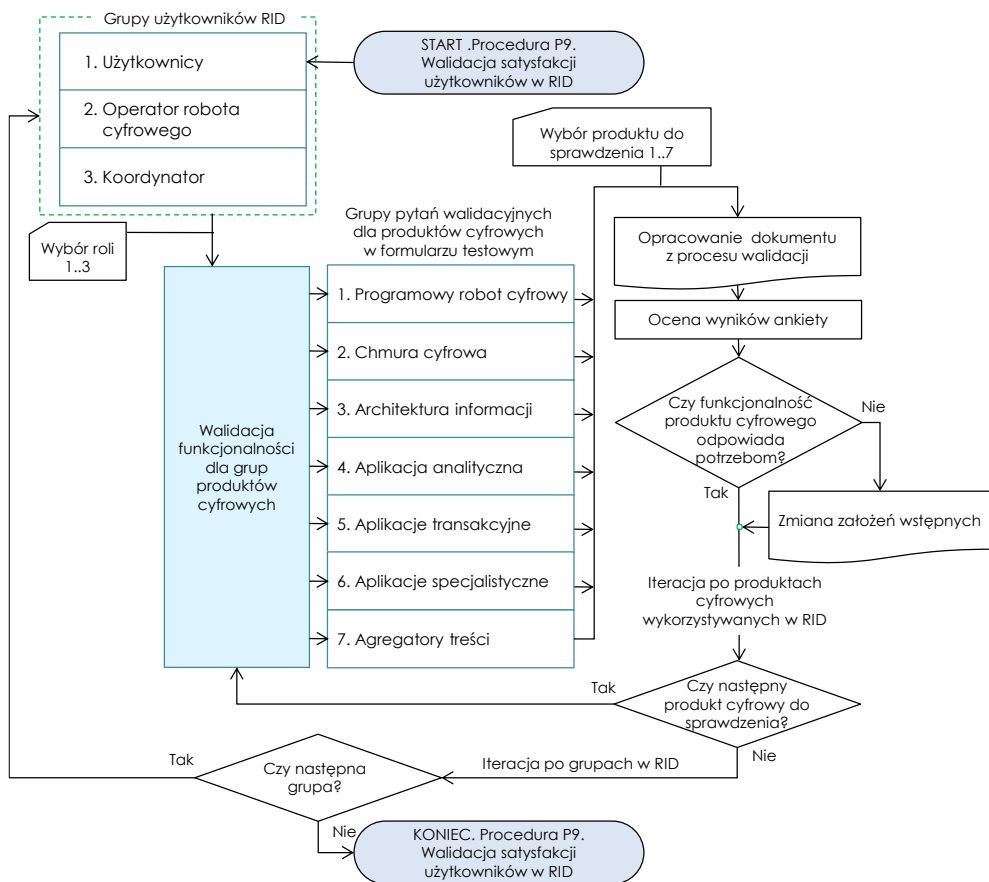
Rys. 1.38. Proces weryfikacji i walidacji badań bibliograficznych



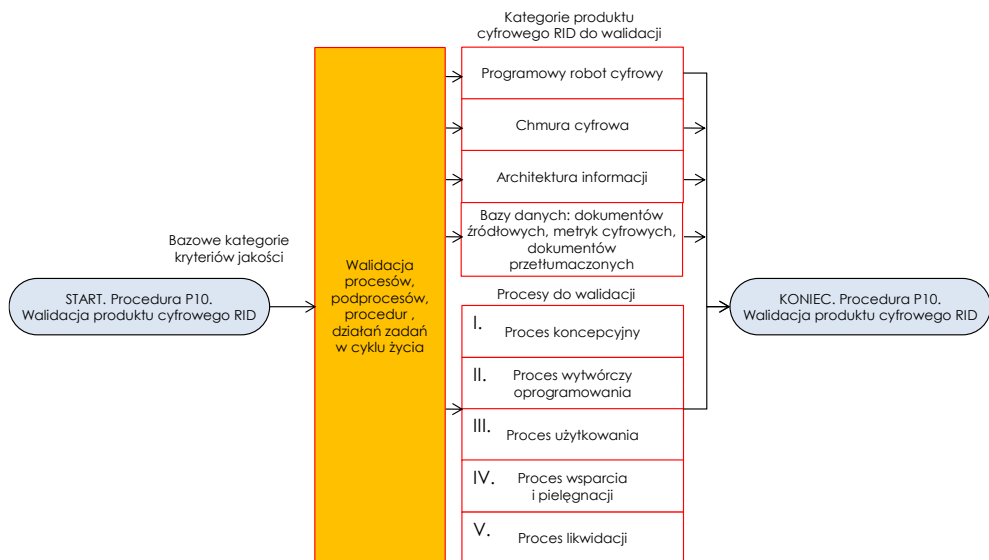
Rys. 1.39. Algorytm procedury P8. Walidacja poszerzona RID

Walidacja poszerzona dotyczy czterech istotnych problematyk w metodyce zautomatyzowanego prowadzenia badań bibliograficznych:

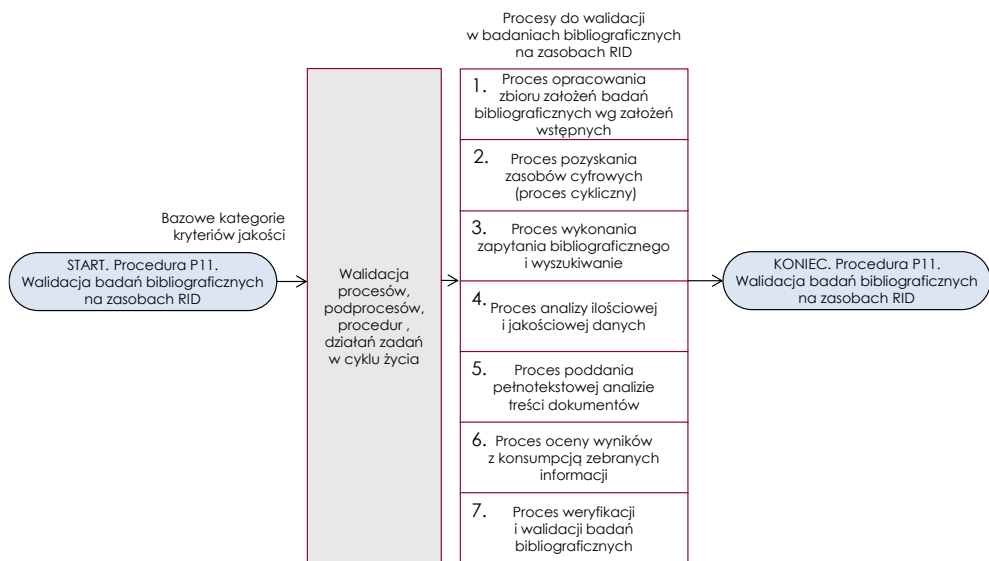
1. Walidacja potrzeb, oczekiwań dla użytkowników w RID (procedura P9. *Walidacja satysfakcji użytkowników w RID*) (rys. 1.40).
2. Walidacja produktu cyfrowego RID (procedura P10. *Walidacja produktu cyfrowego RID*) (rys. 1.41).
3. Walidacja badań bibliograficznych na zasobach RID (procedura P11. *Walidacja badań bibliograficznych na zasobach RID*) (rys. 1.42).
4. Walidacja zdefiniowanych ról w RID (procedura P12. *Walidacja zdefiniowanych ról w RID*) (rys. 1.43).



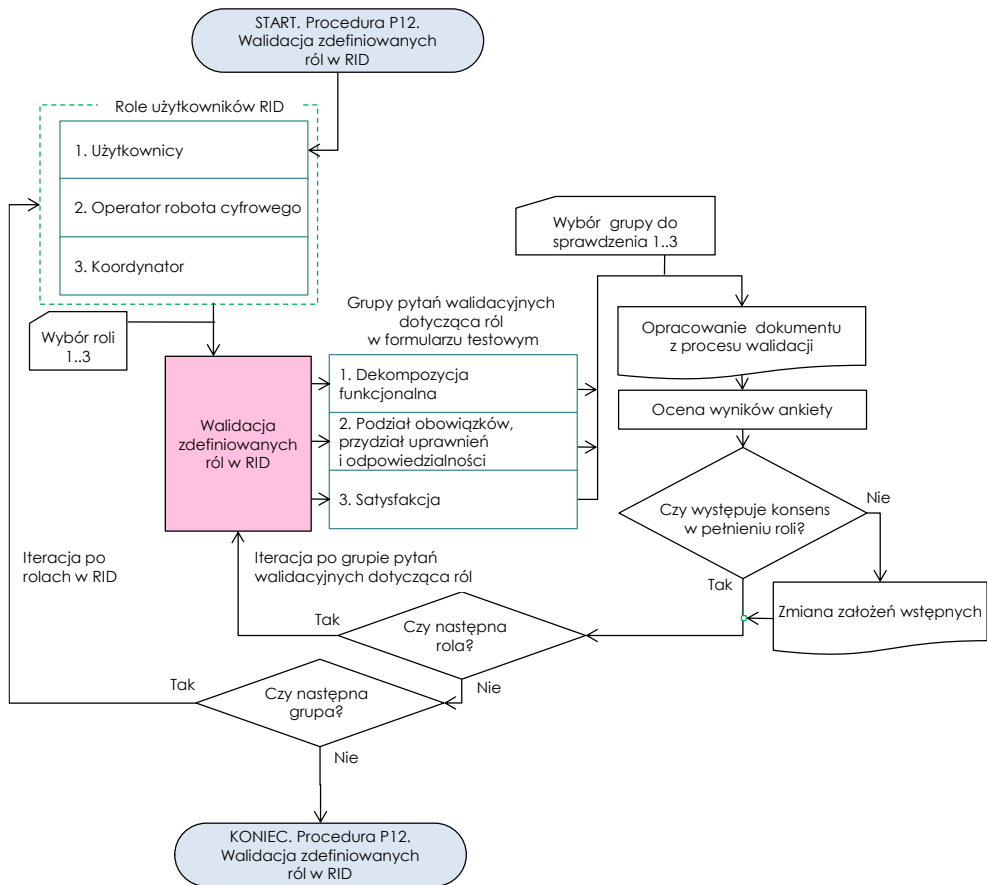
Rys. 1.40. Algorytm procedury P9. Walidacja satysfakcji użytkowników w RID



Rys. 1.41. Algorytm procedury P10. Walidacja produktu cyfrowego RID



Rys. 1.42. Algorytm procedury P11. Walidacja badań bibliograficznych na zasobach RID



Rys. 1.43. Algorytm procedury P12. Walidacja zdefiniowanych ról w RID

W rozdziale pierwszym w warstwie opisowej i ilustracyjnej został przedstawiony cykl życia badań bibliograficznych w cyklu życia RID. W następnym rozdziale zostaną zaprezentowane zagadnienia dotyczące cyklu życia Repozytorium Instytucjonalnego Danych, zarządzanie ciągłością działania oraz zarządzanie jakością produktów cyfrowych, danych i e-usługi w RID.

2. Zarządcze i organizacyjne aspekty wdrażania RID

2.1. Cykl życia produktu cyfrowego RID

Repozytorium Instytucjonalne Danych jest produktem cyfrowym, informatycznym i jego cykl życia wkomponowuje się w cykl życia systemu informatycznego. Cykl życia Repozytorium Instytucjonalnego Danych to zbiór procesów składających się z ciągu wzajemnie spójnych etapów, poczynając od momentu podjęcia decyzji o utworzeniu systemu do jego fizycznego usunięcia z zasobów organizacji, która miała prawa użytkownika. Cykl życia oprogramowania RID wpisuje się w klasyczny sześćoetapowy cykl życia oprogramowania⁴²:

1. Koncepcji – identyfikacja potrzeb interesariuszy, generowania koncepcji, zaproponowanie skutecznych rozwiązań.
2. Wytwarzania – opracowanie systemowych wymagań, tworzenie opisu rozwiązań, budowanie systemu, weryfikacja i walidacja systemu.
3. Produkcji – budowa systemu, weryfikacje i testy systemu.
4. Użytkowanie – praca z systemem zaspokajająca potrzeby użytkowników.
5. Wsparcie – zapewnienie stabilności i trwałości systemu.
6. Likwidacja – tworzenie kopii, archiwizacja, usunięcie fizyczne systemu.

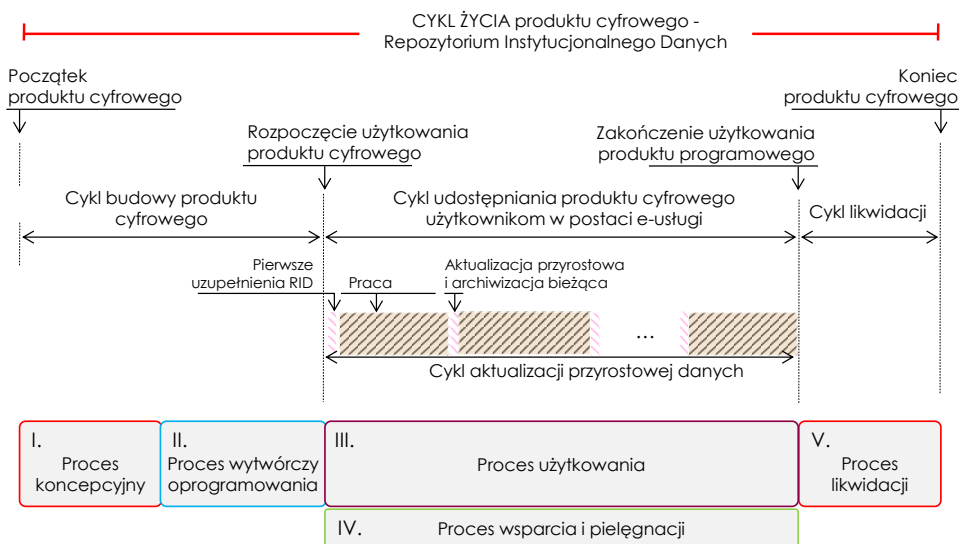
Zastosowanie technologii chmur cyfrowych umożliwia: a) uproszczenie procesu wytwórczego, b) tworzy model przechowywania, przetwarzania danych cyfrowych z zapewnieniem bezpieczeństwa, c) umożliwia zarządzanie użytkownikami i wymianę danych pomiędzy nimi, d) realizuje etap produkcji i częściowo wsparcia w zakresie infrastruktury oraz proces wdrożenia. Wymienione właściwości technologii chmur cyfrowych powoduje redukcję działań ograniczających się wyboru konkretnej platformy chmurowej, utworzenia architektury informacji, skonfigurowania, utworzenie kont użytkownikom i przyporządkowanie uprawnień oraz przetestowania funkcjonalności. W tym konkretnym zastosowaniu technologia chmur spełnia wszystkie wymagania stawiane systemowi informatycznemu w zakresie potrzeb użytkowników, obiegu informacji, udostępniania danych z zapewnieniem bezpieczeństwa informacji, co oznacza ochronę danych przed nieautoryzowanym dostępem do danych,

⁴² INCOSE 2010 *Systems Engineering Handbook* (SEBoK) v.3. s. 3.4/20 oraz ISO/IEC 15288: 2008 *Systems and software engineering–System life cycle processes*.

wykorzystaniem, ujawnianiem, zakłócaniem, zmianą lub zniszczeniem struktury i danych. W przypadku celowego lub przypadkowego usunięcia danych istnieje możliwość ich odzyskania. Również możliwość zakładania kont oraz zarządzanie uprawnieniami, tworzenia statystyk wykorzystania z danych to obecnie standardowe funkcjonalności chmur publicznych. Zastosowanie chmur cyfrowych powoduje, że klasyczny sześćoetapowy cykl życia oprogramowania zostaje zredukowany do pięciu procesów (etapów) w cyklu życia RID:

- I. Proces koncepcyjny.
- II. Proces wytwórczy oprogramowania.
- III. Proces użytkowania.
- IV. Proces wsparcia i pielęgnacji (proces równoległy z użytkowaniem).
- V. Proces likwidacji.

Pięcioetapowy cykl życia produktu cyfrowego ze zdefiniowanymi stanami RID został zaprezentowany na rysunku 2.1.



Rys. 2.1. Stany i procesy w cyklu życia produktu cyfrowego–RID

Cykl budowy Repozytorium Instytucjonalnego Danych zawiera dwa procesy: *I. Proces koncepcyjny*, *II. Proces wytwórczy oprogramowania*. Proces koncepcyjny w oparciu o założenia wstępne dotyczy opracowania wymagań technicznych i funkcjonalnych dla chmury cyfrowej, zgodnie z potrzebami użytkowników oraz modelem prawnym i finansowym udostępnienia zasobów. Proces wytwórczy oprogramowania ograniczył się do wybrania i zainstalowania konkretnego rozwiązania chmury cyfrowej, utworzenia architektury informacji,

utworzenia kont użytkownikom, przydzielenie uprawnień, skonfigurowania parametrów chmury obliczeniowej i przetestowania chmury cyfrowej. Proces wytwórczy obejmuje również działania związane z robotyzacją pierwszego pozyskiwania zasobów, a w procesie użytkowania do aktualizacji przyrostowej oraz innych procesów pomocniczych. Cykl życia produktu cyfrowego – programowalnego robota cyfrowego⁴³ RID obejmuje sześć procesów:

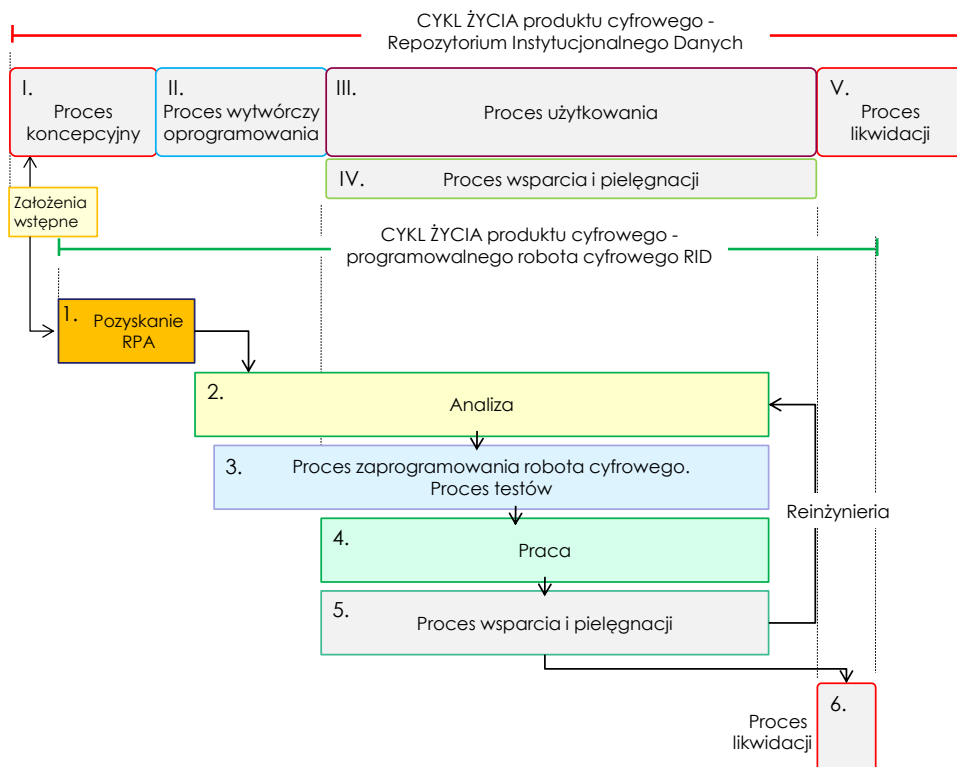
1. Proces pozyskanie technologii RPA. Wybór dostawcy technologii RPA wg założeń wstępnych może uwzględniać⁴⁴: a) Skalowalność rozwiązania, możliwości, cena w oparciu o cyfrową wydajność, b) Analitykę i otwartość platformy, c) Wykorzystanie pulpitu i ocena procesów, d) Wsparcie człowieka. Na przykład w raporcie *The Forrester Wave™: Robotic Process Automation* poddano analizie 15 dostawców RPA w 25 kryteriach pogrupowanych w trzy kategorie: 1. Aktualna oferta, 2. Strategia, 3. Obecność na rynku. W zautomatyzowaniu procesów bibliograficznych można zastosować alternatywne, różne strategie: zakupić licencję, wykorzystać rozwiązania niekomercyjne lub na licencji *Open Source*, wykorzystać bezpłatne wersje próbne lub specjalnego wykorzystania, skorzystać z ogólnodostępnych lub licencjonowanych pakietów biurowych zawierających możliwość programowania w środowisku programistycznym.
2. Proces analizy. Proces dotyczy wykonania dokumentacji zawierającej wszystkie wykorzystywane źródła i zbiorów danych od poszczególnych dostawców danych, ze sposobami dostępu do nich, z podziałem na grupy aktywów intelektualnych z metrykami cyfrowymi (VIII grup). Zbiór indywidualnych procedur dostępowych i pozyskania zasobów oraz procedur związanych z formularzami testowymi w procesie walidacji metodyki tworzy komplet dokumentacji i zarazem stanowi podstawę do procesu programowania robota cyfrowego.

⁴³ Jimenez-Ramirez A., Reijers H.A., Barba I., Del Valle C. (2019) *A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle*. In: Giorgini P., Weber B. (eds) *Advanced Information Systems Engineering*. CAiSE 2019. Lecture Notes in Computer Science, vol 11483. Springer, Cham.

⁴⁴ *The Forrester Wave™: Robotic Process Automation*, Q4 2019, October 23, 2019.

4. Proces zaprogramowania robota cyfrowego, proces testów. Tworzenie aplikacji RPA na wybranej platformie to proces odwzorowania biznesowych przypadków użycia, reguł biznesowych dla różnych rodzajów danych z opracowanej dokumentacji, zaprojektowanie scenariuszy awarii, opracowania koncepcji scenariusza z jego uruchomieniem oraz przetestowania dla sprawdzania prawidłowości działania. Proces zaprogramowania i testowania następuje wg określonej metodyki stworzonej przez twórców robota (platformy technologicznej). Pozytywne testy oprogramowania umożliwiają pierwsze uzupełnienie danymi RID przez robota.
5. Praca – implementacja oprogramowania robota cyfrowego. Proces dotyczy pierwszego uruchomienia dla uzupełnienia RID danymi, aktualizacji przyrostowej danych w procesie użytkowania RID oraz wykorzystania w zbieraniu danych formularzowych w procesie walidacji poszerzonej RID.
6. Proces wsparcia i pielęgnacji. Proces wsparcia dotyczy operatora robota cyfrowego przez stronę dostawcy platformy RPA. Zaplanowane działania w pielęgnacji ściśle związane są z procesem wsparcia i pielęgnacji w cyklu życia produktu cyfrowego – Repozytorium Instytucjonalnego Danych.
7. Proces likwidacji. Proces usunięcia oprogramowania polega na zarchiwizowaniu oprogramowania robota z towarzyszącą dokumentacją oraz fizycznym usunięciu platformy z zasobów cyfrowych RID. Zakłada się, że likwidacja oprogramowania robota cyfrowego powiązana jest z likwidacją programową Repozytorium Instytucjonalnego Danych.

Cykl życia programowalnego robota cyfrowego RID zawiera się w cyklu życia Repozytorium Instytucjonalnego Danych. Może wystąpić wielokrotna migracja oprogramowania z starszej wersji na nowszą lub zmiana platformy spowodowana innymi czynnikami, np. zmiana platformy na bardziej wydajną, korzystniejszą ekonomicznie. Cykl życia produktu cyfrowego – programowalnego robota cyfrowego RID w cyklu życia Repozytorium Instytucjonalnego Danych bez procesu migracji został zaprezentowany na rysunku 2.2.



Rys. 2.2. Cykl życia produktu cyfrowego–programowalnego robota cyfrowego RID w cyklu życia Repozytorium Instytucjonalnego Danych

Przedstawiony cykl życia programowalnego robota cyfrowego jest poszerzony w stosunku do etapów opisanych w literaturze z zakresu RPA, gdzie prezentuje się typowy czteroprosowy (fazowy) cykl życia⁴⁵ lub sześcioprosowy⁴⁶: 1) Analiza⁴⁷, 2) Projektowanie, 3) Budowa, 4) Wdrożenie, 5) Kontrola i monitorowanie, 6) Ewaluacja i wydajność. Nie uwzględnia się procesu decyzyjnego związanego z pozyskaniem robota cyfrowego oraz procesu likwidacji.

⁴⁵ Del Valle, C. (2019). A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle. In *Advanced Information Systems Engineering: 31st International Conference, CAiSE 2019, Rome, Italy, June 3–7, 2019, Proceedings* (Vol. 11483, p. 446). Springer.

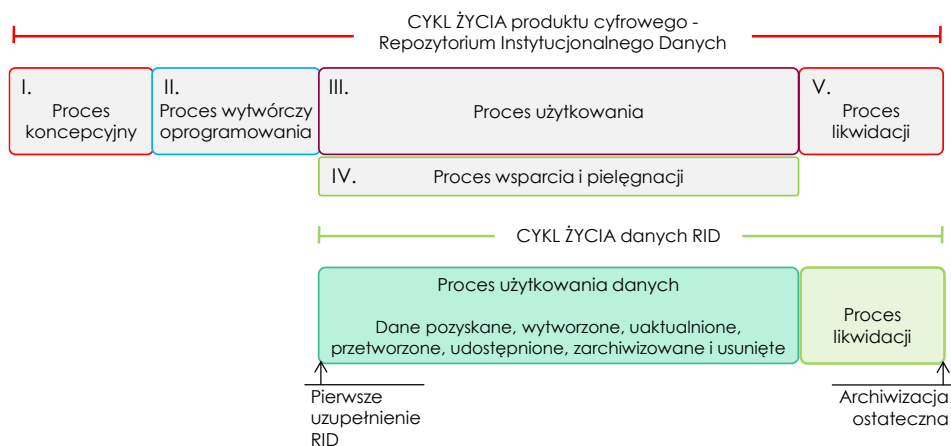
⁴⁶ *Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study*, 10.1109/ACCESS.2020.2974934, s. 39126.

⁴⁷ Enríquez, J.G., Jiménez-Ramírez, A., Dominguez-Mayo, F.J., & García-García, J.A. (2020). *Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study*. IEEE Access, 8, 39113–39129.

Cykl pierwszego uzupełnienia repozytorium danymi przez programowalnego robota cyfrowego wyznacza rozpoczęcie użytkowania produktu cyfrowego RID. Cykl pierwszego uzupełnienia RID obejmuje:

- Wykonania zapytań bibliograficznych do źródeł i zbiorów danych wg założeń.
- Pozyskanie zasobów.
- Modyfikację pozyskanych plików.
- Utworzenie raportu z pozyskania zasobów.
- Archiwizację bieżącą zgromadzonych zasobów RID.

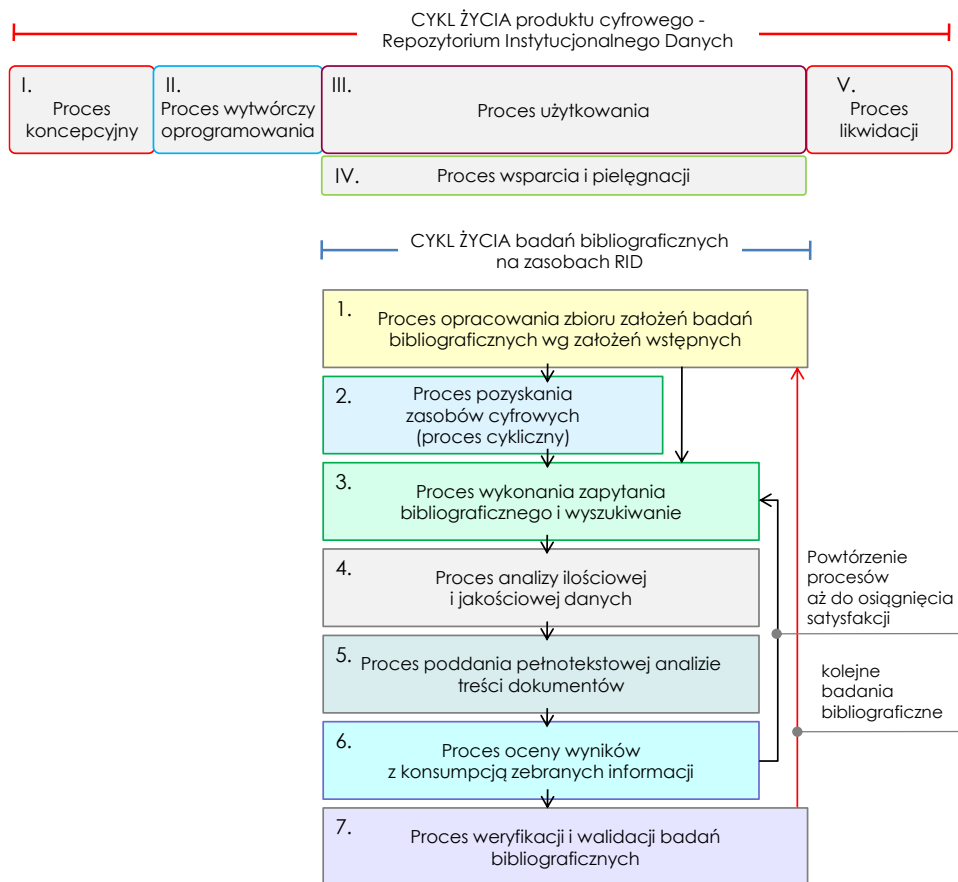
W momencie rozpoczęcia procesu użytkowania również rozpoczyna się cykl życia danych RID (rys. 2.3).



Rys. 2.3. Cykl życia danych RID w cyklu życia Repozytorium Instytucjonalnego Danych

Proces użytkowania danych dotyczy danych pozyskanych, utworzonych, uaktualnionych, przetworzonych, udostępnionych, zarchiwizowanych i usuniętych. Pozyskane lub wytworzone dane były gromadzone, archiwizowane oraz poddane obróbce przez uaktualnienie, przetworzenie, udostępnianie użytkownikom z wykorzystaniem aplikacji analitycznych (wariant I) lub czytniki dokumentów PDF (wariant II). Oddzielenie zasobów źródłowych, przetworzonych i przechowywanych w chmurze cyfrowej, od aplikacji ich obsługujących pozwala na zarządzanie zbiorem dokumentów PDF oraz umożliwia swobodne wykorzystanie zgromadzonych danych przez użytkowników. Użytkownicy korzystają z RID na zasadzie czarnej skrzynki, mają swobodny wybór oprogramowania do wyszukiwania informacji czy też zarządzania bibliografią. Jedynie rozwiązania technologiczne chmury cyfrowej określają granice i możliwości użytkowania, komunikowania się z innymi użytkownikami.

Proces użytkowania RID to wykorzystanie metodyki badań bibliograficznych przez grupy, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów w prowadzeniu badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej i praktycznej. Cykl życia badań bibliograficznych na zasobach RID w cyklu życia Repozytorium Instytucjonalnego Danych został zaprezentowany na rysunku 2.4.

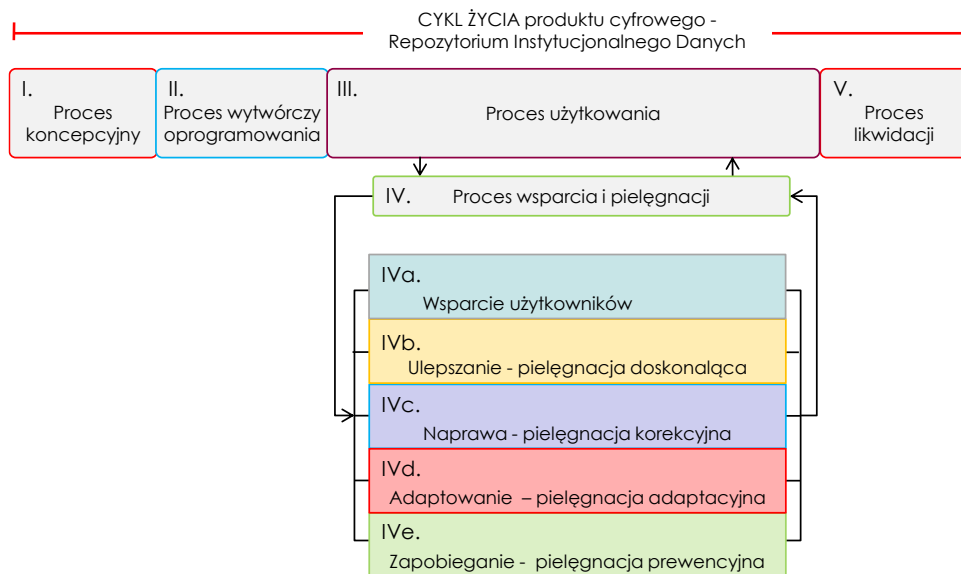


Rys. 2.4. Cykl życia badań bibliograficznych na zasobach RID w cyklu życia Repozytorium Instytucjonalnego Danych

W ramach użytkowania Repozytorium Instytucjonalnego Danych były wykorzystane gotowe programy komputerowe, m.in.: aplikacje analityczne, transakcyjne, specjalistyczne, agregatory treści. Wymienione produkty cyfrowe są aplikacjami gotowymi, nie wymagają procesu ich tworzenia a jedynie skonfigurowania lub w przypadku aplikacji komercyjnych wykupienia licencji i sparametryzowania. Ze względu na funkcjonalność i specjalistyczne przeznaczenie

czenia, niektóre aplikacje wymagają licencjonowania z opłatnością za jego użytkowanie. W procesie użytkowania przewiduje się również archiwizację bieżąca zasobów RID, która jest planowana po zakończeniu procesu aktualizacji przyrostowej. Archiwizacja bieżąca wiąże się z bezpieczeństwem komputerowym i odpowiada potrzebom bezpieczeństwa użytkownika. Pozwala na awaryjne zapoznanie się z zawartością dokumentów i innych plików przez użytkowników, które zostały zapisane w określonym momencie czasowym. Umożliwia również odzyskanie danych a w skrajnym przypadku skopiowanie i przywrócenie całego zasobu lub grupy plików.

Proces użytkowania jest buforowanym procesem wsparcia udzielanym użytkownikom w procesie użytkowania przez operatora robota cyfrowego i koordynatora (proces IV *Proces wsparcia i pielęgnacji* cyklu życia RID, IVa. *Wsparcie użytkowników*) (rys. 2.5).



Rys. 2.5. Proces wsparcia i pielęgnacji w cyklu życia RID

Wdrażanie i wsparcie użytkowników (*help desk*) lub asysta techniczna może dotyczyć problemów z wykorzystania zbioru funkcjonalności dostępnych aplikacji lub chmury cyfrowej, z nieprawidłowym wyświetlaniem dokumentów PDF przez ich uszkodzenie lub znaczną modyfikacją, czy też problemami z indeksami użytkownika i sparametryzowaniem systemu wyszukiwawczego.

W zarządzaniu specjalistami w grupach i zespołach badawczych można wyróżnić cztery kanały dla świadczenia wsparcia synchronicznego i asynchronicznego użytkownikom:

1. E-mail.
2. Telefon, czat z koordynatorem lub operatorem robota cyfrowego, programy do wideokonferencji (*Teams* na platformie *Office 365*, *Google Class Room*).
3. Zgłoszenia on-line bezpośrednio z wykorzystaniem chmury cyfrowej.
4. Funkcje pomocy podnoszące lub rozszerzające kompetencje, tj. transmisje szkoleniowe on-line, filmy z rozwiązaniami najczęstszych problemów, e-kursy, dokumenty elektroniczne zawierające najczęściej zadawane pytania (FAQ).

Drugim procesem buforowanym jest pielęgnacja oprogramowania (proces IV *Proces wsparcia i pielęgnacji* cyklu życia RID, cztery procesy pielęgnacji IVb–IVe), która dotyczy ośmiu produktów cyfrowych: Repozytorium Instytucjonalnego Danych, programowalnego robota cyfrowego RID, chmury cyfrowej, architektury informacji, aplikacji analitycznych, aplikacji transakcyjnych, aplikacji specjalistycznych, agregatorów treści. Pielęgnacja produktów cyfrowych dokonywana jest przez użytkowników, koordynatora, operatora robota cyfrowego i może obejmować cztery kategorie doskonalenia:

1. Ulepszanie – pielęgnacja doskonaląca (proces IVb). Zmiany polegają na ulepszaniu produktów cyfrowych i zasobów, np. zmiana funkcjonalności przez aktualizację wykorzystywanych aplikacji, wymiana aplikacji na alternatywne, zwiększenie użyteczności lub innych cech jakościowych produktów cyfrowych, które bezpośrednio operują na zasobach RID lub dotyczą funkcjonalności chmury cyfrowej.
2. Naprawa – pielęgnacja korekcyjna (proces IVc). Naprawa to zmiany wprowadzone w oprogramowaniu robota przez operatora robota polegające na usunięciu nieprecyzyjności. Działania naprawcze pozwalają na naprawę wykrytych problemów, aby produkt cyfrowy pełnił funkcje, do czego został stworzony i spełniał zakładane wymagania.
3. Adaptowanie – pielęgnacja adaptacyjna (proces IVd). Zmiany w zbiorach danych i dokumentów źródłowych (serwisach źródłowych), zmiana licencji, sposobu autoryzacji lub zmiany w aktywach intelektualnych, metryk cyfrowych, cyfrowych aktywach osobowych wymaga modyfikacji oprogramowania robota. Również zmiany w funkcjonalności chmury cyfrowej (autoryzacji, elektronicznego obiegu dokumentów, ochrona danych) wymusza

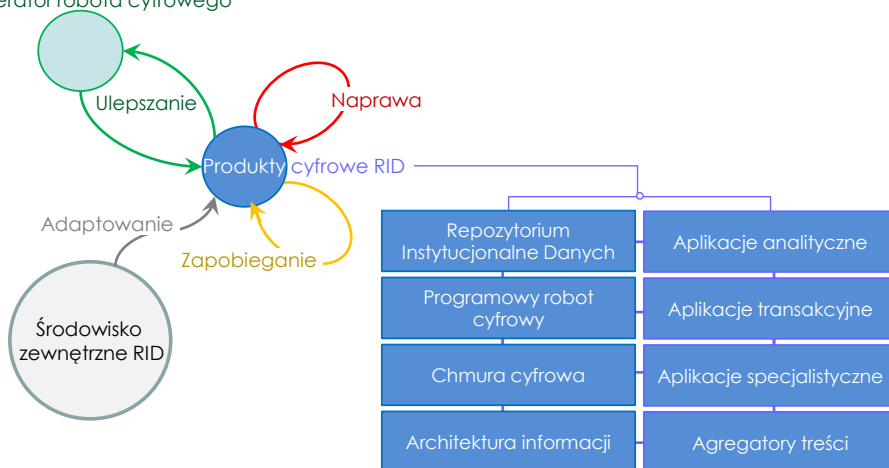
działania adaptacyjne. Ogólnie ujmując, modyfikacja adaptacyjna ma na celu dostosować wykorzystywane produkty cyfrowe, aby stały się użyteczne, spełniające wymagania w zmienionym środowisku działania po stronie użytkownika (*front office*) jak i stronie administracyjnej (*back office*), tj. koordynatora, operatora robota cyfrowego.

4. Zapobieganie – pielęgnacja prewencyjna (proces IVE). Zmiany, które w przyszłości poprawią utrzymanie i niezawodność systemu. W przeciwieństwie do wyżej wymienionych kategorii doskonalenia, zmiany zapobiegawcze w sposób przyszłościowy ułatwią proces zmian w środowisku wewnętrznym lub zewnętrznym. Przykładem pielęgnacji prewencyjnej są planowane i zapowiadane zmiany w funkcjonalności lub struktury w źródłowych serwisach internetowych lub planowane zmiany w wykorzystywanych aplikacjach.

W sposób graficzny kategorie pielęgnacji przedstawiono na rysunku 2.6.

Użytkownicy RID:

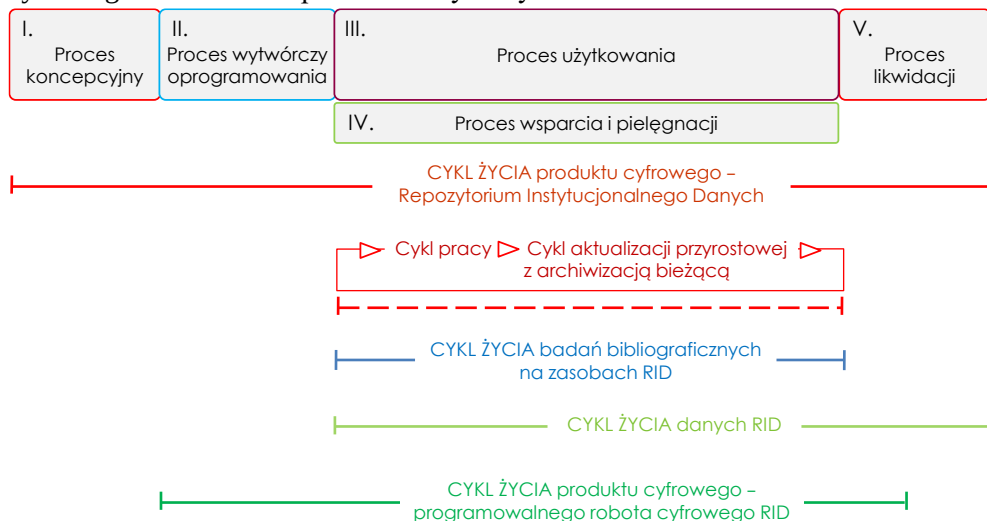
1. Użytkownicy
2. Koordynator
3. Operator robota cyfrowego



Rys. 2.6. Cztery kategorie doskonalenia produktów cyfrowych w procesie pielęgnacji w cyklu życia RID

Proces likwidacji w cyklu życia RID obejmuje dwa działania. Pierwsze to wykonanie raportu końcowego zawierającego spis produktów cyfrowych oraz danych i archiwizacja całej struktury informacji łącznie z danymi. Drugie działanie dotyczy usunięcia architektury informacji oraz usunięcie danych z chmury obliczeniowej.

W ujęciu syntetycznym pięcioprosowy cykl życia produktu cyfrowego – Repozytorium Instytucjonalnego Danych z cyklem badań bibliograficznych na zasobach RID, cyklem życia danych RID programowalnego robota cyfrowego RID został zaprezentowany na rysunku 2.7.



Rys. 2.7. Syntetyczne ujęcie cykli życia w cyklu życia produktu cyfrowego – Repozytorium Instytucjonalnego Danych

W następnym podrozdziale zostanie zaprezentowane zagadnienie zarządzania zasobami ludzkimi, rzeczowymi i informacyjnymi w procesie użytkowania Repozytorium Instytucjonalnego Danych.

2.2. Zarządzanie zasobami RID

Według założeń twórcy RID wprowadzający do organizacji innowację procesową, produktową i organizacyjną został opracowany cykl życia RID, w którym zaplanowano pięć procesów: I. Koncepcyjny, II. Wytwórczy oprogramowania, III. Użytkowania, IV. Wsparcia, V. Likwidacji. Poszczególne procesy zostały już opisane w rozdziale 2.1. *Cykl życia Repozytorium Instytucjonalnego Danych*. Zarządzanie zasobami RID dotyczy zarządzania procesami podstawowymi i zarządczymi w organizacji, w tym w szczególności: zarządzanie wiedzą skodyfikowaną do postaci cyfrowej i produktami cyfrowymi oraz zasobami ludzkimi przez zdefiniowane role i relacje pomiędzy użytkownikami.

Zarządzanie wiedzą skodyfikowaną w postaci cyfrowej i produktami cyfrowymi w organizacji

W procesie użytkowania przyrostowej aktualizacji zasobów RID istotne jest założenie oddzielenia fizycznego i logicznego dwóch głównych grup działań. Pierwsza grupa działań dotyczy pozyskania i zarchiwizowania w chmurze cyfrowej w odpowiedniej architekturze informacji aktywów intelektualnych z metrykami cyfrowymi w szerokim zakresie (aktywa naukowe, badawcze, praktyczne), w którym wykorzystano automatyzację procesów:

- Pozyskiwanie zasobów (aktywa intelektualne, cyfrowe metryki) wg założeń bibliograficznych i zapis w chmurze obliczeniowej, usuwanie duplikatów wg nazwy plików, weryfikacja poprawności pozyskanych zasobów.
- Automatyczna aktualizacja przyrostowa zasobów (proces cykliczny).
- Translacja dokumentów na język polski dla zasobów niepolskojęzycznych.
- Sprawdzenie zabezpieczenia dokumentu PDF przed edycją lub zaznaczeniem treści.
- Umieszczenie metadanych pozyskanych z metryk cyfrowych w dokumentach PDF. Utworzenie opcjonalnie metryki cyfrowej dla dokumentu źródłowego w postaci stowarzyszonego dokumentu PDF.
- Sprawdzenie automatyczne występowania duplikatów dokumentów PDF wg zawartości i w przypadku wystąpienia usunięcie duplikatów.
- W procesie weryfikacji i walidacji poszerzonej RID połączenie indywidualnych ankiet walidacyjnych w zarządzaniu dotyczącym zapewnienia jakości.

Z punktu widzenia robotyzacji procesów wymienione działania w tej grupie dotyczą wykorzystania trzech produktów cyfrowych. Pierwszy produkt programowy to programowalny robot cyfrowy zastosowany w procesach, które może również wykonać człowiek, gdyż procesy te są powtarzalne oraz są oparte na regułach. Drugim i trzecim produktem programowym jest chmura cyfrowa i zorganizowana w niej architektura informacji, uzupełniona zbiorem dokumentów, pojęciowo określana w monografii bazą danych⁴⁸. Ujmując architektonicznie, w jej strukturze umiejscowiono: bazę danych plików źródłowych, bazę plików przetłumaczonych i bazę danych metryk cyfrowych. Do bazy da-

⁴⁸ CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022, <https://www.coretrustseal.org/why-certification/requirements/>, data dostępu: 12.01.2022 r.

nych nie zalicza się oprogramowania umożliwiającego operacje na zgromadzonych zasobach. Jest to wyłącznie zbiór aktywów intelektualnych z metrykami w postaci dokumentów PDF z określonych dziedzin i dyscyplin naukowych.

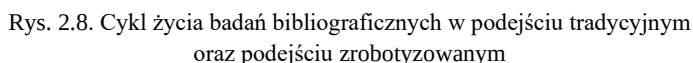
Druga grupa działań w procesie użytkowania dotyczy wyszukiwania i przetwarzania informacji na zgromadzonych dokumentach RID z wykorzystaniem specjalistycznych aplikacji, np. czytników dokumentów PDF lub oprogramowania, które pozwalają na zarządzanie zbiorem dokumentów PDF. Z punktu widzenia informatycznego to umożliwienie użytkownikom korzystania ze zbioru funkcjonalności różnych aplikacji z wykorzystaniem e-usługi.

Rozdzielenie logiczne dwóch grup działań powoduje, że zarządzanie zasobami ludzkimi w organizacji sprowadzone jest do pełnienia określonych ról: użytkownika, koordynatora, operatora robota cyfrowego. Użytkownicy korzystają z RID na zasadzie czarnej skrzynki z zasobami, mają swobodny wybór oprogramowania do wyszukiwania informacji czy też zarządzania bibliografią. Jedynie rozwiązania technologiczne chmury cyfrowej określają granice i możliwości użytkowania, komunikowania się w organizacji i dzielenia się wiedzą z innymi użytkownikami.

Użytkownicy mają określone potrzeby informacyjne oraz oczekiwania jakościowe co do których powinny być one spełnione podczas pracy ze zgromadzonymi zasobami. Eksploracja zasobów RID odbywa się w formie świadczenia e-usługi przez specjalistyczne aplikacje analityczne lub w sposób bezpośredni na zgromadzonych plikach przez czytniki dokumentów PDF z zapewnieniem oczekiwanej przez użytkownika funkcjonalności w spełnieniu potrzeb i oczekiwań. Funkcjonalność oprogramowania jest właściwością i oznacza zdolność produktu do zapewnienia funkcji odpowiadających ustalonym potrzebom informacyjnym, w kontekście jego użycia, zgodnie z wyspecyfikowanym zbiorem założeń użytkownika. Użytkownik i jego potrzeby informacyjne oraz wymagania, wiązka celów do osiągnięcia, kontekst użycia, funkcja elementarna, funkcjonalność to najważniejsze pojęcia w badaniach bibliograficznych. W tym kontekście zalecanym sposobem pracy z zasobami RID jest wykorzystanie aplikacji analitycznej, która posiada co najmniej następujący zbiór funkcji elementarnych:

- umożliwia indeksację różnego typu zasobów z analizą morfologiczną w kilku językach,
- udostępnianie indeksów przez serwer wyszukiwania zdalnego lub możliwość umieszczenia indeksów z chmurze cyfrowej z udostępnieniem innym użytkownikom,
- wyszukiwanie tekstowe z wykorzystaniem języka kwerend,

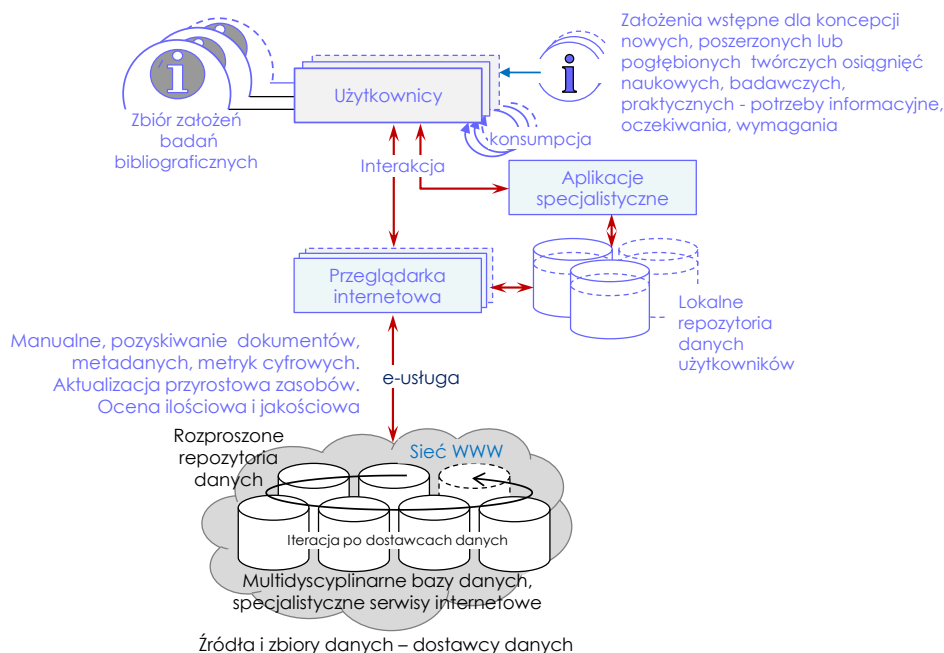
- Zasadniczo, korzystanie z RID nie zmienia metodyki badań bibliograficznych i tym samym cykl życia z niewielkimi usprawnieniami pozostał bez zmian. Proces archiwizowania badania bibliograficznego w uogólnionej metodyce badań bibliograficznych (proces 6) jest zbędny w opracowanej metodyce. Proces wykonania zapytania bibliograficznego, wyszukiwanie na źródłach i zbiorach danych w uogólnionej metodyce badań bibliograficznych (proces 2) dokonuje się kolejno po dostawcach danych, podczas gdy w opracowanej metodyce odbywa się na zasobach RID. Dlatego istotą zmiany jest sposób korzystania ze źródeł, zbiorów danych umożliwiając przede wszystkim zmniejszenie czasu – i pracochłonności zadań w cyklu życia badań bibliograficznych w opracowanej metodyce (rys. 2.8).



W podejściu tradycyjnym użytkownik korzysta z wielu rozproszonych źródeł i zbiorów danych. Tym samym dokonuje następujących działań:

- przeszukuje zawartość zbiorów wg różnych języków kwerend i ustawień,
- tworzy dla swoich celów lokalne repozytorium danych przez manualne pozyskiwanie dokumentów, metadanych, metryk cyfrowych za pomocą przeglądarki internetowej,
- korzysta z lokalnego repozytorium danych z wykorzystaniem aplikacji specjalistycznych,
- przeszukuje zawartość zbiorów za pomocą czytnika dokumentów PDF,
- zarządza zebranymi dokumentami, ewentualnie dokonuje archiwizacji i weryfikacji, walidacji zbiorów oraz spersonalizowanego procesu bibliograficznego.

W zakresie oceny jakościowej danych użytkownik korzysta z tych samych lub dodatkowo innych źródeł danych. Czynności są powtarzane, dopóki użytkownik nie osiągnie satysfakcji z badania bibliograficznego, czego wynikiem są cytowania (odwołania), pozycje bibliograficzne i bibliografia w osiągnięciu twórczym. Ostatnim możliwym etapem wykonanym przez użytkownika po zakończeniu badań jest archiwizacja danych wykorzystywanych w badaniu bibliograficznym. W przypadku kolejnych badań użytkownicy sami dokonują uaktualnień swoich repozytoriów danych i doskonalą swój warsztat pracy. Sposób, metody czy techniki badań bibliograficznych są zindywidualizowane, zależne od doświadczenia i e-kompetencji użytkownika. W przypadku zespołów lub grup specjalistów te same etapy i czynności są wielokrotnie wykonywane analogicznie lub w bardzo podobny sposób. Użytkownicy nie udostępniają swoich zasobów, każdy z nich posiada lokalne repozytorium danych, które jest unikalne i nieporównywalne a zgromadzone zasoby w grupie po połączeniu były w dużej części zduplikowane (rys. 2.9).



Rys. 2.9. Podejście tradycyjne w badaniach bibliograficznych

Zarządzanie zasobami ludzkimi w organizacji z wykorzystaniem zdefiniowanych ról

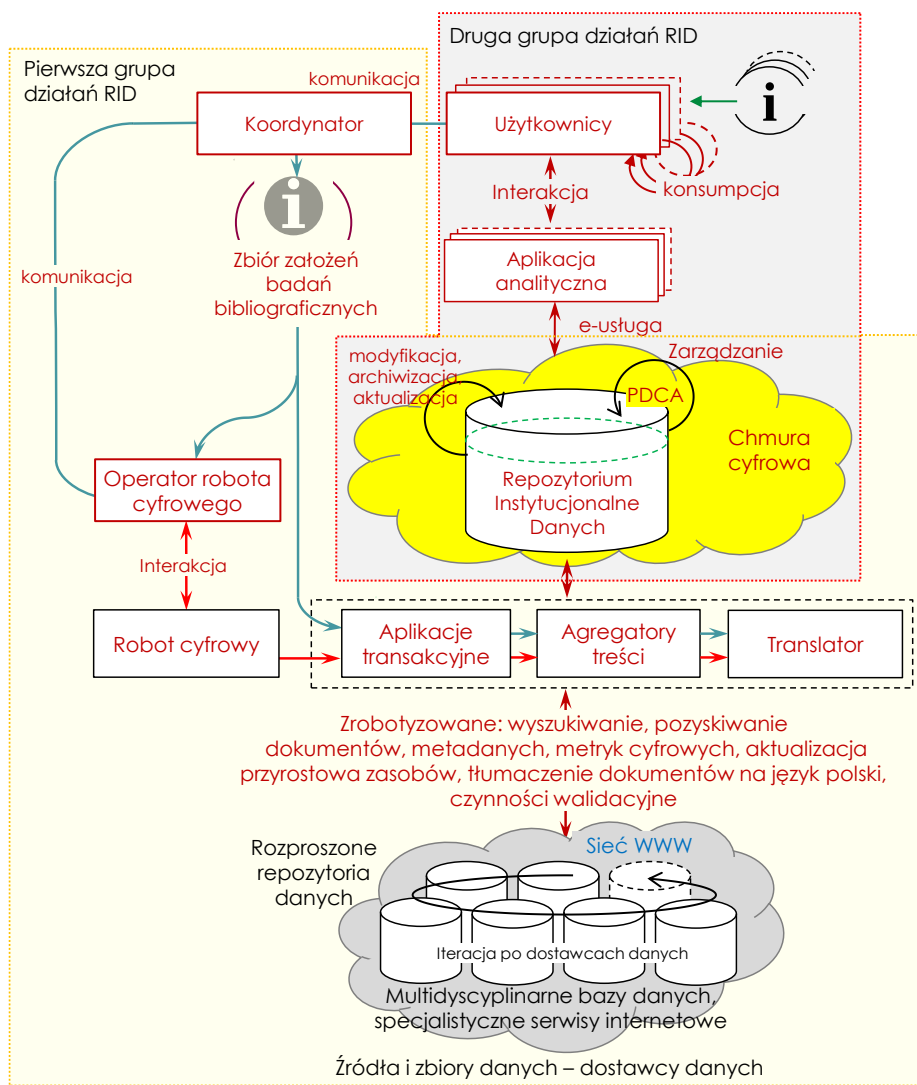
W podejściu zrobotyzowanym korzystanie ze zbiorów scentralizowanego zasobu w chmurze cyfrowej jest z założenia nastawione na wykorzystanie przez grupy, zespoły monodyscyplinarne lub interdyscyplinarne specjalistów. W zarządzaniu zasobami ludzkimi w organizacji dokonano podziału zadań przez wprowadzenie trzech ról użytkowników i określenie relacji pomiędzy użytkownikami.

Rola operatora robota cyfrowego polega na analizie, zaprogramowaniu, testowaniu, obsłudze i kontroli jego działania z procesami wsparcia i pielęgnacji. Pracochłonne, żmudne, powtarzające się procesy zostały przejęte przez programowalnego robota cyfrowego. Robot cyfrowy automatycznie pozyskuje dokumenty i metryki cyfrowe w oparciu o zdefiniowane grupy aktywów intelektualnych odciążając tym samym użytkowników od wielokrotnych czynności w różnych repozytoriach. Robot cyfrowy automatycznie sprawdza zawartość pliku dokumentu PDF występowania metadanych lub metryki cyfrowej dokumentu dla cytowania, a w przypadku jej braku wstawia do dokumentu

odpowiednie metadane. Kolejną czynnością wykonywaną automatycznie przez programowalnego robota cyfrowego jest tłumaczenie dokumentów z różnych języków na język polski. Po przetłumaczeniu plik jest odpowiednio nazwany i umieszczony w zdefiniowanym katalogu. Cały proces pozyskiwania, sprawdzania, translacji jest kontrolowalny oraz zarządzalny a składowane pliki umieszczane są w odpowiednich katalogach w chmurze cyfrowej. Również proces aktualizacji zasobów i indeksacji zbiorów odbywa się wg ściśle określonych zasad. Kolejnym zaprogramowanym procesem robotycznym jest opracowanie dokumentu z procesu walidacji przez połączenie indywidualnych ankiet walidacyjnych.

Użytkownik w roli koordynatora współpracuje z pozostałymi użytkownikami i jest odpowiedzialny m.in. za poprawność, aktualność, spójność dokumentów oraz za proces zarządzania ciągłością działania w chmurze cyfrowej. Wymagania kwalifikacyjne są najwyższe spośród innych użytkowników RID. Szczegółowe zadania oraz role osób w metodyce zautomatyzowanego prowadzenia badań bibliograficznych zostały omówione w podrozdziale 2.3. *Zarządzanie ciągłością działania RID* oraz w rozdziale 3. *Rekomendowane działania w zarządzaniu zasobami ludzkimi*.

Trzecia rola dotyczy użytkowników korzystających z zasobów zgromadzonych w RID. Użytkownik wykonuje zapytania bibliograficzne w zasobach RID z wykorzystaniem aplikacji analitycznej, ma możliwość podglądu dokumentu źródłowego oraz plików stowarzyszonych, cytuje dokument używając aplikacji wspomagającej cytowania lub kopiuje dane z metryczki dokumentu przy odwołaniu do dokumentu. Podczas korzystania z zasobów RID zaplanowano dwa warianty pracy użytkownika. Pierwszy wariant, który jest rekomendowany dotyczy korzystania z licencjonowanych, płatnych aplikacji analitycznych wykorzystujących indeksy. W drugim wariancie zaplanowano wykorzystanie ogólnodostępnych czytników plików PDF z funkcjonalnością indeksowania zasobów, wyszukiwania z zaawansowanymi filtrami. Podejście zautomatyzowane w badaniach bibliograficznych niezależnie od wariantu korzystania zostało zaprezentowane na rysunku 2.10.



Rys. 2.10. Podejście zautomatyzowane w badaniach bibliograficznych

Zarządzanie zasobami ludzkimi RID dotyczy procesów zarządczych w zakresie wykorzystania zasobów sprzętowych, programowych, produktów cyfrowych RID, informacyjnych, finansowych, infrastrukturalnych w aspektach metodycznych⁴⁹, przedsięwzięć projektowych, praktycznych związanych z automatyzacją prowadzenia badań bibliograficznych w ocenie stanu wiedzy

⁴⁹ Data Repositories Requirements, <https://www.coretrustseal.org/why-certification/requirements/>, data dostępu: 12.01.2022 r.

naukowej, badawczej, praktycznej dla grup i zespołów specjalistów, wspomagana programowalnymi robotami cyfrowymi. Dla zapewnienia skuteczności procesów zarządczych powiązano rolę robota cyfrowego z rolami użytkowników, którym przyporządkowano obowiązki, przydzielono uprawnienia i odpowiedzialność. Całościowe spojrzenie na zarządzanie zasobami stanowi kompozycję zamkniętą w metodykę zautomatyzowanego prowadzenia badań bibliograficznych dla grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów. Wykorzystane zasoby wkomponowano w opisane już wzajemne zależne cykle:

- Cykl życia produktu cyfrowego Repozytorium Instytucjonalnego Danych (cykl budowy produktu cyfrowego, cykl udostępniania produktu cyfrowego użytkownikom w postaci e-usługi, cykl likwidacji).
- Cykl życia badań bibliograficznych na zasobach RID (1. Proces opracowania zbioru założeń badań bibliograficznych wg założeń wstępnych, Proces pozyskania zasobów cyfrowych (proces cykliczny), 3. Proces wykonania zapytania bibliograficznego i wyszukiwanie, 4. Proces analizy ilościowej i jakościowej danych, 5. Proces poddania pełnotekstowej analizie treści dokumentów, 6. Proces oceny wyników z konsumpcją zebranych informacji, 7. Proces weryfikacji i walidacji badań bibliograficznych).
- Cykl życia danych RID (proces pierwszego uzupełnienia RID, proces użytkowania danych, proces likwidacji).
- Cykl życia programowalnego robota cyfrowego (pozyskanie RPA, analiza, proces zaprogramowania robota cyfrowego i proces testów, praca, proces wsparcia i pielęgnacji, proces likwidacji).
- Cykl aktualizacji przyrostowej danych (cykl pierwszego uzupełnienia RID, cykl pracy, cykl aktualizacji przyrostowej i archiwizacja bieżąca).

W następnym podrozdziale zostanie zaprezentowane zagadnienie zarządzania ciągłością działania w procesie użytkowania Repozytorium Instytucjonalnego Danych.

2.3. Zarządzanie ciągłością działania RID

Zarządzanie ciągłością działania (BCM) to proces, dzięki któremu organizacja przygotowuje się na przyszłe incydenty lub sytuacje kryzysowe mogące zagrozić podstawowej misji organizacji oraz jej krótko – i długoterminowej

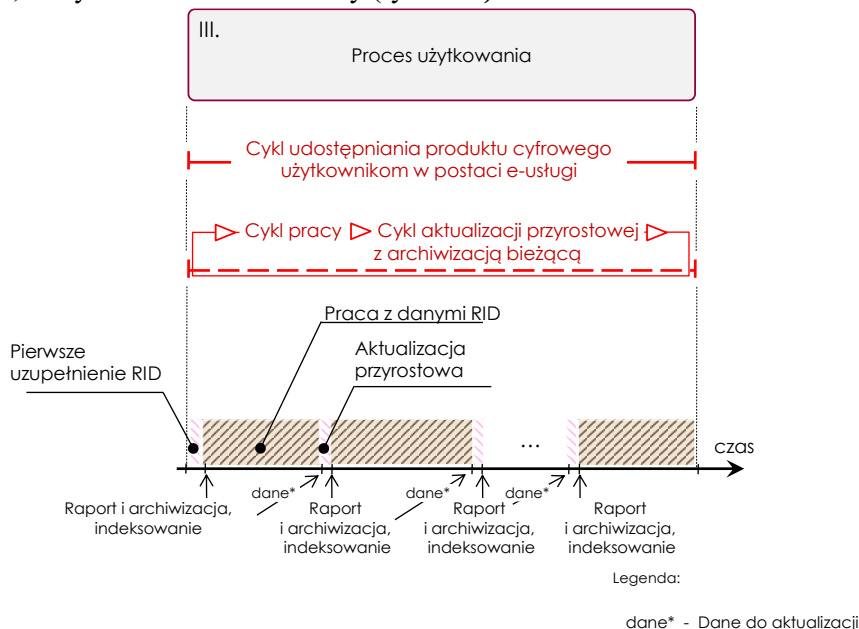
zdolności do ciągłości działania oraz sprostania oczekiwaniom interesariuszy⁵⁰. Do głównych zadań BCM należy: wsparcie kierownictwa, oszacowanie i ograniczanie ryzyka, analiza wpływu na działalność, utrzymanie, obsługa, konserwacja w budowie odporności organizacji⁵¹. W poprzednich rozdziałach przedstawione były elementy związane ze wsparciem użytkowników, pielęgnacja oraz zarządzanie ryzykiem. Podstawowym działaniem RID jest dostarczenie aktualnych zasobów dla procesów eksploracji, ekstrakcji wiedzy, które będzie miało istotne znaczenie w zarządzaniu ciągłością działania. Utworzenie i pierwsze uzupełnienie Repozytorium Instytucjonalnego Danych jest aktem jednorazowym w wyniku, którego uzyskuje się aktualny zbiór plików źródłowych w momencie ich pozyskania. Cykl pierwszego uzupełnienia RID rozpoczyna się zapytaniem bibliograficznym do źródeł i zbiorów danych. Następne działania dotyczą pozyskania zasobów i ich weryfikacji. Modyfikacja zasobów dotyczy tworzenia plików stowarzyszonych, uzupełniania metadanych, usuwanie zbędnych lub zduplikowanych zasobów. Cykl pierwszego uzupełnienia kończy się utworzeniem raportu z pozyskanych zasobów, jego archiwizacji oraz wykonania indeksacji zasobów (wariant I). Po zakończeniu cyklu rozpoczyna się kolejny cykl, który dotyczy procesu pracy ze zgromadzonymi dokumentami. Użytkownicy w e-usłudze zadają pytania bibliograficzne, otrzymują odpowiedzi, mają możliwość podglądu dokumentów źródłowych, korzystają z dokumentów stowarzyszonych i konsumują zebrane informacje. Istotnym problemem badań bibliograficznych jest aktualność zasobów, na podstawie których stwierdza się określony stan wiedzy w określonym temacie w dyscyplinie wiedzy. W określonym momencie należy dokonać aktualizacji zasobów o nowe, jeszcze niezarchiwizowane zasoby od momentu ich zapisania w RID lub pozyskania nowych zasobów wiedzy w oparciu o nowe zapotrzebowania ze strony użytkowników. Częstotliwość wykonywania aktualizacji jest wg potrzeb i na żądanie użytkownika lub działaniem inicjowanym przez koordynatora. Interwał aktualizacji jest ustalany na podstawie zarchiwizowanej daty ostatniego pozyskania zasobów z raportu pozyskanych zasobów. Aktualizacja może dotyczyć wszystkich elementów, tj. aktywów intelektualnych, metryk cyfrowych profili autorów osiągnięć twórczych, specjalistycznych

⁵⁰ *Międzynarodowe Ramowe Praktyki Zawodowe (IPPF)–Praktyczny przewodnik Zarządzanie ciągłością działania*. Instytut Audytorów Wewnętrznych, 2014, s. 2.

⁵¹ Sobczak A. (2022). *Robotic Process Automation as a Digital Transformation Tool for Increasing Organizational Resilience in Polish Enterprises*. Sustainability. 14(3):1333. <https://doi.org/10.3390/su14031333>.

metryk cyfrowych lub tylko wybranych elementów. W tym sensie aktualizacja określonego źródła i zbioru danych może być niezależna od innych aktualizacji i innych źródeł. Przyrostowa aktualizacja zasobów RID to zapewnienie ciągłości wiedzy skodyfikowanej i wiąże się z uzyskaniem ciągłości działania procesów biznesowych. Z punktu widzenia zarządzania, aktualizacja zasobów jest funkcją biznesową, a więc zestawem procesów wspierających osiągnięcie określonego celu biznesowego. Produktem biznesowym jest RID, który jest równocześnie produktem cyfrowym a aktualność zasobów odgrywa biznesową rolę.

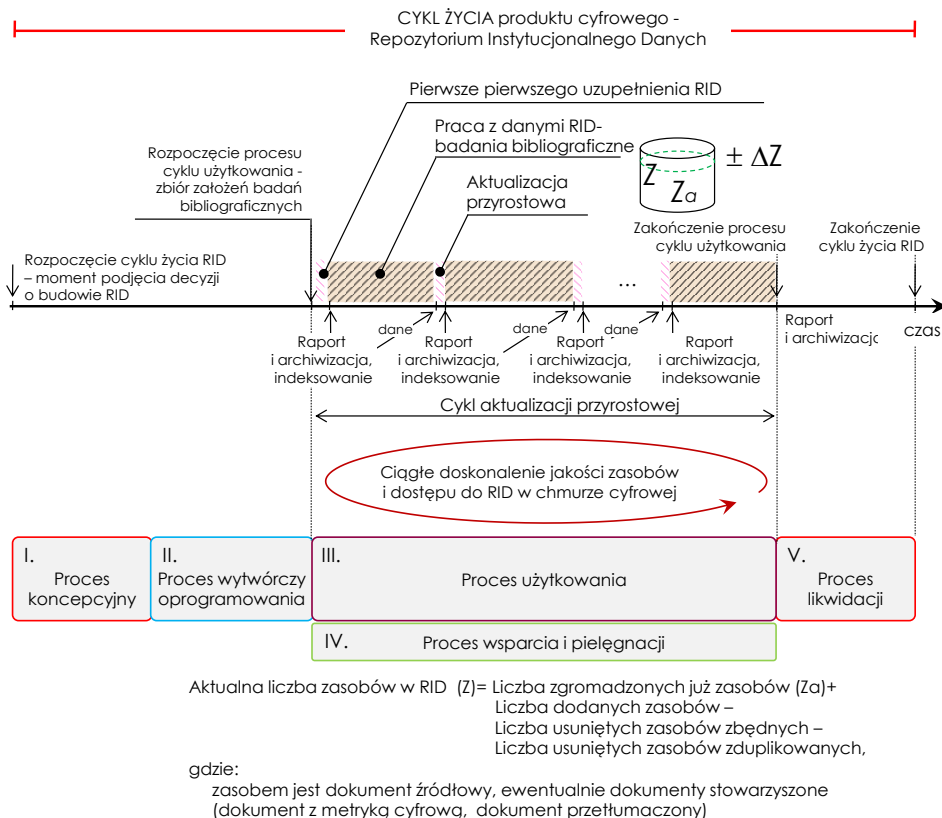
Analizując proces użytkowania RID w zarządzaniu ciągłością działania można wyróżnić dwa cykle które są naprzemienne: cykl pracy oraz cykl aktualizacji przyrostowej z archiwizacją bieżącą. Kolejny cykl aktualizacji przyrostowej odbywa się z wykorzystaniem danych z raportu z pozyskanych zasobów, który został zarchiwizowany (rys. 2.11).



Rys. 2.11. Zarządzanie ciągłością działania w procesie użytkowania
Repozytorium Instytucjonalnego Danych

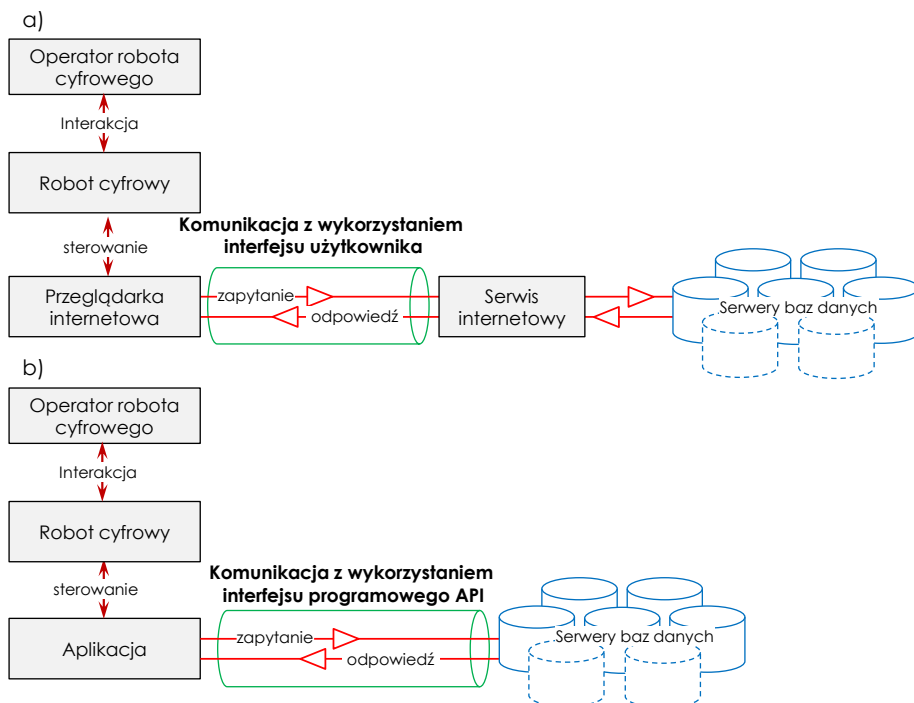
Cykl aktualizacji to proces ciągły z momentami dyskretnymi w postaci zbioru działań dyskretnych które skrótowo obejmują: Zapytania → Pozyskanie → Modyfikacja → Raport → Archiwizacja. Modelowe zapewnianie ciągłości działania RID w prowadzeniu badań bibliograficznych w ocenie stanu wiedzy

naukowej, badawczej i praktycznej zostało graficznie zaprezentowane na rysunku 2.12.



Rys. 2.12. Modelowe zapewnianie ciągłości działania RID
w prowadzeniu badań bibliograficznych

Aktualna liczba zasobów zgromadzonych w RID (Z) wynika z aktualnej liczby zgromadzonych zasobów (Z_a), zasobów dodanych (ΔZ), pomniejszona o liczbę usuniętych zasobów zbędnych oraz liczbę usuniętych zasobów zduplikowanych. Zasobem jest dokument źródłowy, ewentualnie dokument pozyskany lub dodany w postaci metryki cyfrowej lub dokument przetłumaczony dla zasobu niebędącego w języku polskim. Kluczowym procesem w uzyskaniu ciągłości działania jest technika pozyskania nowych zasobów przez ponowne zapytanie, które było zarchiwizowane lub przez nowe zapytania. Aktualizacja zasobów RID może się odbywać za pomocą interfejsu użytkownika lub komunikacji bezpośredniej z wykorzystaniem API przez użytkowników, koordynatora lub operatora przez zaprogramowanie robota cyfrowego (rys. 2.13).



Rys. 2.13. Możliwości aktualizacji zasobów RID z wykorzystaniem robotyzacji procesów:
a) za pomocą interfejsu użytkownika, b) komunikacji bezpośredniej przez API

Wymienione niewykluczające się możliwości aktualizacji zasobów RID zostały już zaprezentowane, ale z punktu widzenia automatyzacji procesów biznesowych istotna jest niezmiennosć źródła danych co implikuje sprawność pozyskania danych. Nie można jednoznacznie stwierdzić, który sposób aktualizacji zasobów w danym momencie jest najsprawniejszy. W komunikacji za pomocą GUI do aktualizacji zasobów można wykorzystać subskrypcję informacji, o ile serwis internetowy posiada taką funkcjonalność. Natomiast w komunikacji API można zaprogramować horyzonty zdarzeń aktualizacji zasobów już wykorzystywanych serwisów na wiele miesięcy przez określenie horyzontu czasowego kwerendy.

Z zarządzaniem ciągłością działania wiążą się zadania zapewniające ciągłość w powiązaniu z rolami osób w organizacji. W zakresie tych działań są zadania zapewniające skuteczne zarządzanie procesowe, regularne monitorowanie przebiegu procesu, mierzenie parametrów realizacji czynności wchodzących

w skład danego procesu i ich analizowanie⁵². Wymienione zadania dotyczą ról użytkowników w procesach, a więc podziału obowiązków, przydział uprawnień i odpowiedzialności, która wynika z logicznej dekompozycji funkcjonalnej, dzieląc procesy, systemy, obszary funkcjonalne lub produkty na części składowe. Role osób w procesach oraz niewyczerpująca lista czynności w metodyce zautomatyzowanego prowadzenia badań bibliograficznych zostały zaprezentowane w tabelach 2.1–2.4.

Tabela 2.1. Rola koordynatora w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Koordinator – może pełnić rolę użytkownika lub operatora robota cyfrowego

Do obowiązków należy:

- Opracowanie struktury założeń bibliograficznych, pozyskiwanie, dokumentowanie i uzgadnianie wymagań w ścisłej współpracy z użytkownikiem i operatorem robota cyfrowego. Wynikiem działań jest opracowanie dokumentu zawierającego zbiory założeń badań bibliograficznych na podstawie założeń wstępnych.
- Inicjacja procesów związanych z produktami cyfrowymi.
- Manualna weryfikacja zasobów w RID.
- Weryfikacja dokumentów PDF po translacji.
- Weryfikacja dokumentów niepoddających się utworzeniu w formacie PDF. Obróbka cyfrowa dokumentów (dzielenie plików, rozpoznawanie tekstu oprogramowaniem typu OCR).
- Zarządzania indeksem RID (wariant I korzystania z zasobów RID): tworzenie, aktualizacja, udostępnianie, import indeksów użytkowników.
- Zarządzania ciągłością działania: aktualizacja zasobów, wykonanie raportu z pozyskania zasobów, archiwizacja raportów.
- Wykonanie raportu z przyrostowej aktualizacji zasobów i zarchiwizowanie raportu. Dane z archiwizacji są wykorzystywane do następnej aktualizacji przyrostowej.
- Walidacja oraz ewaluacja dokumentów, zarządzanie raportami.
- Zapewnienie procesu uzyskania ciągłości działania (BCM) w prowadzeniu badań bibliograficznych przez użytkowników.
- Asysta techniczna w procesie wsparcia użytkownika przez zdefiniowanie i podane do wiadomości użytkownikom kanały dostępowe.

⁵² Kotnis M., *Modele zarządzania ryzykiem w warunkach niepewności*. Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 802 Finanse, Rynki Finansowe, Ubezpieczenia nr 65 (2014), s. 669–678.

c.d. Tabela 2.1. Rola koordynatora w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

- Pielęgnacja produktów cyfrowych przez zdefiniowane cztery kategorie doskonalenia: ulepszanie, naprawa, adaptowanie, zapobieganie.
- Współpraca z operatorem robota cyfrowego oraz użytkownikami.
- Zarządzanie procesami, procesami podstawowymi i pomocniczymi w RID.
- Ciągłe doskonalenie swojego warsztatu, podnoszenia jakości zasobów przez mechanizmy zarządzania zasobami RID.

Posiada uprawnienia:

- Administratora chmury cyfrowej.
- Autoryzacji do multidyscyplinarnych bazy danych, specjalistycznych serwisów internetowych.

Koordinator ma realizować działania w zakresie zarządzania infrastrukturą chmury cyfrowej w cyklu życia produktu cyfrowego:

- Utworzenie architektury informacji.
- Udostępnianie danych oraz zarządzanie uprawnieniami użytkowników.
- Zarządzanie danymi w chmurze cyfrowej.
- Odzyskiwanie danych usuniętych przypadkowo lub celowo.
- Raportowanie aktywności – tworzenie statystyk korzystania z zasobów.
- Archiwizacja bieżąca danych.
- Zarządzanie bezpieczeństwem informacji, m.in. sprawdzenia integralności, poufności, dostępności danych.
- Kontrola jakości zasobów chmury cyfrowej.

Koordinator odpowiedzialny jest za zadania:

- Zdefiniowanie dziedzin i dyscyplin naukowych.
- Opracowanie listy aktywów intelektualnych.
- Opracowanie listy multidyscyplinarnych baz danych i specjalistycznych serwisów internetowych.
- Opracowanie listy aktywów dostępnych przez interfejs użytkownika Web lub API.
- Opracowanie listy cyfrowych aktywów osobowych.
- Opracowanie listy metryk cyfrowych z podziałem na dane statyczne, dynamiczne.
- Opracowanie listy rekomendowanych grupy produktów cyfrowych wykorzystywanych.
- Opracowanie listy pełnionych ról w metodyce zautomatyzowanego prowadzenia badań bibliograficznych.

Tabela 2.2. Rola operatora robota cyfrowego w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Operator robota cyfrowego – osoba, która programuje robota, uruchamia i nadzorując jego pracę. Koordynator może pełnić rolę operatora robota cyfrowego
Operator jest odpowiedzialny m.in. za działania robota cyfrowego w zakresie: • Postępowania zgodnie z metodyką opracowaną przez twórców platformy RPA. • Pozyskania aktywów intelektualne w poszczególnych grupach (grupa I–VII) i metryk cyfrowych (grupa VIII) ze zbiorów danych i dokumentów źródłowych przez interfejs użytkownika Web lub API. • Weryfikacja i walidacja działania robota cyfrowego wg metodyki twórców platformy RPA lub medyki zmodyfikowanej. Operator jest odpowiedzialny m.in. prawne pozyskanie i oznakowanie aktywów intelektualnych oraz działań dotyczących: • Zarządzania własnością intelektualną. • Zarządzania prawami własności intelektualnej. Operator ma realizować działania w zakresie: • Procesów zarządzania oraz procesów podstawowych i pomocniczych w RID. • Ciągłego doskonalenia swojego warsztatu, podnoszenia jakości zasobów przez mechanizmy zarządzania zasobami RID. • Wykorzystania dostępnych metod, technik narzędzi w procesie wsparcia i pielęgnacji w procesie użytkowania RID ze współpracy z koordynatorem.

Tabela 2.3. Funkcje robota cyfrowego w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Programowalny robot cyfrowy – realizuje zadania zaprogramowane przez operatora robota cyfrowego
Robot zaprogramowany ma realizować działania: • Automatycznego pozyskiwania zasobów wg założeń bibliograficznych, zasobów dostępnych przez interfejs użytkownika Web lub API i zapis w chmurze obliczeniowej. • Automatycznego pozyskiwania metryk cyfrowych profili autorów. • Automatycznego pozyskiwania specjalistycznych metryk cyfrowych. • Translacji zbiorów dokumentów na język polski. • Sprawdzenie zabezpieczenia dokumentu PDF przed edycją lub zaznaczaniem treści.

c.d. Tabela 2.3. Funkcje robota cyfrowego w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

- Umieszczenie metadanych z metryk cyfrowych w dokumentach PDF, jeżeli dokument ich nie zawierał.
- Utworzenie opcjonalnie metryki cyfrowej dla dokumentu źródłowego w postaci dokumentu PDF.
- Sprawdzenie automatyczne występowania duplikatów dokumentów typu PDF wg zawartości i w przypadku wystąpienia usunięcie duplikatów.
- W procesie weryfikacji i walidacji poszerzonej RID połączenie indywidualnych ankiet walidacyjnych dotyczących: 1. Walidacji produktu cyfrowego, 2. Walidacji badań bibliograficznych na zasobach, 3. Walidacji zdefiniowanych ról, 4. Walidacji potrzeb, oczekiwań dla użytkowników.

Tabela 2.4. Rola użytkownika w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

Użytkownicy – grupy, zespoły dyscyplinarne lub interdyscyplinarne specjalistów dokonujące oceny stanu wiedzy naukowej, badawczej i praktycznej w pracach nad nowymi osiągnięciami twórczymi

Zadaniem użytkownika jest zbierania faktów w celu zdobywania, zgłębiania wiedzy o stanie nauki i techniki dla określenia parametrów granicznych badań dla tworzenia nowej, rozszerzonej lub pogłębionej wiedzy w procesie eksploracji i ekstrakcji wiedzy z zasobów RID. Użytkownicy realizują zadania wg metodycznego podejścia do badania bibliograficznego w cyklu życia badań bibliograficznych na zasobach RID w zakresie:

- Dokonują identyfikacji, dokumentowania swoich wymagań lub zespołu, którym kierują oraz sporządzają dokument zawierający założenia dla badań bibliograficznych.
- Zarządzania indeksami użytkownika z wykorzystaniem rekomendowanej aplikacji analitycznej dla tworzenia, aktualizacji, modyfikowania, eksportu, importu, usuwania indeksów.
- Zarządzania indeksami użytkownika z możliwością udostępnienia innym użytkownikom lub włączenia indeksu z zasobami źródłowymi do indeksu repozytorium RID.
- Zarządzanie wyszukiwaniem informacji w indeksach RID: z wykorzystaniem kwerend (budowa, modyfikacja, archiwizacja kwerend), zmiany parametrów wyszukiwania w aplikacji analitycznej, wyszukiwanie proste i złożone.
- Zarządzanie rezultatami wyszukiwań: selekcja i rankingowanie rezultatów, drążenie rezultatów wyszukiwania, eksport wyników, sortowanie i filtrowanie.

c.d. Tabela 2.4. Rola użytkownika w metodyce zautomatyzowanego prowadzenia badań bibliograficznych

- Ekstrakcji informacji, analizy danych i interpretacji (analiz tabelarycznych i statycznych wyników zapytań, podglądu dokumentów źródłowych, korzystania z istniejących metryk cyfrowych, oceny ilościowej i jakościowej, syntezy wyników).
- Kodyfikacji pozyskanej i wytworzonej przez siebie wiedzy.
- Konsumpcji wiedzy pozyskanej z zasobów RID przez cytowania, zestawienia cytowań, sporządzanie bibliografii dla nowego osiągnięcia naukowego, badawczego, praktycznego. Użytkownicy mogą stosować alternatywne aplikacje do zarządzania bibliografią i cytowaniami, np. EndNote, Mendeley, Zotero⁵³.
- Utworzenia nowych osiągnięć twórczych, w których wykorzystywany był etap badań bibliograficznych.
- Przechowywania, weryfikacji i walidacji, upublicznienia oraz wykorzystania wiedzy – wykonanie i archiwizacja raportów z badań bibliograficznych.
- Korzystania z funkcjonalności rekomendowanej chmury cyfrowej oraz rekomendowanych aplikacji.
- Zarządzania zasobami prywatnymi z możliwością ich włączenia do zasobów RID przez operacje: scalania dokumentów, dzielenia dokumentu na kilka plików, wstawiania obiektów w dokumentach na własny użytek, komentowanie, tworzenia zakładek, zakreslanie ważnych fragmentów treści.
- Upowszechnienie własnego dorobku naukowego, badawczego i praktycznego przez włączenie osiągnięcia twórczego w określone miejsce struktury architektury informacji RID.
- Zgłaszanie potrzeb w zakresie nowych źródeł i zbiorów danych lub funkcjonalności chmury cyfrowej albo aplikacji RID.
- Ciągłego doskonalenia swojego warsztatu, podnoszenia jakości zasobów przez mechanizmy zarządzania zasobami RID.

Dodatkowo oprócz wymienionych działań w rolach użytkownika, operatora robota cyfrowego, koordynatora należy zapewnić możliwość doskonalenia swojego warsztatu pracy i podnoszenia jakości zasobów RID. W związku z tym, jakość osiągniętych wyników będzie w znacznym stopniu określona przez system zarządzania jakością w organizacji. Zagadnienia jakości zostaną zaprezentowane w następnym podrozdziale.

⁵³ Comparison of reference management software.

https://en.wikipedia.org/wiki/Comparison_of_reference_management_software, data dostępu 15.12.2021 r.

2.4. Zarządzanie jakością produktów cyfrowych, danych i e-usługi

Sukces innowacji wprowadzony do badań bibliograficznych zależy od wypracowanego unikatowego *know-how*, zdobycia wiedzy i wdrożenia na jej podstawie odpowiednich narzędzi i na końcu zadowolenia i satysfakcji końcowych użytkowników. Zadowolenie to percepcja użytkownika dotycząca stopnia w jakim jego wymagania zostały spełnione. Dla uzyskania satysfakcji użytkownika niezbędne jest systematyczne doskonalenie w zapewnieniu ciągłości działania procesów biznesowych. Przy czym ze względu na podział ról w RID, można wyróżnić trzy podejścia rozumienia procesów biznesowych i pojęcia jakości. Dla koordynatora proces biznesowy to jakość zgromadzonych zasobów, jakość dostępu do zasobów (zarządzanie profilami cyfrowymi, uprawnieniami dostępu, integralność, dostępność, redundancja). Dla operatora robota cyfrowego to niezawodność działania oprogramowania robota dla aktualizacji zasobów w uzyskaniu ciągłości zasobów skodyfikowanej wiedzy. Dla użytkownika to jakość zgromadzonych zasobów (aktualność, integralność, dostępność, aplikowalność) i jakość wykorzystywanych aplikacji, np. analitycznej (funkcjonalność, łatwość uczenia się).

Zeprezentowane w niniejszym podrozdziale zagadnienie zarządzanie jakością produktów cyfrowych, danych i e-usługi w Repozytorium Instytucjonalnego Danych oparte są na ośmiu zasadach zarządzania jakością zawartych w normach serii ISO 9000 oraz normach szczegółowych zagadnień związanych z zarządzaniem jakością i elementami systemu zarządzania jakością ujętych w normach serii ISO 100xx⁵⁴. Zasady mają ułatwić osiągnięcie celów oraz w działaniach organizacji w kierunku doskonalenia jego funkcjonowania poprzez⁵⁵:

- Zasada 1. Orientacja na klienta.
- Zasada 2. Przywództwo.
- Zasada 3. Zaangażowanie ludzi.
- Zasada 4. Podejście procesowe.
- Zasada 5. Podejście systemowe do zarządzania.

⁵⁴ ISO 10006:2017 *Quality management systems—Guidelines for quality management in Project*, ISO 10014:2006 *Quality management—Guidelines for realizing financial and economic benefits*, ISO 10015:2019 *Guidelines for competence management and people development*.

⁵⁵ ISO 9000:2015 *Quality management systems—Fundamentals and vocabulary*.

- Zasada 6. Ciągłe doskonalenie.
- Zasada 7. Podejmowanie decyzji na podstawie faktów.
- Zasada 8. Wzajemnie korzystne powiązania z dostawcami.

Opisane zasady przede wszystkim dotyczą głównie sektora przedsiębiorstw nastawionych na zysk, jednak są na tyle uniwersalne, że można je wykorzystać w każdej organizacji dążącej do stworzenia sprawnego systemu zarządzania jakością, w tym także do repozytoriów instytucjonalnych organizacji sektora finansów publicznych.

Zasada 1. Orientacja na klienta

Odbiorcą metodyki automatyzacji pozyskania aktywów intelektualnych z rozproszonych źródeł danych dla utworzenia scentralizowanego zasobu dokumentów w organizacji są osoby, grupy, zespoły monodyscyplinarne lub interdyscyplinarne specjalistów. Zastosowanie zasady „orientacja na klienta” umożliwia: a) zrozumienie potrzeb informacyjnych i oczekiwań użytkowników końcowych, b) zapewnienie powiązania wiązki celów utworzonych w RID, wiązki celów organizacji z potrzebami i oczekiwaniami użytkowników, c) systematyczne zarządzanie relacjami, d) zaplanowanie i realizację pomiarów zadowolenia i satysfakcji z wykorzystaniem walidacji poszerzonej (Procedura P8). W szerszym ujęciu, tworzenie, aktualizacja zasobów RID dokonuje się w sieci wartości. Sieć wartości można rozpatrywać z punktu widzenia ustanowionych ogniw w łańcuchu logistycznym w pięciu perspektywach:

1. Perspektywa twórców aktywów intelektualnych, którzy wytwarzają skodyfikowaną wiedzę i na określonych prawach licencyjnych przekazują dostawcom dane lub klientom.
2. Perspektywa dostawców danych, od których pozyskuje się zasoby do RID z rozproszonych repozytoriów danych. Dostawcy nie są zainteresowani konkurencją oraz dostarczaniem danych masowych bezpłatnie. Oferują usługi płatności za przygotowanie masowych danych, np. *Scopus Custom*

Data⁵⁶ lub w modelu subskrypcyjnym albo przez partnerstwo danych, np. Semantic Scholar⁵⁷.

3. Perspektywa dostawców produktów cyfrowych, którzy są producentami specjalistycznych aplikacji oraz są dostawcami technologii chmur cyfrowych.
4. Perspektywa brokera informacji, który pozyskuje zasoby od dostawców danych lub twórców aktywów intelektualnych i przechowuje zasoby w RID. W tym sensie instytucjonalne repozytorium jest platformą cyfrową pośredniczącą pomiędzy twórcami, dostawcami producentami i konsumentami.
5. Perspektywa użytkownika końcowego RID, którego celem jest wyszukiwanie skodyfikowanej wiedzy w zgromadzonych aktywach intelektualnych z wykorzystaniem e-usługi. Jest on konsumentem danych i zarazem klientem RID. W tym przypadku broker informacji oraz użytkownik końcowy należy do tej samej organizacji.

Analiza barier o różnym charakterze przedstawiona we wprowadzeniu do monografii według zasady orientacji na klienta oraz wyartykułowane potrzeby specjalistów w organizacji, możliwości, ograniczenia i planowane koszty, korzyści⁵⁸, wpłynęły na rozwiązania techniczno-organizacyjne i programistyczne. Wybór zakresu robotyzacji oraz sposobu wykorzystania programowalnego robota do zadań realizowanych manualnie na etapie badań bibliograficznych wynikał z zasad: – myślenia o potrzebach użytkownika i realizacji działań w kierunku zaspokojenia jego potrzeb (*User Centred Design*, UCD), – projektowania opartym na doświadczeniu użytkownika (*User Experience*, UX) – automatyzacji zadań procesowych opartych na regułach w celu zwiększenia wydajności procesu i obniżenia kosztów procesu badań bibliograficznych⁵⁹. Zaprezentowane bariery, w szczególności problemów odnoszących się do zarządzania specjalistami w organizacji dotyczą:

- braku podaży danych naukowych w sposób scentralizowany i jednolitego sposobu wyszukiwania,
- braku opisu do osiągnięć twórczych w postaci metadanych,

⁵⁶ Scopus Custom Data, https://www.elsevier.com/data/assets/pdf_file/0003/810579/ELSV-13178-ACAD-SC-Custom-Poster-r0c-print.pdf, data dostępu: 16.12.2021 r.

⁵⁷ Semantic Scholar, <https://www.semanticscholar.org/>, data dostępu: 16.12.2021 r.

⁵⁸ Aguirre, S., & Rodríguez, A. (2017). *Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study*. WEA.

⁵⁹ Wewerka, J., & Reichert, M. (2021). *Robotic process automation-a systematic mapping study and classification framework*. Enterprise Information Systems, 1–38.

- braku struktur typu repozytoria danych z możliwością dzielenia się doświadczeniem i wiedzą,
- braku możliwości wykorzystania specjalistycznego oprogramowania ograniczającej specjalistów organizacji w osiągnięciu wysoko punktowanych osiągnięć twórczych.

Geneza problemów i orientacja na użytkownika zapoczątkowała proces koncepcyjny i realizacyjny opracowania metodyki zautomatyzowanego prowadzenia badań bibliograficznych w postaci RID przechowującej szeroki zakres aktywów intelektualnych z grupą metryk cyfrowych. Ostatecznie zrealizowane rozwiązanie techniczno–organizacyjne, programistyczne zostały oparte na analizie rzeczywistych potrzeb odbiorców (UCD, UX)) i utworzenie RID w powiązaniu z platformą chmury cyfrowej był podyktowane myśleniem zorientowanym na użytkownika – kliencie jako odbiorcy produktów cyfrowych.

Zasada 2. Przywództwo

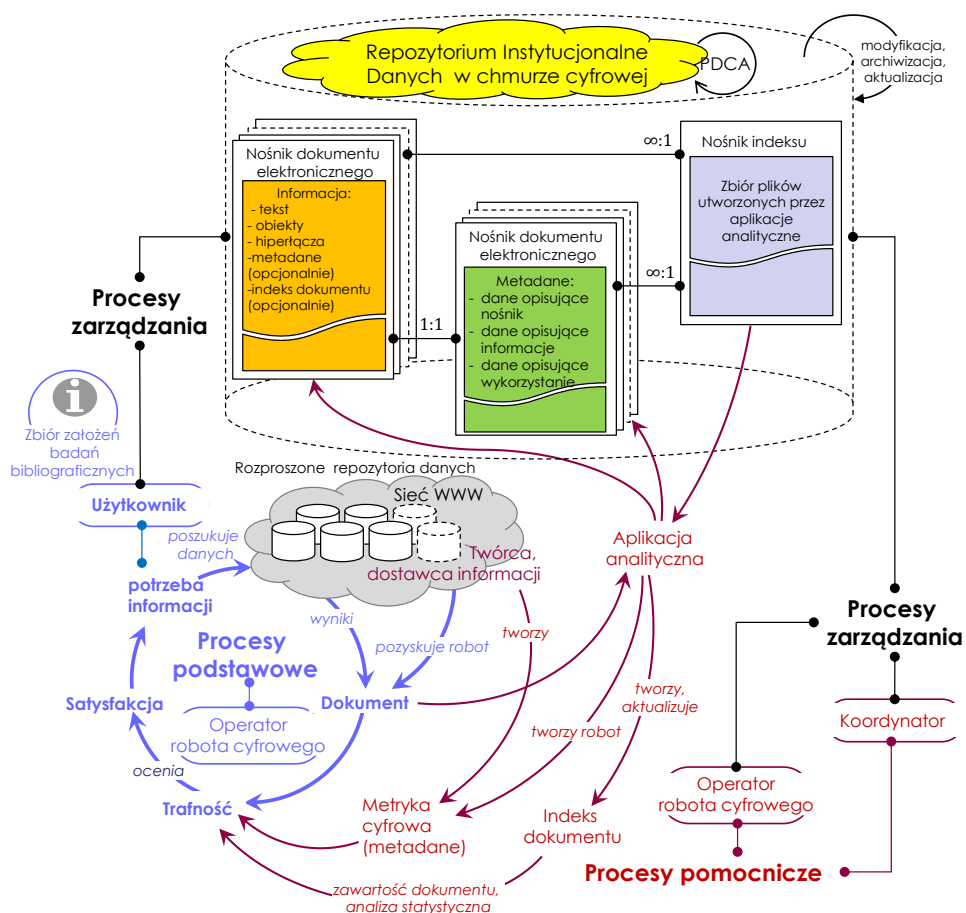
Zasada przywództwa uwzględnia potrzeby wszystkich zainteresowanych stron, umożliwia określenie wiązki celów i zadań do zrealizowania, zapewnienia użytkownikom wymaganych zasobów, szkolenia i swobody działania w ramach ich odpowiedzialności i uprawnień. Wprowadzenie czterech ról w metodyce zautomatyzowanego prowadzenia badań bibliograficznych, w tym trzy role dla użytkowników RID wynika z logicznej dekompozycji funkcjonalnej całego zagadnienia, dzieląc procesy, systemy, obszary funkcjonalne lub produkty na części składowe. W sposób kompleksowy przyporządkowano obowiązki, przydzielono uprawnienia i odpowiedzialność. Opracowane graficzne algorytmy jednoznacznie dookreślają zakres działań do pełnionych ról w przeprowadzaniu badań bibliograficznych. Rolą nadrzędną jest funkcja koordynatora, którego głównym zadaniem jest rozwijanie kluczowych procesów biznesowych, inspirowania, zachęcania do dzielenia się wiedzą i doświadczeniem. Szczegółowe wymagania kompetencyjne zostały przedstawione w rekomendacjach a lista czynności w metodyce zautomatyzowanego prowadzenia badań bibliograficznych została zaprezentowana w podrozdziale 2.3. *Zarządzanie ciągłością działania RID.*

Zasada 3. Zaangażowanie ludzi

W gospodarce opartej na wiedzy najważniejszym zasobem są ludzie. Zaangażowanie ludzi dotyczy wszystkich użytkowników zaangażowanych w rozwój produktów cyfrowych RID, które są przez nich wykorzystywane. Podnoszenie e-kompetencji użytkowników wpływa na skuteczność wydobywania wiedzy z zasobów RID, zmniejszając czas wyszukania. Istnieje więc dodatnie sprzężenie zwrotne. Zapewnienie użytkownikom możliwości rozwoju w organizacji, podnoszenie kompetencji i doskonalenia pozytywnie wpływa na efekty badań bibliograficznych. Utworzenie mechanizmów dzielenia się wiedzą i wymiany doświadczeń, np. z wykorzystaniem funkcjonalności chmury cyfrowej czy też możliwość komentowania, tworzenie zakładek tematycznych, zakreślanie ważnych fragmentów treści w dokumencie PDF przyczynia się do ciągłego doskonalenia.

Zasada 4. Podejście procesowe

Z punktu widzenia użytkownika jego praca z danymi RID ma przyspieszyć badania bibliograficzne przez pracę się w cyklu procesów powtarzalnych, do momentu osiągnięcia satysfakcji z badań bibliograficznych. W tym sensie potrzeby informacyjne są przyczynkiem do poszukiwania interesującej wiedzy w dokumentach zgromadzonych w RID. Trafność otrzymanych wyników jest podstawą oceny satysfakcji z wyszukania i osiągnięcia stawianych celów. Proces poszukiwania jest powtarzany iteracyjnie do zakończenia badań. Projakościowa orientacja wymusza stosowanie modelu zarządzania i usprawniania procesów RID. Dla sprawnego zarządzania RID wyróżnia się trzy procesy: podstawowe, pomocnicze i zarządzania. Są to procesy przynależne modelowi zarządzania w ujęciu procesowym. Procesy podstawowe o decydującym znaczeniu związane z realizacją misji, wizji, wartościami organizacji, pomocnicze, związane z automatyzacją żmudnych monotonicznych czynności, odciążające użytkowników, zarządzania, dotyczące procesów decyzyjnych, sterowania organizacją i oceną jakości procesów. Całokształt procesów przynależnych modelowi zarządzania w ujęciu procesowym został zaprezentowany na rysunku 2.14.



Rys. 2.14. Procesy zarządzania, podstawowe i pomocnicze w systemie zarządzania jakością

Realizowane procesy wpływają na osiągnięcie efektów bezpośrednich w postaci nowych osiągnięć twórczych i uwzględniają oczekiwania, potrzeby informacyjne użytkowników.

Zasada 5. Podejście systemowe do zarządzania

Podejście systemowe umożliwia zrozumienie, zarządzanie wzajemnie powiązanymi procesami jako systemem i przyczynia się do zwiększenia skuteczności zarządzania zasobami w osiąganiu celów.

Zasada 6. Ciągłe doskonalenie

Ciągłe doskonalenie funkcjonowania organizacji jest jednym ze stawianych celów. Ciągłe doskonalenie dotyczy: a) szkolenia użytkowników – specjalistów w zakresie metod i narzędzi ciągłego doskonalenia, b) doskonalenia produktów cyfrowych, danych, e-usługi dostępowej do zasobów RID, c) procesów zarządzania, podstawowych i pomocniczych, d) sposobów pomiaru, uznawania i potwierdzania doskonalenia przez odpowiednie procesy.

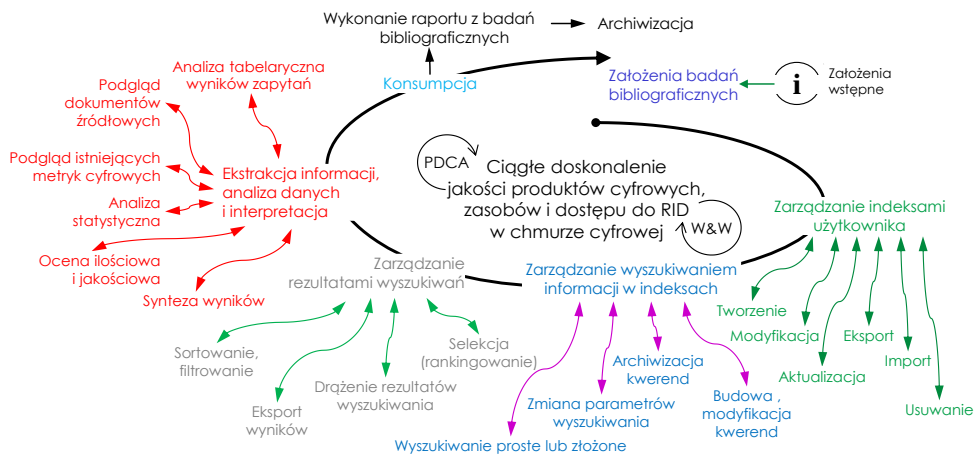
Wdrożenie innowacji w organizacji wymaga od koordynatora analizowania potrzeb użytkowników związanych z kompetencjami. Wykorzystywane aplikacje są bardzo specjalistyczne, oparte na płatnych licencjach i tym samym są w niewielkim stopniu znane i wykorzystywane. Dlatego istotna jest identyfikacja i analiza potrzeb użytkowników oraz zaplanowanie poszerzenia, pogłębienia kompetencji w zakresie metodycznego podejścia w badaniach bibliograficznych oraz kompetencji z technik informacyjno-komunikacyjnych (ICT). Zmiana kompetencji możliwa jest przez ciągłe doskonalenie poprzez trening oparty na faktach i szkolenia. W systemie zarządzania jakością określono e-kompetencje adekwatne do pełnione ról, które zostały zidentyfikowane i przedstawione w rekomendacjach. Natomiast efektywność i skuteczność inwestycji w kapitał ludzki (użytkowników) podlega monitorowaniu w procesie użytkowania, wsparcia i pielęgnacji jest weryfikowana w osiąganiu zamierzonych celów. Proces jest monitorowany przez Proces 7. *Proces weryfikacji i walidacji badań bibliograficznych* oraz procedurę *Walidacja poszerzona* (Procedura P8).

W celu zapewnienia systemowego podejścia do zarządzania jakością produktów cyfrowych, danych i dostępu do repozytorium RID przez e-usługę w chmurze cyfrowej należy wyodrębnić elementy składowe. W skład systemu zarządzania jakością wchodzi następujące elementów składowe RID: produkty cyfrowe w RID, role użytkowników w RID, wysokopoziomowa architektura RID, procesy RID przynależne modelowi zarządzania, procesy w cyklu życia RID, zasoby RID (rys. 2.15).

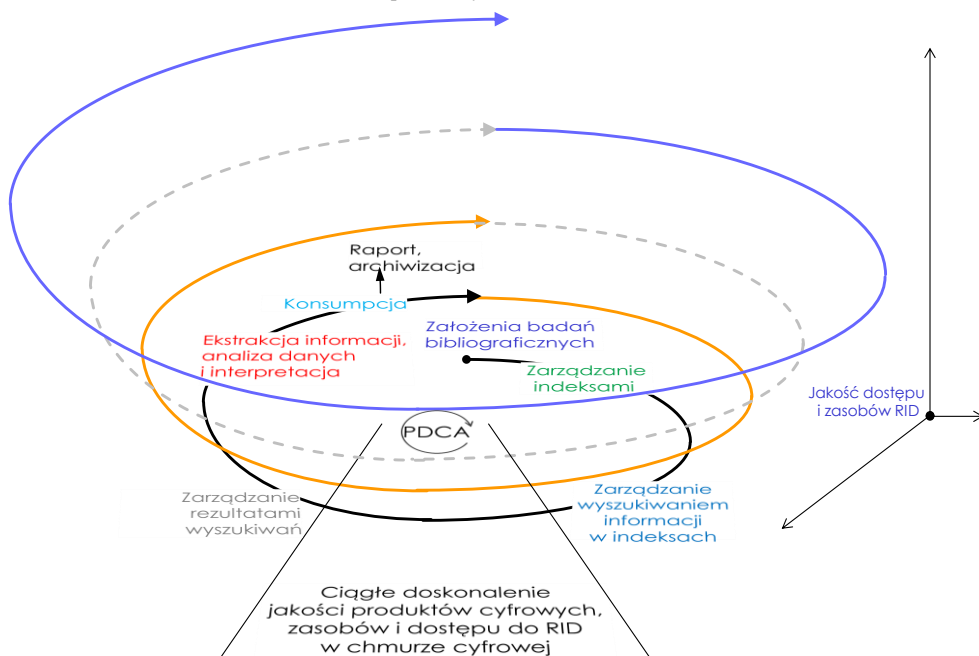
Elementy składowe w ciągłym doskonaleniu jakości produktów cyfrowych, zasobów i dostępu do RID w chmurze cyfrowej		
Produkty cyfrowe w RID	Rola użytkowników w RID	Wysokopoziomowa architektura RID
Repozytorium Instytucjonalne Danych - bazy danych: plików źródłowych, metryk cyfrowych, plików przetłumaczonych, indeksów	Użytkownik końcowy	Warstwa e-usług
Programowy robot cyfrowy	Operator robota cyfrowego	Warstwa aplikacji i prezentacji
Chmura cyfrowa	Koordynator	Warstwa danych przetworzonych
Architektura informacji	Rola cyfrowa w RID	Warstwa danych pozyskanych
Zestaw programów komputerowych: aplikacje analityczne, agregatory treści, transakcyjne, specjalistyczne, oprogramowanie robota	Programowalny robot cyfrowy	Warstwa danych źródłowych do eksploracji
Procesy RID przynależne modelowi zarządzania	Procesy w cyklu życia RID	Zasoby RID
Podstawowe	Cykl życia Repozytorium Instytucjonalnego Danych	Produkty cyfrowe
Pomocnicze	Cyklus życia badań bibliograficznych na zasobach RID	Użytkownicy
Zarządzania	Cyklus życia danych RID	Finansowe
	Cyklus życia programowego robota cyfrowego RID	Rzeczowe

Rys. 2.15. Elementy składowe w ciągłym doskonaleniu jakości produktów cyfrowych, zasobów i dostępu do RID w chmurze cyfrowej

Jakość osiągniętych wyników będzie w znacznym stopniu określona przez rodzaj i siłę związków pomiędzy poszczególnymi, zależnymi procesami. Jednak ze względu na podział ról należy wydzielić procesy przynależne użytkownikom oraz koordynatorowi i operatorowi programowalnego robota cyfrowego. Ciągłe doskonalenie jakości produktów cyfrowych, zasobów i dostępu do Repozytorium Instytucjonalnego Danych (RID) w chmurze cyfrowej przez użytkownika została zaprezentowana na rysunku 2.16–2.17. Całość cyklu najlepiej spina idea ciągłego doskonalenia *Planuj – Wykonaj – Sprawdź – Działaj* według klasycznego cyklu W. E. Deminga (PDCA). *Planowanie* dotyczy ustalenia celów i procesów niezbędnych do dostarczenia wyników zgodnych z wymaganiami użytkownika, zdefiniowania i przeanalizowania problemu, określenie rozwiązania i przygotowanie planu działania. *Wykonaj* dotyczy wdrożenia procesów. *Sprawdź* pozwala na monitorowanie i mierzenie procesów. *Działaj* dotyczy działań korygujących, standaryzacji procesów, ciągłego doskonalenia funkcjonowania procesów.

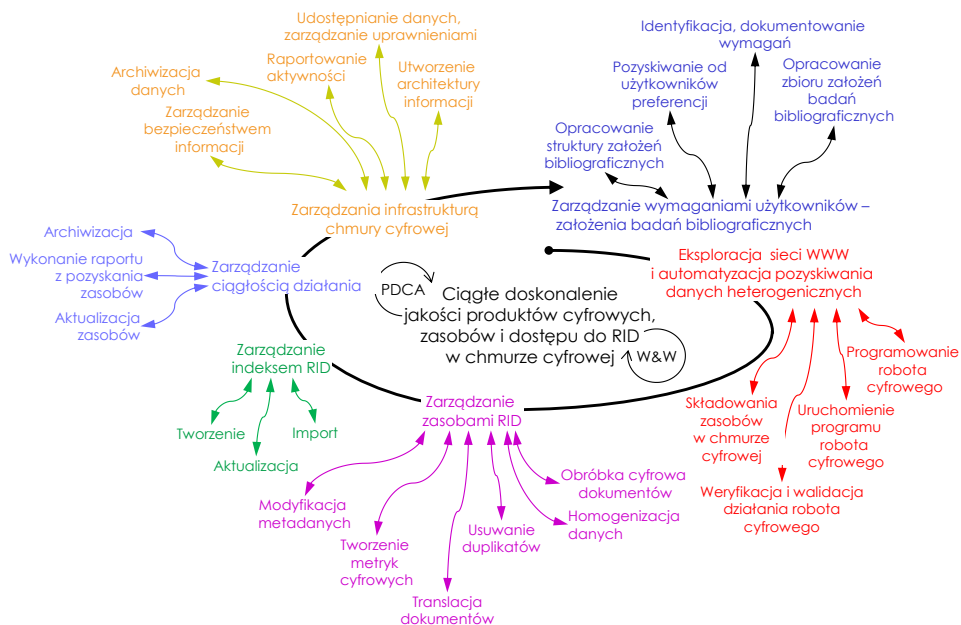


Rys. 2.16. Ciągłe doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez użytkownika

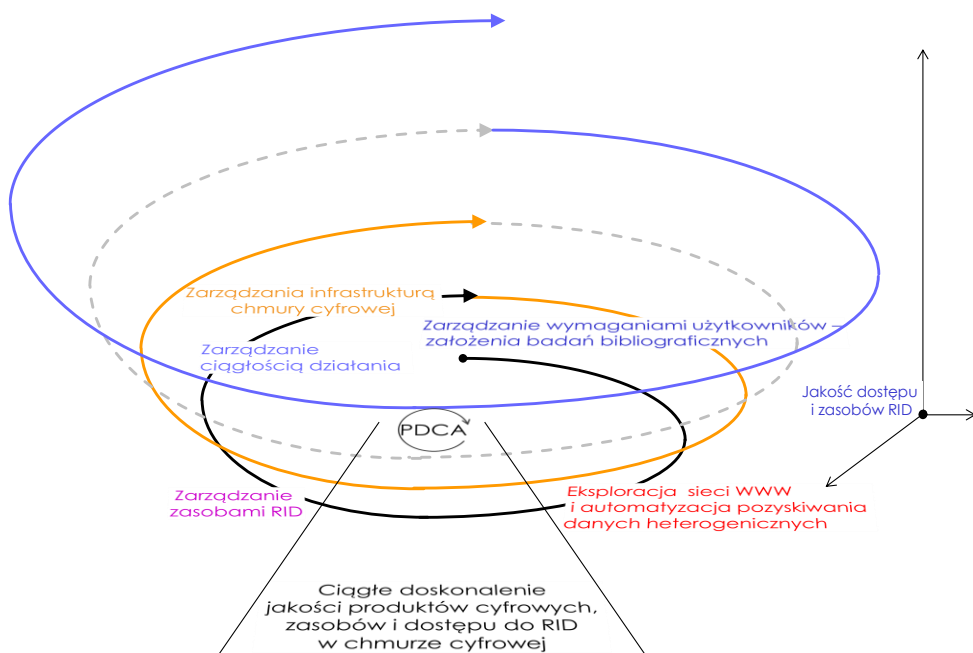


Rys. 2.17. Rozwój, doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez użytkownika

Ciągłe doskonalenie jakości produktów cyfrowych, zasobów i dostępu do Repozytorium Instytucjonalnego Danych (RID) w chmurze cyfrowej przez koordynatora i operatora programowalnego robota cyfrowego została zaprezentowana na rysunku 2.18–2.19.



Rys. 2.18. Ciągłe doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez koordynatora i operatora programowalnego robota cyfrowego



Rys. 2.19. Rozwój, doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez koordynatora i operatora programowalnego robota cyfrowego

Zasada 7. Podejmowanie decyzji na podstawie faktów

Aktualność, kompletność zasobów RID ma kluczowe znaczenie dla tworzenia nowej, rozszerzonej lub pogłębionej wiedzy na podstawie zgromadzonych zasobów wtórnych o stanie nauki i techniki. Wykorzystanie różnego rodzaju aplikacji do analizy zasobów w podejmowaniu decyzji na podstawie faktów wynika z gromadzenia danych w sposób przemysłany w zakresie liczebności zbioru, rodzaju pozyskanych, przetwarzanych dokumentów, sposobów ich przechowywania, weryfikacji, metody analizy. Aktualizacja przyrostowa zwiększa dokładność i wiarygodność zasobów przyczyniając się do podejmowania decyzji opartych na wiarygodnych informacjach. Udostępnianie zasobów z wykorzystaniem e-usługi następuje przez specjalistów, którzy ich potrzebują a analiza informacji następuje przy użyciu metodyki badań bibliograficznych.

Zasada 8. Wzajemnie korzystne powiązania z dostawcami

Pozyskiwanie zasobów i ich przechowywania w RID jest zależne od dostawców danych z rozproszonych repozytoriów danych. W tym sensie istnieje zależność jednokierunkowa pomiędzy dostawcami dysponującymi źródłami i zbiorami danych. RID nie może istnieć bez dostawców, gdyż organizacja wytwarza zbyt mało własnych, oryginalnych zasobów wiedzy, a utworzone stowarzyszone zasoby są uzależnione od danych źródłowych. Natomiast dostawcy danych funkcjonują według własnego modelu działania i nie są zainteresowani zasobami RID. Dlatego zasada wzajemności korzystnych powiązań jest ograniczona wyłącznie do jednej strony.

3. Rekomendowane działania w zarządzaniu zasobami ludzkimi

Przeprowadzona analiza wprowadzenia innowacji w procesie badań bibliograficznych w oparciu o przetworzone zarchiwizowane dane upoważnia do sformułowania rekomendacji praktycznych skierowanych do osób w organizacji pełniących role: bezpośrednich użytkowników, koordynatora, operatora robota cyfrowego.

3.1. Rekomendacje dla użytkowników

Rekomendacje dla użytkowników obejmują zagadnienia wymagań kompetencyjnych, zarządzania indeksami użytkownika, konsumpcji wiedzy z RID i są następujące:

1. Wykorzystanie zasobów Repozytorium Instytucjonalnego Danych wymaga rozwoju e-kompetencji użytkowników. Podwyższenie e-kompetencji powinno dotyczyć wykonywania zadań w cyklu życia badań bibliograficznych oraz działań operacyjnych na dokumentach PDF. W zakresie wykonywania zadań e-kompetencje powinny zostać rozszerzone lub pogłębione o następujące zagadnienia:
 - Metodycznego podejścia do badania bibliograficznego w cyklu życia badań bibliograficznych na zasobach RID.
 - Sposobów wykonywania identyfikacji, dokumentowania indywidualnych lub zespołowych wymagań oraz sporządzania dokumentu zawierającego założenia dla badań bibliograficznych.
 - Zarządzania indeksami użytkownika z wykorzystaniem rekomendowanej aplikacji analitycznej dla tworzenia, aktualizacji, modyfikowania, eksportu, importu, usuwania indeksów.
 - Zarządzanie wyszukiwaniem informacji w indeksach: z wykorzystaniem kwerend (budowa, modyfikacja, archiwizacja kwerend), zmiany parametrów wyszukiwania w aplikacji analitycznej, wyszukiwanie proste i złożone.
 - Zarządzanie rezultatami wyszukiwań: selekcja i rankingowanie rezultatów, drążenie rezultatów wyszukiwania, eksport wyników, sortowanie i filtrowanie.
 - Umiejętności ekstrakcji informacji, analizy danych i interpretacji (analiz tabelarycznych i statycznych wyników zapytań, podglądu

dokumentów źródłowych, korzystania z istniejących metryk cyfrowych, oceny ilościowej i jakościowej, syntezy wyników).

- Kodyfikacji pozyskanej i wytworzonej przez siebie wiedzy.
- Konsumpcji wiedzy, cytowania, zestawienia cytowań, sporządzanie bibliografii.
- Wykonania i archiwizacji raportu z badań bibliograficznych.
- Ciągłego doskonalenia swojego warsztatu, podnoszenia jakości zasobów i dostępu do Repozytorium Instytucjonalnego Danych (RID) w chmurze cyfrowej.
- Korzystania z funkcjonalności rekomendowanej chmury cyfrowej.

W zakresie pogłębiania e-kompetencji w działaniach operacyjnych na pojedynczych dokumentach lub zbiorach dokumentów PDF użytkownik powinien wykonywać operacje: scalania dokumentów, dzielenia dokumentu na kilka plików, wstawiania obiektów w dokumentach na własny użytek, komentowanie, tworzenia zakładek, zakreślanie ważnych fragmentów treści.

2. Korzystanie z danych wtórnych przechowywanych w RID nie zwalnia użytkowników z zasad prawidłowego korzystania z utworów, które są chronione odpowiednimi prawami autorskimi i handlowymi. Zasady, sposób i korzystanie z utworów archiwalnych są identyczne jak ich użycie z pierwotnych źródeł. Użytkownik może w każdej chwili zapoznać się z zakresem wykorzystania dokumentu, gdyż metryki cyfrowe zawierają informacje o prawach ponownego wykorzystania.
3. Dokumenty PDF oraz metryki cyfrowe w RID pozyskane są ze zbiorów i źródeł danych serwisów internetowych na podstawie zbioru słów kluczowych lub kwerend bez ich wartościowania, np. oceny treści, użytych metod, analizy ilościowej czy jakościowej. Pozyskane są wszystkie możliwe dokumenty z rezultatów wyszukiwania i na tym etapie są sobie równoważne. Wartościowanie treści następuje na etapie korzystania z zarchiwizowanych dokumentów wyszukanych przez kwerendę w aplikacji analitycznej.
4. Dobór aplikacji analitycznej innej niż rekomendowana użytkownikom powinien uwzględniać budowę kwerend, który jest zgodny z językiem zapytań internetowych wyszukiwarki Google. Zwiększy się tym samym efektywności wykorzystania zasobów zgromadzonych w RID oraz stworzy się możliwość udostępniania już wykorzystanych kwerend innym użytkownikom dla ich wykorzystania lub jako podstawa do modyfikacji.

Również funkcjonalność aplikacji analitycznej powinna uwzględniać możliwość tworzenia indeksu z dowolnego cyfrowego zasobu wewnętrznego lub zewnętrznego. Oznacza to, że użytkownik może również tworzyć swoje prywatne indeksy, bez możliwości ich współdzielenia, np. indeksując wiadomości i skrzynki pocztowe, dokumenty elektroniczne, bazy danych, dokumenty XML, archiwa WWW (MIME HTML, MHT), itp. Podczas przeszukiwania użytkownik powinien mieć możliwość zaznaczenia i łączenia wybranych indeksów do przeszukania z wykorzystaniem jednej zbudowanej kwerendy. Swoboda wyboru indeksu do wyszukiwania oznacza, że użytkownik może nie korzystać z indeksu RID a wyłącznie z indeksu lub indeksów prywatnych lub udostępnionych indeksów przez innych użytkowników.

5. Korzystanie z metadanych, które są wbudowane w dokumenty PDF lub z metryk cyfrowych jest opcjonalne. Użytkownicy mogą i stosują alternatywne sposoby korzystania z dokumentów, np. aplikacji do zarządzania bibliografią i cytowaniami, np. EndNote, Mendeley, Zotero.
6. Tworzenie, korzystanie z RID to upowszechnienie dorobku naukowego, badawczego i praktycznego osób nie tylko zewnętrznych, ale również pracowników własnej organizacji. Osiągnięcia twórcze pracowników organizacji są umieszczonej w odpowiedniej strukturze danych i skatalogowane w oddzielne, ale zasoby są włączone do indeksu RID.
7. Korzystanie ze zbioru dokumentów źródłowych oraz stowarzyszonych i praca nad nowymi osiągnięciami w poznaniu naukowym, badawczym i stanu techniki nie wyklucza innych działań na innych zbiorach, baz danych, specjalistycznych serwisów internetowych. Również nie zwalnia się użytkowników, koordynatora i operatora robota cyfrowego przed poszukiwaniem nowych źródeł i zbiorów danych. Powstające serwisy lub nowe funkcjonalności mogą być również włączone do algorytmów programowalnego robota cyfrowego, doskonaląc w ten sposób metodykę pozyskiwania i wykorzystania aktywów intelektualnych.

3.2. Rekomendacje dla koordynatora

Rekomendacje dla koordynatora obejmują następujące zagadnienia: wymagań kompetencyjnych w pełnieniu roli koordynatora, zarządzanie indeksem RID i aktualizacji przyrostowej zasobów. Rekomendacje dla koordynatora są następujące:

1. Zarządzanie zasobami Repozytorium Instytucjonalnego Danych wymaga rozwoju e-kompetencji użytkownika pełniącego rolę koordynatora. Rozszerzone lub pogłębienie e-kompetencje powinno dotyczyć wykonywania zadań i działań w cyklu życia RID, a w szczególności:
 - Myślenia projektowego dla wsparcia strategii organizacji przez identyfikację, zarządzanie, udoskonalanie i rozwijanie kluczowych procesów biznesowych. Dwa spośród najważniejszych czynników w rozwoju procesów biznesowych to innowacyjność oraz potrzeba stałej kontroli jakości zasobów i dostępu do RID w chmurze cyfrowej.
 - Kompetencji interpersonalnych wymagających współpracy z operatorem robota cyfrowego oraz użytkownikami i samoorganizacyjnymi.
 - Stosowania udokumentowanego procesu zarządzania zasobami i systemowego podejścia do weryfikacji i walidacji w cyklu życia produktu cyfrowego – Repozytorium Instytucjonalnego Danych, cyklu życia danych RID.
 - Wykorzystania dostępnych metod, technik narzędzi w procesie wsparcia i pielęgnacji w procesie użytkowania RID.
 - Wykorzystanie metod i technik zarządzania wymaganiami użytkowników, w tym opracowanie struktury założeń bibliograficznych, pozyskiwanie, dokumentowanie i uzgadnianie wymagań w ścisłej współpracy z użytkownikiem i operatorem robota cyfrowego, sporządzania dokumentu zawierającego założenia badań bibliograficznych.
 - Zarządzania indeksem RID (wariant I korzystania z zasobów RID): tworzenie, aktualizacja, import indeksów użytkowników.
 - Zarządzania ciągłością działania: aktualizacja zasobów, wykonanie raportu z pozyskania zasobów, archiwizacja raportów.
 - Zarządzania infrastrukturą chmury cyfrowej w cyklu życia produktu cyfrowego: utworzenie architektury informacji, udostępnianie danych, zarządzanie uprawnieniami, raportowanie aktywności, archiwizacja danych, zarządzanie bezpieczeństwem informacji.

- Ciągłego doskonalenia swojego warsztatu, podnoszenia jakości zasobów i dostępu do Repozytorium Instytucjonalnego Danych (RID) w chmurze cyfrowej.
2. Koordynatorem powinien zostać specjalista, który jest w stanie pozyskać cyfrowe aktywa intelektualne i wykorzystać wiedzę z zakresu zbiorów danych i dokumentów źródłowych. Z założenia systemowego RID, archiwizacji podlegają bardzo szerokokorozumiane skodyfikowane aktywa intelektualne oraz specyficzne informacje dotyczące metryk cyfrowych dla wszystkich pozyskiwanych osiem grup (7 grup aktywów intelektualnych, grupa metryk cyfrowych). Dlatego kompetencje koordynatora muszą dotyczyć wielu zagadnień i znajomości cech poszczególnych grup aktywów intelektualnych.
 3. Importowanie indeksów danych do RID Danych bez dokumentów będących podstawą indeksowania jest praktyką niezalecaną. Indeksacji podlegają nośniki danych ich umiejscowienie, atrybuty nośnika i dokumentu oraz dane tekstowe, np. tekst w dokumencie, w tabelach tekstowych lub liczbowych, opisy tekstowe w obiektach graficznych o ile rozpoznanie tekstu z grafiki jest możliwe. Dlatego zaleca się dodanie dokumentów i utworzenie nowego indeksu lub aktualizację indeksu RID. Włączenie dokumentów źródłowych, zaindeksowanie i wyszukanie odpowiedniego dokumentu pozwala użytkownikom na zapoznanie się z pełną treścią dokumentu oryginalnego. Ma to szczególne znaczenie w zasobach patentowych, gdzie istotą rozwiązań technicznych są różnego rodzaju rysunki, schematy i tabele oraz opisy niepoddające się obróbce OCR. Importowanie tylko indeksów może mieć uzasadnienie dla sprawdzenia lub zwiększenia pewności przeszukania różnych indywidualnych zbiorów użytkowników. W przypadku wyszukania interesującego użytkownika zasobu, zasób ten powinien zostać skopiowany do RID i zaindeksowany przez użytkownika w indeks użytkownika lub w przypadku korzystania z wielu osób zaindeksowane w indeksie RID lub umieszczonym w oddzielnym indeksie i umieszczone na liście indeksów przez koordynatora.
 4. Wykonanie raportu z pozyskania zasobów umożliwia precyzyjną aktualizację zasobów, pokrywając zakres czasowy badań w źródłach i zbiorach danych umożliwiając zastosowanie zarządzania jakością w ciągłości działania. Zarchiwizowanie raportu umożliwia wykorzystania danych w aktualizacji przyrostowej. Fakt systematycznego uzupełniania RID wiąże się z ponowną indeksacją dokumentów.

5. Aktualizacja przyrostowa zasobów w RID wymaga ponownego indeksowania dokumentów. W zależności od zastosowanego programu analitycznego, aktualizacja indeksu może być przyrostowa (indeksowanie tylko zmodyfikowanych i nowych dokumentów) lub ponowna dla wszystkich zasobów z RID. Ponowne indeksowanie zasobów jest czynnością bardzo czasochłonną, zależną od ustawień parametrów indeksowania, np. zapamiętanie pozycji słów z uwzględnieniem odległości między słowami, ustawienie przetwarzania morfologicznego dla kilku języków, ustawienie kilku sposobów kodowania dokumentów, wypakowanie dokumentów z archiwów, rozpoznawanie tekstów, dodawanie katalogów do indeksu, ustawienie dużej liczby długości słów do analizy. Na przykład przy ustawieniach na wartościach optymalnych na zbiorze danych zawierający 99 985 dokumentów o objętości 86,83 GB (ok. 93 GB zajętości na dysku) czas indeksowania na lokalnym komputerze całego indeksu wynosił 58 godz. 1 min i 40 sek. na komputerze z procesorem i7 10-tej generacji przy 32 GB pamięci operacyjnej. Aplikacja Acrobat Pro indeksowała ok. 1h 2906 dokumentów o objętości 8,15 GB (ok. 8,75 GB zajętości na dysku). W czasie trwania indeksacji całego zbioru uniemożliwia się korzystanie z zasobów RID. Dlatego zalecaną pracą przy aktualizacji zasobów jest indeksowanie przyrostowe lub indeksowanie zasobów na innym komputerze i importowanie indeksu wraz z danymi źródłowymi. Aktualizacja indeksów może zostać zrealizowana w ściśle określonych w godzinach lub w momencie, gdy użytkownicy nie korzystają z RID. Również, jeżeli aplikacja indeksująca posiada harmonogram automatycznej aktualizacji to można ustawić parametry wykonywania indeksacji, np. o określonej godzinie lub podczas bezczynności aplikacji.
6. Różne aplikacje analityczne lub czytniki dokumentów PDF umożliwiają tworzenie i wyszukiwanie na indeksach dokumentów. Różnice w parametrach ustawień programu oraz szybkość indeksowania zasobów, a także czas wyszukiwania są znaczące. Alternatywą dla wyszukiwania w indeksach jest wyszukiwanie bezpośrednio w zbiorach dokumentów PDF (wariant II korzystania z zasobów RID). Jednak czas przeszukiwania jest bardzo długi i może wynosić dla jednego zapytania nawet kilka godzin. Uruchamianie kilku czytników PDF i wyszukiwanie równoległe powoduje znaczne obciążenie zasobów pamięci oraz procesora, co skutkuje opóźnieniem reakcji komputera, na którym te procesy są uruchomione.

3.3. Rekomendacje dla operatora robota cyfrowego

Rekomendacje dla operatora robota cyfrowego obejmują zagadnienia wymagań kompetencyjnych, zarządzania programowalnego robota cyfrowego RID w cyklu jego życia i są następujące:

1. Aktualizacja przyrostowa zasobów, tworzenie dokumentów stowarzyszonych i zarządzanie programowalnym robotem cyfrowym RID wymaga rozwoju e-kompetencji użytkownika pełniącego rolę operatora robota cyfrowego. Rozszerzone lub pogłębienie e-kompetencje powinno dotyczyć wykonywania zadań i działań w cyklu życia programowalnego robota cyfrowego, a w szczególności:
 - Umiejętności analizy procesów biznesowych przez identyfikację, dokumentowanie optymalizację kluczowych procesów biznesowych.
 - Kompetencji interpersonalnych wymagających współpracy z koordynatorem oraz użytkownikami.
 - Kompetencji samo organizacyjnych w zarządzaniu czasem.
 - Znajomości interfejsu użytkownika dla wyspecyfikowanych zbiorów danych i dokumentów źródłowych, translacji dokumentów, w tym w rozwiązania z wykorzystaniem API oraz rozwiązań API jako formy komunikacji pomiędzy aplikacjami (budowanie kwerendy – pozyskanie – przetworzenie – ocena).
 - Wykorzystania dostępnych metod, technik narzędzi w procesie wsparcia i pielęgnacji w procesie użytkowania RID ze współpracy z koordynatorem.
 - Zarządzania ciągłością działania: aktualizacja zasobów, wykorzystania danych z raportu z pozyskania zasobów.
 - Zarządzania infrastrukturą chmury cyfrowej w cyklu życia produktu cyfrowego: utworzenie architektury informacji, udostępnianie danych, zarządzanie uprawnieniami, raportowanie aktywności, archiwizacja danych, zarządzanie bezpieczeństwem informacji.
 - Ciągłego doskonalenia swojego warsztatu, podnoszenia jakości zasobów i dostępu do Repozytorium Instytucjonalnego Danych (RID) w chmurze cyfrowej.
2. Operatorem robota cyfrowego powinien zostać specjalista, który posiada doświadczenie implementacyjne w technologii RPA, w szczególności potrafi odwzorować biznesowe przypadki użycia, reguł biznesowych dla różnych

rodzajów danych, zaprojektować scenariusze awarii oraz mieć wsparcie ze strony dostawcy platformy RPA.

3. Zasoby pozyskane z wykorzystaniem programowalnego robota cyfrowego i udostępniane użytkownikom są przedmiotami handlowymi dystrybuowanymi na określonych prawach autorskich przez dostawców. Licencje określają możliwości ingerencji i sposób oraz warunki udostępniania innym użytkownikom. Działania podejmowane w stosunku do wykorzystania zasobów muszą respektować zarówno prawa autorskie jak i handlowe do utworu.
4. Pozyskiwanie zasobów z multidyscyplinarnych baz danych może odbywać z wykorzystaniem interfejsu użytkownika lub interfejsu programowania aplikacji. Efektywniejszym sposobem jest możliwość bezpośredniego pozyskiwania danych w czasie rzeczywistym z wykorzystaniem API. Na przykład w serwisie Scopus można skorzystać z *Scopus API*, *SciVal API* bezpłatnie na użytek niekomercyjny dla subskrybentów. Dla oceny ilościowej i jakościowej pobieranie danych jest szybsze i powtarzalne, np. wskaźników dla instytucji i naukowców przez *SciVal API*. W przypadku aktualizacji zasobów można w prosty sposób zaprogramować interwał czasowy pobrania cyklicznego np. kwartalnego, półrocznego lub dowolnego przedziału czasu. Mniej efektywniejszym sposobem jest eksport danych lub kopiowanie danych bezpośrednio z okna przeglądarki internetowej do dokumentu.
5. Dużym utrudnieniem w automatyzacji pozyskiwania zbiorów ze specjalistycznych serwisów jest zastosowanie techniki Captcha oraz autoryzacja. Dla korzystania z niektórych danych jest wymagane jest sprawdzenie czy użytkownikiem jest człowiek, a nie robot programowy. Najczęściej wykorzystywane są zdjęcia lub obrazki wyświetlane losowo, a odczytanie ich jest proste przez człowieka natomiast dla robota rozpoznawanie obrazów lub tekstów zniekształconych jest znacznie utrudnione. Jednak techniki Captcha często utrudniają, a nawet uniemożliwiają odczytanie zawartości obrazka także ludziom pełnosprawnym. Ma to znaczenie dla osób z niepełnosprawnościami, gdyż Captcha może uniemożliwić korzystania z serwisu i zarazem z danych. Zastosowanie mechanizmów autoryzacji dla potwierdzenia subskrypcji danych utrudnia proces pozyskania zasobów cyfrowych. Każde pozyskanie zasobów wymaga potwierdzenia tożsamości cyfrowej, a to skutkuje za każdym razem zmianą hasła dostępu do serwisu.

6. Trend związany ze zwiększeniem dostępności danych, tj. otwartym dostępem dla danych (otwarte cyfrowe archiwa i bazy danych) powoduje, że należy uwzględniać rodzaj użytego nośnika danych. Coraz mniej jest przypadków serwisów lub baz danych udostępniających treść dokumentu tylko na ekranie, bez możliwości skopiowania, edycji, komentowania czy też pobrania danych i zapisania na określony nośnik danych. Znacząca większość serwisów zawiera zasoby informacyjne i posiada funkcjonalność zapisu dokumentów w różnych formatach. Preferowanym formatem pliku do pozyskania jest dokument w formacie PDF, który jest standardem wymiany elektronicznych kopii dokumentów opartych na stronach. PDF to format pliku stworzony przez firmę Adobe Systems na początku lat 90. XX wieku, którego głównym celem było stworzenie dokumentu cyfrowego zasadniczo podobnymi do ich papierowych odpowiedników. Zalety PDF obejmuje:

- kompatybilność na wszystkich platformach,
- możliwość tworzenia PDF z dokumentów źródłowych (skanowanie dokumentów lub obiektów),
- łatwość tworzenia PDF przez zapis bezpośrednio do PDF lub przez polecenie eksportu danych do PDF,
- możliwość zapisania metadanych w odpowiednich polach dokumentu PDF,
- podatność do modyfikacji i przydatność do ponownego użycia,
- możliwość osadzenia w nim innych plików,
- wyszukiwanie pełnotekstowe, włączając również metryki cyfrowe i indeks danych,
- respektowanie praw autorskich,
- wymagania dotyczące oprogramowania do przeglądania, edycji, modyfikacji plików lub konwersji na inne formaty.

Bezpośrednie pozyskanie dokumentów w formacie PDF zmniejsza pracochłonność w dostosowaniu różnego formatu plików do standardu przechowywania plików w RID.

7. Dokumenty PDF pozyskane przed zapisem do RID podlegają sprawdzeniu poprawności odczytu i edycji w sposób automatyczny. W przypadku pozytywnego zweryfikowania dokumentu zostaje on zapisany w zasobach RID. Sprawdzenie zabezpieczeń jest istotne, gdyż eliminuje się późniejsze błędy podczas pracy użytkownika nad dokumentem oraz umożliwia się translacji lub przekształceń przez użytkownika.

8. Użytkownik może dokonywać edycji dokumentu PDF przechowywanego w RID przez wstawienie obiektów w treści dokumentu, np. recenzje, komentarze i znaczniki, które mogą być dalej modyfikowane lub usuwane przez użytkowników. Wstawienie lub edycja recenzji, komentarzy, znaczników lub metadanych do dokumentu nie powoduje automatycznej indeksacji zbioru i tym samym, zmiany w wyszukiwaniu nie są uwzględniane. Wykorzystanie aplikacji analitycznej Acrobat⁶⁰ umożliwia hostowanie udostępnionych recenzji w SharePoint lub Office 365. W przypadku korzystania z Acrobat Pro istnieje dodatkowa możliwość udostępniania recenzji. Komentarze użytkowników są przechowywane w repozytorium na serwerze wewnętrznym. W przypadku recenzji udostępnionej użytkownicy mogą dołączyć do recenzji, udostępniać swoje komentarze, śledzić recenzje i otrzymywać regularne aktualizacje. Recenzenci są powiadamiani o nowych komentarzach w miarę ich dodawania i mogą przeglądać komentarze innych recenzentów i odpowiadać na nie. Ze względu na różne wersje i licencje oprogramowania zaawansowane funkcje opisane są na stronie producenta oprogramowania Acrobat⁶¹.
9. Wśród specjalistycznych serwisów internetowych przechowujących dokumenty źródłowe są przypadki udostępniania wszystkich zgromadzonych zasobów, ale bez możliwości selekcji lub wyszukiwania dokumentów, metadanych, gdyż serwis nie zawiera systemu wyszukiwawczego w dostępie on-line, a jedynie po zalogowaniu, np. serwis *Journal of Business Models*⁶². W takim przypadku są trzy rozwiązania. Pierwsze i drugie rozwiązanie dotyczy automatyzacji pozyskiwania dokumentów przez indywidualne zaprogramowanie robota cyfrowego. W pierwszym przypadku program robota cyfrowego wyszukuje na stronie hiperłącza i pobiera dokumenty, w drugim przypadku należy zaprogramować robota, który uruchomi menedżera pobierania plików i zapisze pozyskane pliki w odpowiednie miejsce w architekturze informacji RID. Trzecim rozwiązaniem jest założenie konta w serwisie i zaprogramowanie robota z podaniem hasła dostępu. Kolejnym nietypowym serwisem jest baza „Polona” Biblioteki Narodowej⁶³, który posiada system wyszukiwaczy, którego wyniki wyszu-

⁶⁰ Adobe Systems Incorporated, 345 Park Avenue, San Jose, California 95110, USA.

⁶¹ Introduction to Acrobat, <https://helpx.adobe.com/acrobat/user-guide.html>, data dostępu 5.01.2023 r.

⁶² *Journal of Business Models*, <http://journalofbusinessmodels.com>, data dostępu 5.01.2023 r.

⁶³ Cyfrowa Biblioteka Narodowa Polona, <http://www.polona.pl/dlibra>, data dostępu 5.01.2023 r.

kania są prezentowane wyłącznie na ekranie. Brak możliwości eksportu metadanych powoduje, że serwis nie jest interoperacyjny, przyjazny w użytkowaniu i w zasadzie nieprzydatny w pozyskiwaniu danych zarówno manualnie jak i w sposób zautomatyzowany.

10. Masowe, bezpłatne pozyskanie zasobów, np. patentów *USPTO*, *esp@cenet* oraz *Google Patents* nie jest możliwe, gdyż bazy nie oferują możliwości masowego ich pozyskania. Należy wykorzystać aplikacje transakcyjne oferujące automatyzację ich pozyskania, np. aplikacja *HTTrack Website Copier*⁶⁴. Warunkiem użycia jest brak ograniczeń licencyjnych. Najczęściej nie jest dozwolone masowe pozyskanie bardzo dużej liczby dokumentów w krótkich odstępach czasu, np. w serwisie *Semantic Scholar*⁶⁵ dostęp do zasobów przez API jest bezpłatny, ale występuje ograniczenie pobiera, co skutkuje odpowiedzią serwera statusem HTTP 429 „Zbyt wiele żądań”, jeśli limit zostanie przekroczony (100 żądań na 5 minut na adres IP). W serwisie *CORE* dostęp przez API jest bezpłatny, jeżeli nie ma żądania zbiorczego lub liczba żądań nie przekracza 5 pojedynczych żądań na 10 sekund.
11. Pozyskiwanie dokumentów z multidyscyplinarnych baz danych może odbywać się w modelu usługi płatności za przygotowanie masowych danych, np. *Scopus Custom Data* lub w modelu subskrypcyjnym, co wymusza odpowiednie ustawienie parametrów dostępu do zasobów, np. WoS.
12. W celu zautomatyzowania procesów na etapie badań bibliograficznych można:
 - Zakupić licencję dla programowalnego robota, która kwotowo jak na polskie warunki jest kosztem bardzo znacznym i praktycznie rozwiązaniem nieakceptowanym finansowo. I tak np. licencja roczna na *Blue Prism* to wydatek ok. 8 tys. funtów dla robota z opłatą utrzymaniową, cena licencji na *UiPath* zaczyna się od 6000 USD na robota w wersji nadzorowanej, ceny rozwiązań firmy *Automation Anywhere* zaczynają się od 5000 USD.
 - Wykorzystać darmowe rozwiązania platform RPA lub rozwiązania klasy RPA w modelu *Open Source*, np. *Robocorp* oferuje zbiór bibliotek *Open Source* i narzędzi do RPA dla osób indywidualnych i małych

⁶⁴ *HTTrack Website Copier*, <https://www.httrack.com>, data dostępu 5.01.2023 r.

⁶⁵ *Semantic Scholar*, <https://www.semanticscholar.org/>, data dostępu 5.01.2023 r.

firm, *Robot Framework* to ogólna platforma automatyzacji typu *Open Source*.

- Wykorzystać bezpłatne wersje próbne lub specjalnego wykorzystania: *VisualCron* 45 dni, *AssistEdge Community Edition* oferuje bezpłatną platformę automatyzacji dla studentów i niezależnych programistów, *UiPath Community Edition* dla indywidualnych programistów RPA i małych zespołów, *UiPath* oferuje produkt freemium, *Blue Prism* do 30 dni lub w wersji edukacyjnej 180 dni, *Workfusion RPA Express* wyłącznie do użytku indywidualnego.
- Wykorzystać języki programowania, np. *Automagica* udostępnia bibliotekę *Open Source Phyton RPA* na *Github*, *Taskt* to klient automatyzacji procesów zbudowany na platformie *.NET Framework* w języku C#.
- Skorzystać z ogólnodostępnych lub licencjonowanych pakietów biurowych zawierających możliwość programowania w środowisku programistycznym, np. wykorzystanie z wbudowanego języka programowania deklaratywnego i VBA w arkuszu kalkulacyjnym MS Excel.

Przedstawione rekomendacje dla trzech grup użytkowników w zarządzaniu zasobami ludzkimi dla zapewnienia rozwoju e-usług RID i robotyzacji procesów wykonywalnych manualnie są jednocześnie przesłanką zrealizowania zdefiniowanego dodatkowego celu pracy.

W celu zachęcenia pracowników do dołączenia się do RID można w pierwszej kolejności, pokazać korzyści z indeksowania zbiorów i sposób ich przeszukiwania na lokalnych zasobach na przykładzie zgromadzonych przez użytkownika zasobów. W ten sposób użytkownik znając własne zasoby jest w stanie docenić pracę analityczną oprogramowania, sposób wydobywania wiedzy i korzystania z raportów. Drugiej kolejności należy zademonstrować przechowywanie zasobów w aplikacji MS TEAMS. Microsoft Teams w wersji bezpłatnej umożliwia pracę do 500 tys. użytkowników w organizacji, w zakresie przechowywania plików udostępnia 2 GB na użytkownika i 10 GB magazynu udostępnionego. W ten sposób na znanej użytkownikowi platformie chmurowej można wykazać praktyczne zastosowanie koncepcji RID w korzystaniu i udostępnieniu zasobów, włączając również inne funkcje, jak: grupowe i indywidualne (1:1) połączenia audio i wideo online, spotkania na kanałach, udostępnianie ekranu, zaplanowane spotkania, nagrywanie spotkań.

Zakończenie

Zastosowanie programowalnych robotów cyfrowych w automatyzacji pozyskiwania zasobów dla utworzenia archiwalnego zbioru danych z badań jakościowych stwarza nowe możliwości specjalistom zainteresowanych prowadzeniem badań bibliograficznych dla potrzeb oceny stanu wiedzy naukowej, badawczej i praktycznej oraz w pobudzaniu innowacyjności opartej na ocenie ilościowej i jakościowej zgromadzonych zasobów. Przedstawiony autorski sposób wykorzystania robotyki w badaniach bibliograficznych skutecznie łączy rekurencyjne pozyskiwanie zasobów w sposób zautomatyzowany z aktualizacją zgromadzonych zasobów oraz mapowaniem tematycznym zainteresowanych użytkowników w zespołach monodyscyplinarnych i interdyscyplinarnych. Zgromadzone zasoby w instytucjonalnym repozytorium są indeksowane wg parametrów i ustawień aplikacji indeksującej. Instytucjonalne repozytorium nie jest statyczną bazą dokumentów a dynamiczną platformą danych w chmurze cyfrowej z możliwością odkrywania wiedzy i powiązań niedostępnych w multidyscyplinarnych bazach danych lub na specjalistycznych serwisach internetowych. Jest to masowy zbiór danych typu *Open Access* lub na licencji *Creative Commons* z różnego rodzaju aplikacjami umożliwiającymi wzajemne uczenie się, dzielenie wiedzą w grupach lub zespołach specjalistów. Dodatkowym atutem wydobywania wiedzy jest możliwość łączenia osobistych zaindeksowanych zasobów przechowywanych w lokalnych zasobach z zaindeksowanymi zasobami repozytorium tworząc jedną indywidualną przestrzeń przeszukania wg jednego schematu zapytań. Stworzenie i wykorzystanie instytucjonalnego repozytorium danych jest działaniem komplementarnym a nie konkurencyjnym w stosunku do innych serwisów udostępniających dokumenty źródłowe lub metryki cyfrowe. Dla specjalistów, którzy nie posiadają właściwych e-kompetencji w przeszukiwaniu wielu różnych rozproszonych repozytoriów, o nieujednoliconej strukturze danych, sposobach dostępu, w tym autoryzacji, prawach własności czy sposobów odpłatności stanowi problem zarówno metodyczny, organizacyjny jak i techniczny. Problem ten jest rozwiązywalny przez wykorzystanie zasobów RID i zastosowanie jednego zunifikowanego schematu wyszukiwania informacji w wielu obszarach wiedzy. Tym samym jest to innowacyjny, uniwersalny pod względem zainteresowań odbiorców sposób odkrywania wiedzy w grupach lub zespołach badawczych. Korzystanie z zasobów RID powoduje oszczędność czasu w zbieraniu informacji i pozwala na skoncentrowaniu się na wykonywaniu zapytań bibliograficznych, ocenie ilościowej, jakościowej wyników oraz na konsumpcji tej wiedzy. W tym sensie

potwierdzają się badania, które wykazują, że robotyzacja procesów umożliwia również efektywną pracę specjalistów o mniejszych e-kompetencjach.

Każde repozytorium ma ograniczoną wartość w eksploracji i ekstrakcji danych, w aspektach poznawczych i przyswajania wiedzy w zarządzaniu grupami lub zespołami badawczymi. Najważniejszymi wartościami dodanymi dla użytkowników są następujące zagadnienia:

1. Automatyzacja pozyskiwania aktywów intelektualnych dla aktualizacji zasobów przez programowalnego robota cyfrowego w siedmiu grupach aktywów intelektualnych, cyfrowych metryk umożliwiała utworzenie produktu cyfrowego RID rozumianej pojęciowo jako bazy danych. Baza zapewnia jednoczesny zdalny dostęp *on-line* do zawartości w formie e-usługi, tj. edytowalnych pełnotekstowych plików PDF z metadanymi oraz do dokumentów zawierających metryki cyfrowe i do dokumentów przetłumaczonych. Istnieje również możliwość replikacji zasobów do innych internetowych miejsc udostępniania lub przerzucenia ich na lokalne nośniki pamięci i korzystanie w trybie *off-line*.
2. Bezpośredni dostęp do wszystkich zaindeksowanych zasobów w wariancie I (z indeksami danych) pozwala użytkownikom na zastosowanie jednolitego schematu formułowania zapytań bibliograficznych. Użytkownicy uzyskują w sposób całościowy wgląd w aktywa intelektualne, otrzymują analizę ilościową oraz jakościową. Technicznie, otrzymują wykonaną pełną analizę treści z analizą morfologiczną pełnotekstowych plików PDF wszystkich dokumentów zgromadzonych w RID i tym samym, w sposób analityczny zwiększono im dostęp do zgromadzonej skodyfikowanej wiedzy. Korzystanie z indeksów danych RID jest fakultatywne. Użytkownicy mogą w sposób bezpośredni przeszukiwać zasoby źródłowe z wykorzystaniem rekomendowanej aplikacji analitycznej lub niezależnych aplikacji, które posiadają funkcjonalność przeszukiwania w wielu plikach PDF. Zmiana przyzwyczajeń użytkownika jest barierą do wykorzystania zasobów RID, dlatego rozpoczęcie pracy z RID powinno być poprzedzone zaindeksowaniem zasobów zgromadzonych przez użytkownika. W ten sposób przez przykład na własnych zasobach, z których się korzysta prowadzi do zrozumienia sposobu pracy z wykorzystaniem aplikacji analitycznej. Następnym etapem jest włączenie kolejnych indeksów do przeszukania. Dalsze postępowanie powinno być skierowanie do wykorzystania zasobów w chmurze cyfrowej i korzystania z e-usług *on-line*.
3. Dostęp do systemu wyszukiwawczego w aplikacji analitycznej, umożliwiający użytkownikom budowanie kwerend zgodnych z językiem zapytań

internetowych wyszukiwarki Google z analizą morfologiczną w kilku językach w trybie udostępniania indeksów w chmurze cyfrowej lub przez serwer wyszukania zdalnego ułatwia formułowanie i stosowanie zapytań bibliograficznych. Analiza morfologiczna zasobów jest unikalną cechą zasobów powiązaną z systemem wyszukiwawczym i stanowi *know-how* nie tylko RID, ale jest również wartością unikalną niektórych specjalistycznych serwisów internetowych.

4. Możliwość łączenia i wiązania repozytoriów użytkownika, udostępniania ich lub włączenia do instytucjonalnego repozytorium stanowi dodatkową wartość pożądaną w utworzeniu jednej indywidualnej przestrzeni przeszukiwania wg tego samego schematu zapytań.
5. Analityczny dostęp do zasobów RID przez aplikację umożliwia drążenia wyników zapytań lub stopniowe doprecyzowanie zapytań bibliograficznych, co następnie pozwala na uzyskanie większej trafności uzyskanych wyników i uzyskania szybszej satysfakcji z badania bibliograficznego. Drążenie wyników zapytań stanowi unikalną cechą RID w zakresie systemu wyszukiwawczego, niespotykaną w innych serwisach specjalistycznych.
6. Wykorzystanie narzędzia komentowania, zadawania pytań przez mechanizmy chmury cyfrowej pozwala na dzielenie się i wymianę wiedzy. Zwłaszcza, że istnieje dodatkowa możliwość wstawiania obiektów w treści dokumentu PDF, np. recenzje, komentarze, zakładki i znaczniki, które mogą być dodawane, modyfikowane lub usuwane przez użytkowników.
7. Dostarczenie użytkownikom produktów programowych z zasadami ich wykorzystania, z rekomendacjami w wykorzystaniu oprogramowania w rozwiązywaniu konkretnych problemów w sposób przyjazny pozwala na szybsze zapoznanie się z możliwościami pozyskiwania wiedzy przy mniejszej e-kompetencji. Pozwala to na ocenie stanu wiedzy niezależnie od doświadczenia użytkownika. Pracownik nowozatrudniony po zapoznaniu się dokumentacją RID i przypadkami użycia opartymi na doświadczeniach użytkowników (UX) jest w stanie skutecznie wykorzystać zgromadzone zasoby, co pracownik z wieloletnim doświadczeniem. Efekt zrównywanie szans jest kolejną wartością dodaną RID. Użytkownicy do swojej dyspozycji otrzymują produkt cyfrowy do wykorzystania wysokiej jakości, możliwej do uzyskania w przy zastosowaniu wybranych rozwiązań informacyjno-organizacyjnych i programistycznych.
8. Możliwość użytkowania zasobów RID nie wyklucza możliwości korzystania ze specjalistycznych serwisów, które udostępniają zasoby z zaawansowanymi możliwościami: łącznia i grupowania danych w specjalistyczne

zbiory danych, kanałów badawczych przez spersonalizowanie rekomendacji opartych na sztucznej inteligencji, np. Semantic Scholar.

9. Z założenia systemowego, operator robota cyfrowego ma za zadanie zaprogramować robota na pozyskiwanie konkretnych zasobów oraz nadzorować jego pracę. Rolę operatora mogą przejąć również inni użytkownicy. Mogą spersonalizować zadania robota cyfrowego dla własnych zasobów i zastosować w dowolnym momencie do aktualizacji tylko określonych zasobów. W ten sposób wykorzystanie zaprogramowanego robota z parametryzacją zwiększa efektywność jego wykorzystania.
10. Istnieje sześć głównych powodów, dla których zastosowanie RPA powoduje istotne korzyści w procesie prowadzenia badań bibliograficznych z wykorzystaniem RID: (1) skrócenie czasu pozyskiwania dokumentów źródłowych i stowarzyszonych, (2) zapewnienie ciągłości w aktualizacji zasobów, skrócenie czasu aktualizacji, (3) zwiększenie ilości gromadzonych zasobów, (4) poprawienie efektywności badań bibliograficznych przez skrócenie czasu procesu ich prowadzenia, (5) poprawienie precyzjności pozyskiwania, (6) zwiększenie dostępności e-usług i informacji, uzyskiwanie znacznego przybliżenia narracji realiów uczestników badań bibliograficznych.

Opisane wartości dodane dla użytkowników nie wyczerpują w pełni zagadnienia wykorzystania zasobów RID dla grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów.

Reasumując, tematyka badawcza obejmowała aspekty metodyczne, przedsięwzięcia projektowe, praktyczne związane z automatyzacją prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej, praktycznej dla grup i zespołów specjalistów, wspomagana programowalnymi robotami cyfrowymi. W sposób kompleksowy zaprezentowano sposób zarządzania użytkownikami przez powiązanie roli operatora robota cyfrowego z rolami użytkowników którym przyporządkowano obowiązki, przydzielono uprawnienia i odpowiedzialność. W ujęciu struktur zarządzania w organizacji, jest to struktura sieciowa oparta na funkcjach równorzędnych pełniących ról. Typowe funkcje zarządcze, tj. planowanie, organizowanie, przewodzenie, kontrolowanie zostały przeniesione na poziom wyższy, poza zdefiniowane role w RID.

Również w sposób kompleksowy przedstawiono wszystkie niezbędne procesy w cyklu życia RID, które stanowią kompozycję zamkniętą w metodyce

zautomatyzowanego prowadzenia badań bibliograficznych dla grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów. Na metodykę prowadzenia badań bibliograficznych złożyły się wzajemne zależne cykle życia RID: produktu cyfrowego RID, badań bibliograficznych na zasobach RID, danych RID, programowalnego robota cyfrowego, aktualizacji przyrostowej danych.

Zaproponowane rozwiązania informacyjno-organizacyjne i programistyczne dla problemów przedstawionych we wprowadzeniu monografii w postaci opracowanej metodyki zautomatyzowanego prowadzenia badań bibliograficznych z wykorzystaniem RPA dla grup, zespołów monodyscyplinarnych lub interdyscyplinarnych specjalistów jest jednocześnie przesłanką zrealizowania zdefiniowanego celu naukowego pracy. Większość zaprezentowanych problemów oraz barier w prowadzeniu badań bibliograficznych, które uniemożliwiają szybkie i wydajne ich przeprowadzenie zostały likwidowane lub zniwelowane do poziomu akceptowalnego przez użytkowników.

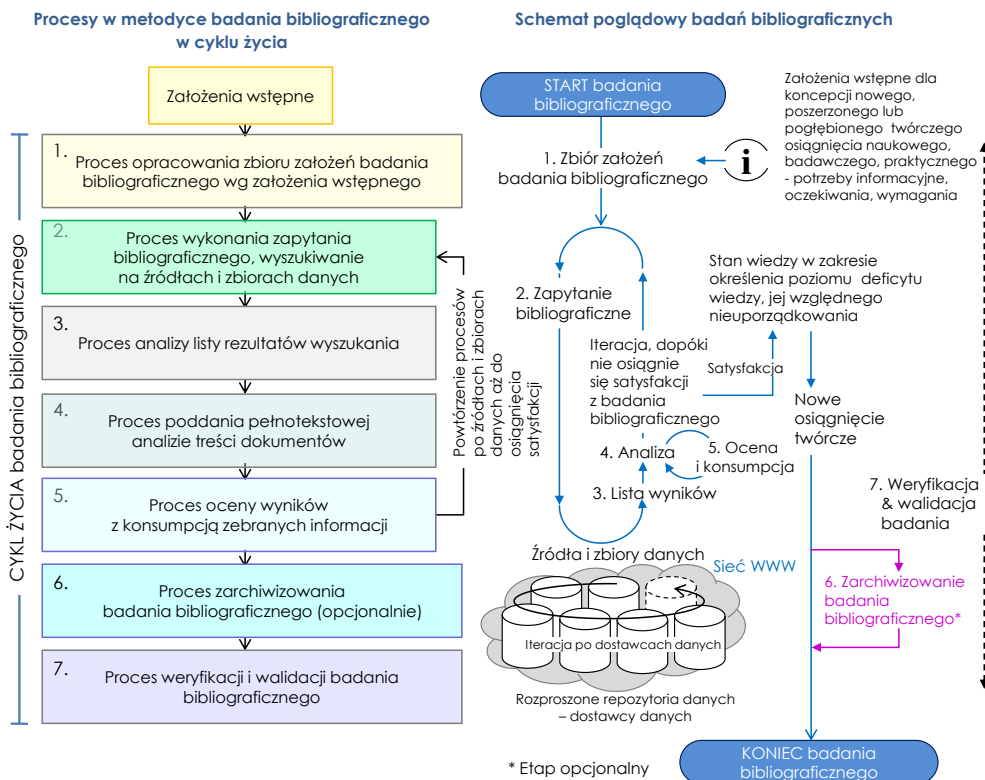
Głównymi efektami pracy na poziomie teoriopoznawczym i aplikacyjnym, które można przypisać do dyscypliny Nauki o zarządzaniu i jakości, należy zaliczyć:

- Opracowanie metodyki zautomatyzowanego prowadzenia badań bibliograficznych przez grupy lub zespoły specjalistów, która obejmuje opracowanie procesów, działań w cyklu życia produktu programowego RID, ze szczególnym uwzględnieniem procesów, które mogą być zautomatyzowane przy wykorzystaniu programowalnego robota cyfrowego.
- Opracowanie modelu zarządzania zasobami ludzkimi w ujęciu procesowym. Wypracowanie rekomendacji zarządzaniu zasobami ludzkimi dla zapewnienia rozwoju robotyzacji procesów dotyczących prac manualnych.

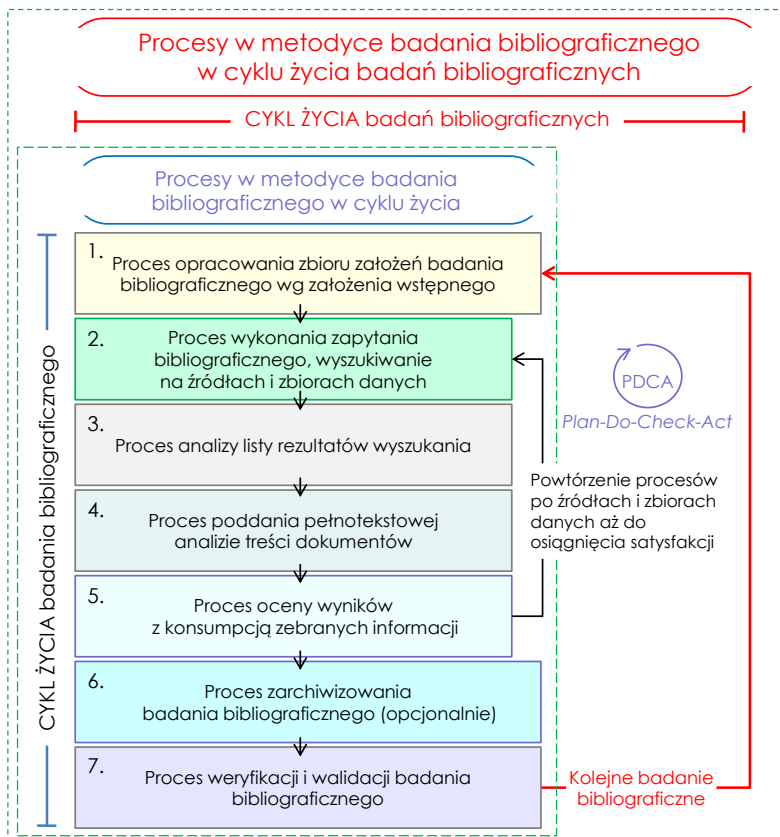
W zarządzaniu zasobami ludzkimi, całokształt podejmowanych przedsięwzięć innowacyjnych umożliwiających sprawne osiągnięcie celów dla nowych osiągnięć twórczych leży w domenie nauki o zarządzaniu i jakości. Wykorzystanie dorobku i wiedzy z nauk o zarządzaniu ukierunkowuje postępowanie ludzi do osiągnięcia wyznaczanych celów przy efektywnym wykorzystaniu zasobów. Właściwe użyte zasoby ludzkie, finansowe, rzeczowe, informacyjne oraz należyty dobór metod, narzędzi technik w RID przyczyniły się do zauważalnego skrócenia cyklu życia badania bibliograficznego i stworzyły specjalistom komfort psychiczny oraz komfort pracy pobudzający innowacyjność opartą na ocenie ilościowej i jakościowej zgromadzonych zasobów.

Załącznik A. Uogólniona metodyka badania bibliograficznego w cyklu życia badań bibliograficznych

Uogólniona metodyka badania bibliograficznego

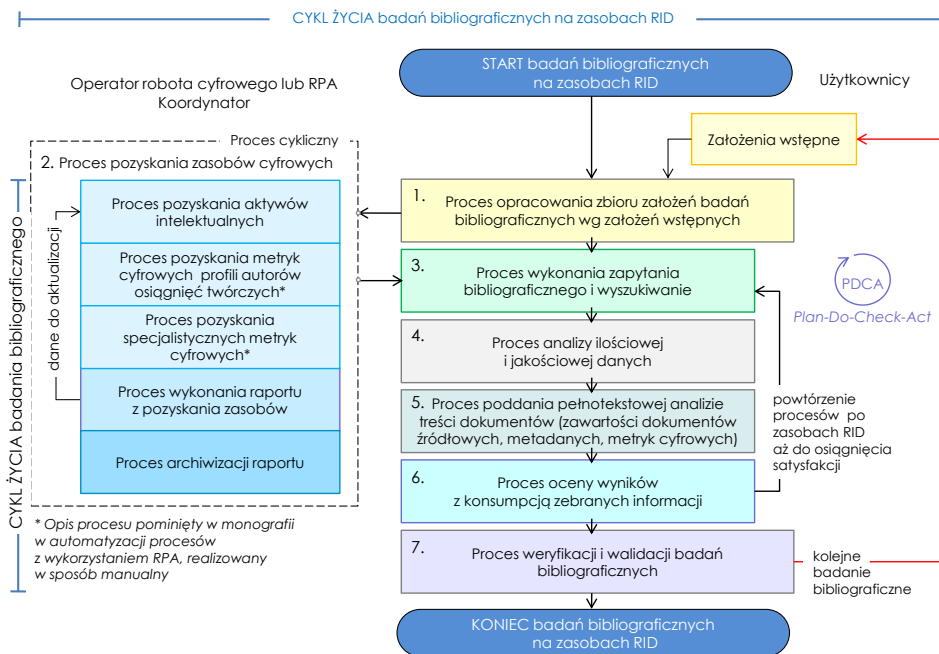


Uogólniona metodyka badań bibliograficznych w cyklu życia badań w ujęciu procesowym

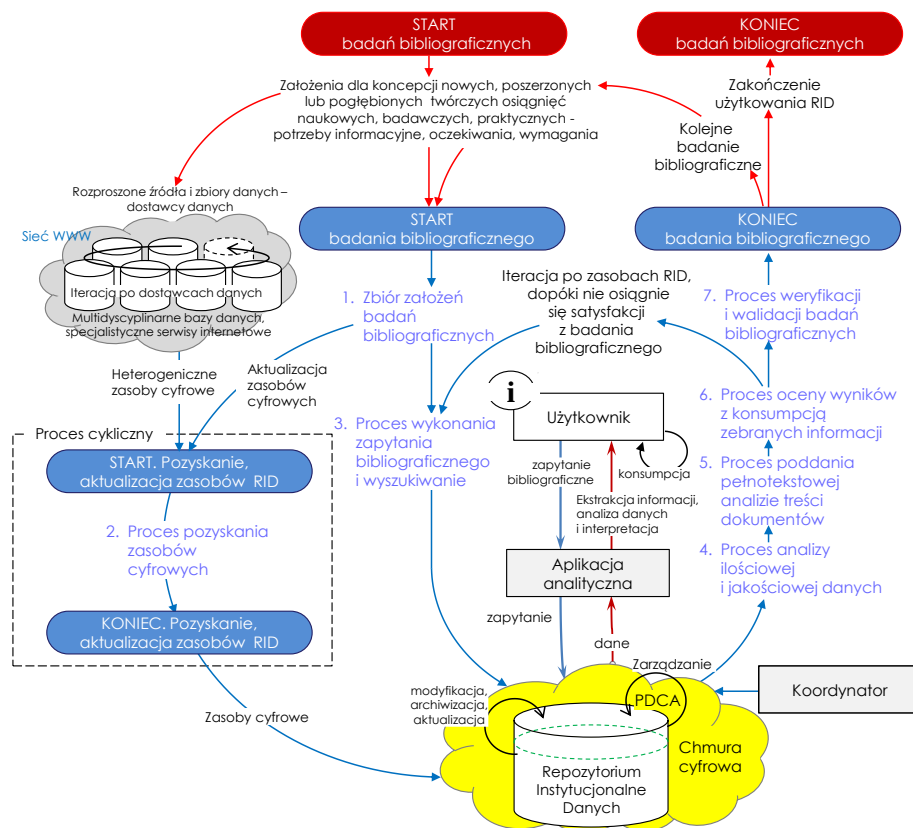


Załącznik B. Metodyka zautomatyzowanego prowadzenia badań bibliograficznych z wykorzystaniem RPA

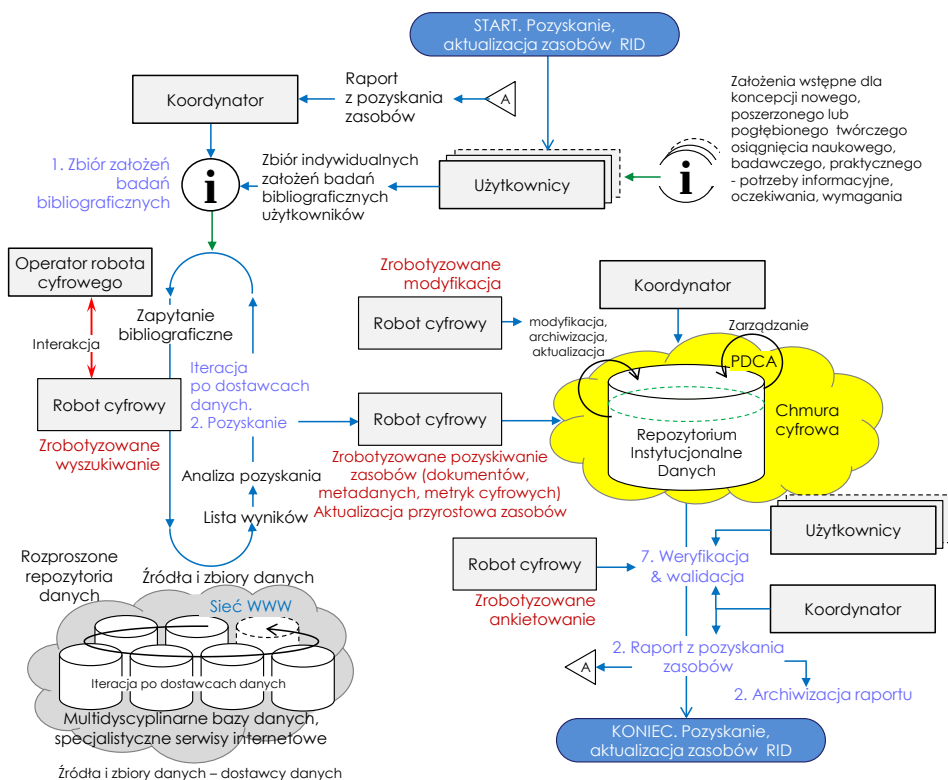
Metodyka zautomatyzowania prowadzenia badań bibliograficznych w ujęciu procesowym



Schemat poglądowy metodyki zautomatyzowania prowadzenia badań bibliograficznych

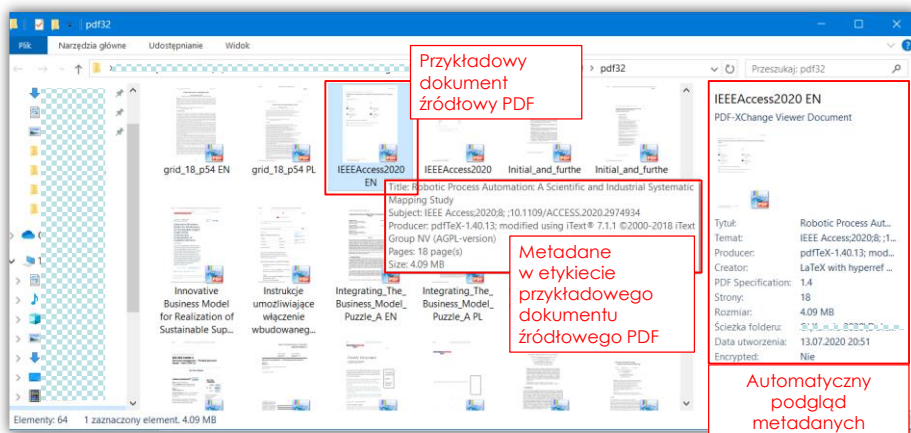


Proces cykliczny pozyskania, aktualizacji zasobów w chmurze cyfrowej

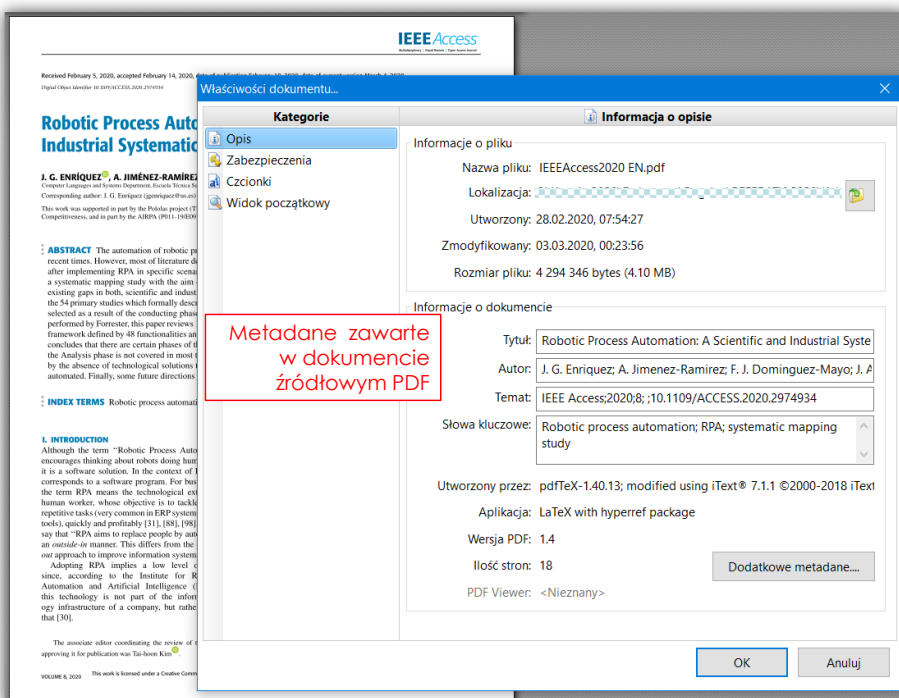


Załącznik C. Przykłady funkcjonalności aplikacji analitycznej w badaniach bibliograficznych na zasobach RID

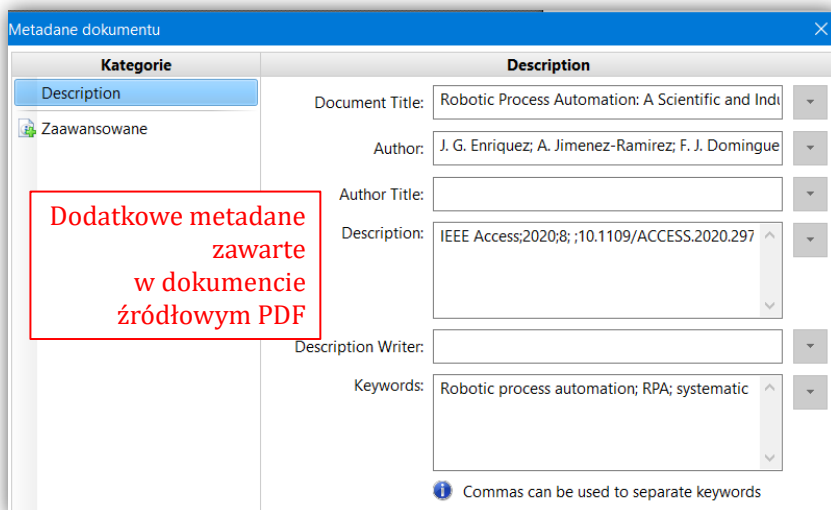
Dokumenty źródłowe PDF pozyskane od dostawców danych



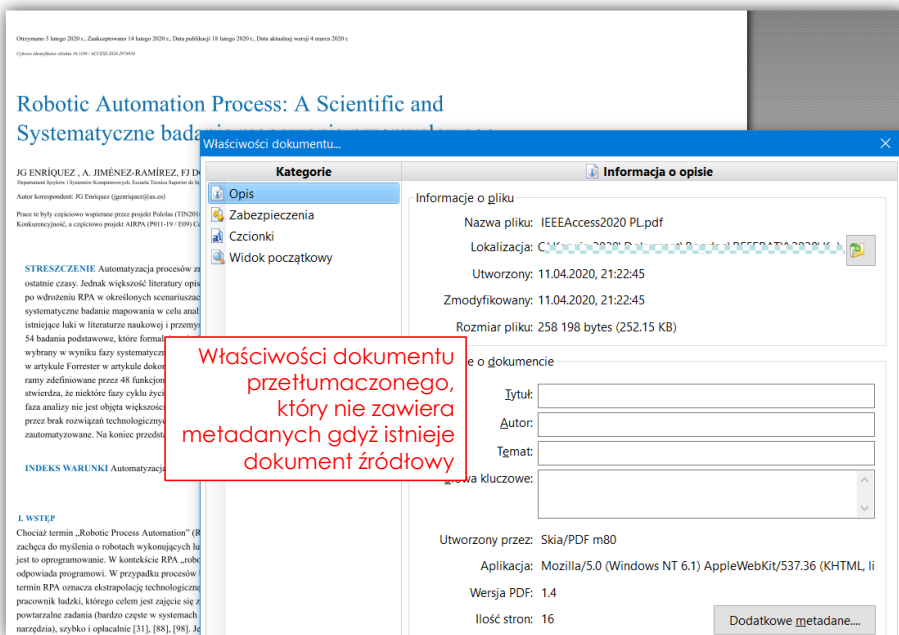
Dokument źródłowy PDF pozyskany od dostawcy danych z metadanymi



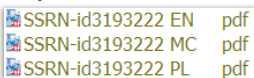
Dokument źródłowy PDF pozyskany od dostawcy danych z dodatkowymi metadanymi



Przełumaczony dokument PDF, słowarzyszony z dokumentem źródłowym

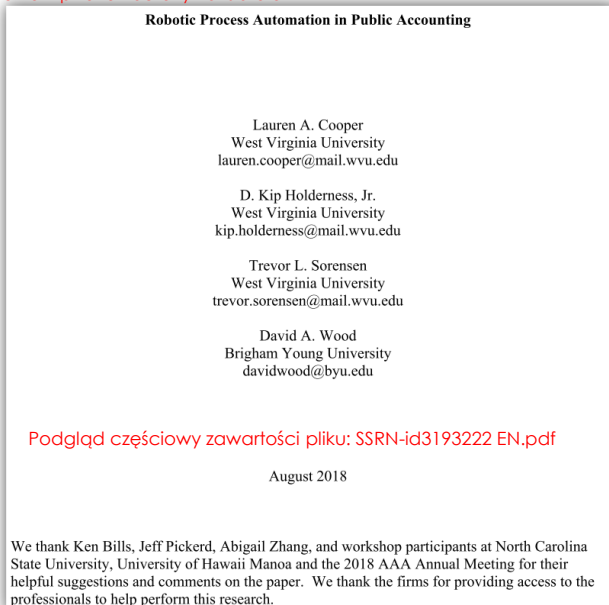


Przykład zasobu zarchiwizowanego w RID



Dokument źródłowy, oryginalny w jęz. ang.
 Metryka cyfrowa dokumentu z sugestią cytowania
 Dokument przełumaczony tłumaczem

Dokument źródłowy,
 oryginalny w jęz. ang.



Download This Paper

Open PDF in Browser

Add Paper to My Library

Share:

Robotic Process Automation in Public Accounting

Accounting Horizons, 33 (4): 15-35

62 Pages • Posted: 11 Sep 2018 • Last revised: 15 Jul 2020

Lauren Cooper
West Virginia University

Kip Holderness
West Virginia University

Trevor Sorensen
West Virginia University - Department of Accounting and Management Information Systems

David A. Wood
Brigham Young University - School of Accountancy

Date Written: 2019

Abstract

This study investigates the adoption and use of Robotic Process Automation (RPA) software—often referred to as bots—in the public accounting industry. Accounting firms use RPA software to automate the input, processing, and output of data across computer applications in order to streamline repetitive and mundane

Paper statistics

DOWNLOADS	ABSTRACT VIEWS	RANK
1,144	4,295	20,074

8 Citations

40 References

Related ejournals

Applied Accounting - Practitioner ejournal

Accounting Technology & Information Systems ejournal

Keywords: Robotic Process Automation (RPA), Automation, Accounting Efficiency, Accounting Effectiveness, Hiring Decisions, Offshoring

JEL Classification: M4, M40, M41, M42, M43

Suggested Citation:

Cooper, Lauren and Holderness, Darin Kip and Sorensen, Trevor and Wood, David A., Robotic Process Automation in Public Accounting (2019). Accounting Horizons, 33 (4): 15-35. Available at SSRN: <https://ssrn.com/abstract/3193222> or <http://dx.doi.org/10.2139/ssrn.3193222>

PlumX Metrics

Metadane dynamiczne

Metadane statyczne

Metryka cyfrowa dokumentu z sugestią cytowania

[illegible]

Metryka cyfrowa dokumentu

The screenshot shows a SpringerLink page for the chapter "Towards Understanding Collaboration Within Circular Business Models". Annotations include:

- Hiperłącze do dokumentu źródłowego**: Points to the "Cite as" link.
- Metadane dynamiczne**: Points to the "Download book" and "Cite chapter" buttons.
- Metadane statyczne**: Points to the "First Online" date (29 June 2018) and the "DOI" (https://doi.org/10.1007/978-3-319-73503-0_9).
- Metryka cyfrowa dokumentu z sugestią cytowania**: Points to the "Cite this chapter as:" section, which provides a citation template: Brown P., Bocken N., Balkenende R. (2018) Towards Understanding Collaboration Within Circular Business Models. In: Moratis L., Melissen F., Idowu S. (eds) Sustainable Business Models. CSR, Sustainability, Ethics & Governance. Springer, Cham. https://doi.org/10.1007/978-3-319-73503-0_9.

Przykład funkcjonalności aplikacji analitycznej – interfejs użytkownika wyszukiwania

The screenshot shows a search application interface with the following annotations:

- Wyszukiwanie proste w indeksie RID**: Points to the "Enter query" field where "Robotic Process Automation" is entered.
- Okienko z wynikami wyszukiwania**: Points to the list of search results on the left, including files like "IEEEAccess2020 EN.pdf" and "RPA-SystematicLiteratureSurvey EN.pdf".
- Okienko podglądu dokumentu zaindeksowanego**: Points to the preview window on the right, showing the content of the selected document, "Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study".

Przykład funkcjonalności aplikacji analitycznej – podgląd dokumentu zaindeksowanego

File Edit Search View Bookmarks Tools Highlight

robotic process automation

Creator: LaTeX with hyperref package
Producer: pdfTeX-1.40.13; modified using iText® 7.1.1 ©2000-2018 iText Group NV (AGPL-version)
Title: **Robotic Process Automation**: A Scientific and Industrial Systematic Mapping Study
Subject: IEEE Access, 2020, 8, ,10.1109/ACCESS.2020.2974934

View publication stats

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339341848>
Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study
Article in **IEEE Access** · February 2020
DOI: 10.1109/ACCESS.2020.2974934

CITATIONS
0
4 authors:
J. G. Enríquez Universidad de Sevilla 33 PUBLICATIONS 65 CITATIONS
[SEE PROFILE](#)
Francisco José Domínguez Mayo Universidad de Sevilla 79 PUBLICATIONS 308 CITATIONS
[SEE PROFILE](#)

READS
157
Andrés Jiménez Ramírez Universidad de Sevilla 27 PUBLICATIONS 118 CITATIONS
[SEE PROFILE](#)
J. A. García-García Universidad de Sevilla 51 PUBLICATIONS 276 CITATIONS
[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects: TemproS- Early Testing and hybrid simulation models in software production (TIN2010-20057-C03-02) View project **NDT 4.0**. Mecanismos para el diseño y gestión de software orientados al usuario View project

All content following this page was uploaded by J. G. Enríquez on 05 March 2020.
The user has requested enhancement of the downloaded file.

Received February 5, 2020, accepted February 14, 2020, date of publication February 18, 2020, date of current version March 4, 2020. Digital Object Identifier 10.1109/ACCESS.2020.2974934
Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study
J. G. ENRÍQUEZ, A. JIMÉNEZ-RAMÍREZ, F. J. DOMÍNGUEZ-MAYO, AND J. A. GARCÍA-GARCÍA
Computer Languages and Systems Department, Escuela Técnica Superior de Ingeniería Informática, 41012 Sevilla, Spain
Corresponding author: J. G. Enríquez (jgenriquez@us.es) This work was supported in part by the Pololas project (TIN2016-76956-C3-2-R) of the Spanish Ministry of Economy and Competitiveness, and in part by the **AIRPA** (P011-19/E09) project of the Centro para el Desarrollo Tecnológico Industrial (CDTI) of Spain.
ABSTRACT The automation of robotic processes has been experiencing an increasing trend of interest in recent times. However, most of literature describes only theoretical foundations on **RPA** or industrial results after implementing **RPA** in specific scenarios, especially in finance and outsourcing. This paper presents a systematic mapping study with the aim of analyzing the current state-of-the-art of **RPA** and identifying existing gaps in both, scientific and industrial literature. Firstly, this study presents an in-depth analysis of the 54 primary studies which formally describe the current state of the art of **RPA**. These primary studies were selected as a result of the conducting phase of the systematic review. Secondly, considering the **RPA** study performed by Forrester, this paper reviews 14 of the main commercial tools of **RPA**, based on a classification framework defined by 48 functionalities and evaluating the coverage of each of them. The result of the study concludes that there are certain phases of the **RPA** lifecycle that are already solved in the market. However, the Analysis phase is not covered in most tools. The lack of automation in such a phase is mainly reflected by the absence of technological solutions to look for the best candidate processes of an organization to be automated. Finally, some future directions and challenges are presented.

Line: 3 Page: 1 File 1 from 50 Entry 1 from 58

Podgląd
pełnotekstowy
zaindeksowanego
w RID dokumentu
źródłowego

Przykład funkcjonalności aplikacji analitycznej – podgląd metadanych

Dokument źródłowy

Metadane zaindeksowanego dokumentu źródłowego

Podgląd pełnotekstowy dokumentu źródłowego zaindeksowanego w RID

Przykład funkcjonalności aplikacji analitycznej – podgląd statystyk

Dokument źródłowy

Metadane zaindeksowanego dokumentu źródłowego

Podgląd pełnotekstowy dokumentu źródłowego zaindeksowanego w RID

Przykład funkcjonalności aplikacji analitycznej – podgląd plików indeksów

Baza indeksów RID

Metadane indeksów

Nazwa indeksu

Metadane wybranego indeksu RID, zakładka Ogólne

Lista indeksów

Baza dokumentów źródłowych PDF i dokumentów stowarzyszonych PDF zaindeksowanych

Documents	Size	File name	Wor
101 813	3 756 383 237	0.a4	
5 575	302 475 580	3.a4	
70	8 929 454	6.a4	
26	3 739 087	1.a4	
67	10 871 030	9.a4	
9 910	490 107 214	5.a4	
6	1 870 867	7.a4	
123 050	255 667 438	8.a4	

Przykład funkcjonalności aplikacji analitycznej – operacje na plikach indeksów

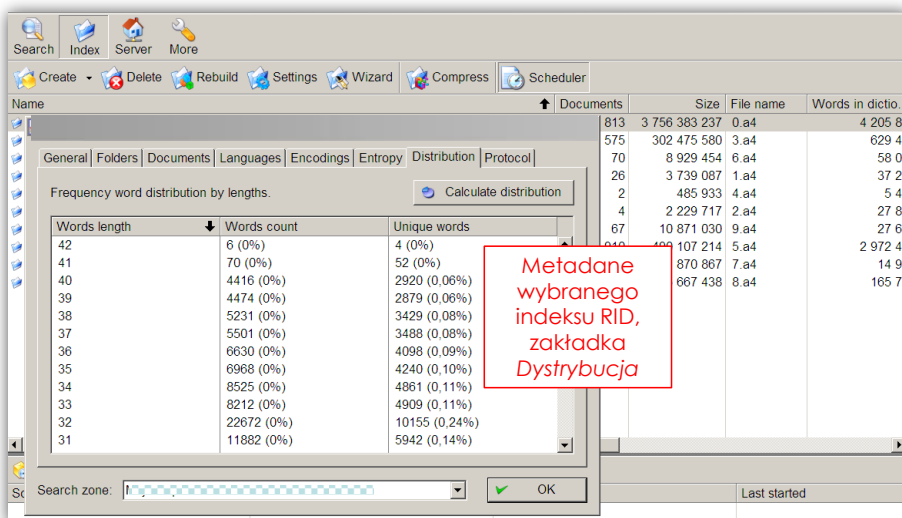
Zadania na indeksach

Metadane wybranego indeksu RID, zakładka Języki

Okno z indeksami danych

Languages	Count	Percents
Unrecognised	97644	95,90%
English	2156	2,11%
Polish	1991	1,95%
Dutch	10	0%
German	2	0%
Spanish	2	0%
Czech	2	0%
Norwegian	2	0%
Korean	2	0%
French	1	0%
Chinese	1	0%
Total	101 813	100%

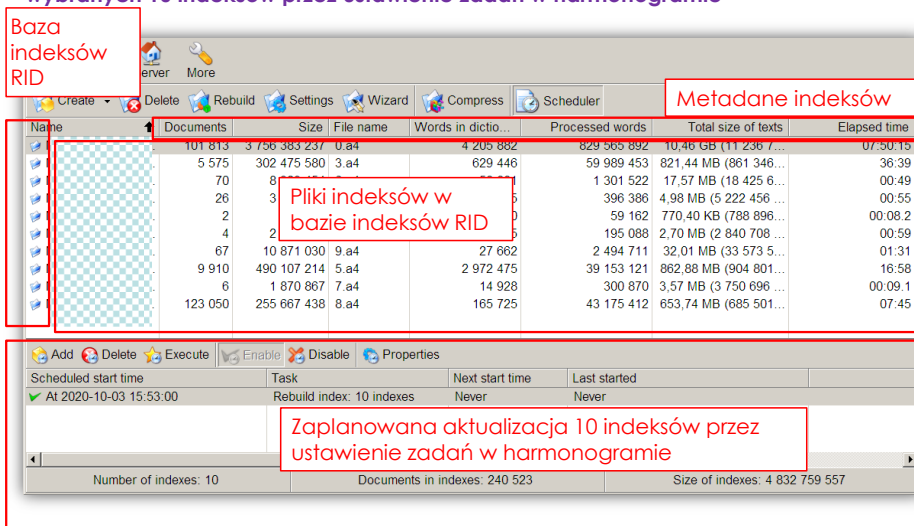
Przykład funkcjonalności aplikacji analitycznej – podgląd metadanych indeksu



Metadane wybranego indeksu RID, zakładka Dystrybucja

Words length	Words count	Unique words
42	6 (0%)	4 (0%)
41	70 (0%)	52 (0%)
40	4416 (0%)	2920 (0,06%)
39	4474 (0%)	2879 (0,06%)
38	5231 (0%)	3429 (0,08%)
37	5501 (0%)	3488 (0,08%)
36	6630 (0%)	4098 (0,09%)
35	6968 (0%)	4240 (0,10%)
34	8525 (0%)	4861 (0,11%)
33	8212 (0%)	4909 (0,11%)
32	22672 (0%)	10155 (0,24%)
31	11882 (0%)	5942 (0,14%)

Przykład funkcjonalności aplikacji analitycznej - automatyczna aktualizacji wybranych 10 indeksów przez ustawienie zadań w harmonogramie



Baza indeksów RID

Metadane indeksów

Pliki indeksów w bazie indeksów RID

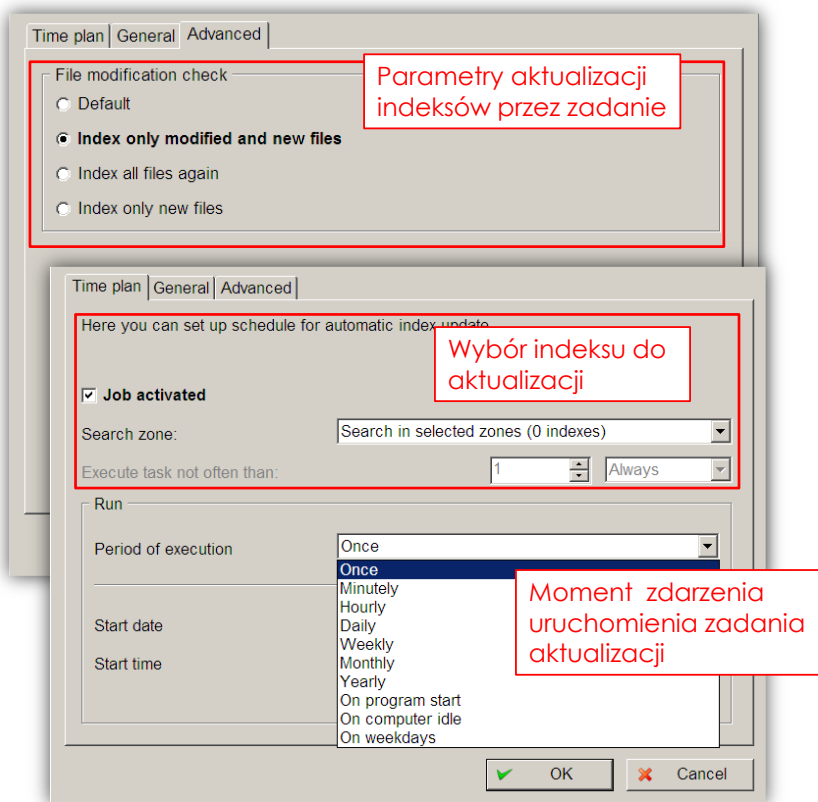
Zaplanowana aktualizacja 10 indeksów przez ustawienie zadań w harmonogramie

Name	Documents	Size	File name	Words in dictio...	Processed words	Total size of texts	Elapsed time
101 813	3 756 383 237	0.a4		4 205 862	829 565 892	10,46 GB (11 236 7...	07:30:15
5 575	302 475 580	3.a4		629 446	59 989 453	821,44 MB (861 346...	36:39
70					1 301 522	17,57 MB (18 425 6...	00:49
26					396 386	4,98 MB (5 222 456...	00:55
2					59 162	770,40 KB (788 896...	00:08.2
4					195 088	2,70 MB (2 840 708...	00:59
67	10 871 030	9.a4		27 662	2 494 711	32,01 MB (33 573 5...	01:31
9 910	490 107 214	5.a4		2 972 475	39 153 121	862,88 MB (904 801...	16:58
6	1 870 867	7.a4		14 928	300 870	3,57 MB (3 750 696...	00:09.1
123 050	255 667 438	8.a4		165 725	43 175 412	653,74 MB (685 501...	07:45

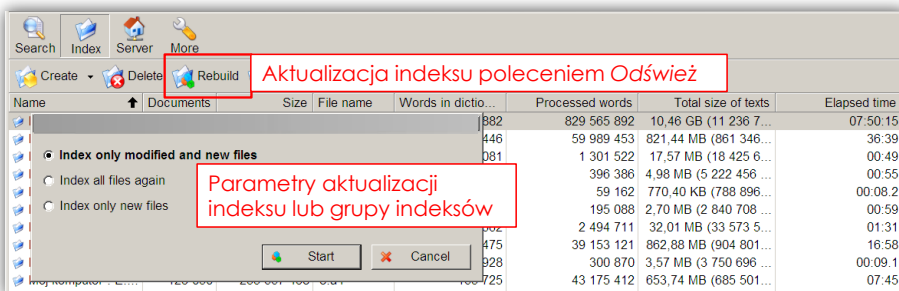
Scheduled start time	Task	Next start time	Last started
At 2020-10-03 15:53:00	Rebuild index: 10 indexes	Never	Never

Number of indexes: 10 Documents in indexes: 240 523 Size of indexes: 4 832 759 557

Przykład funkcjonalności aplikacji analitycznej – parametry indeksowania



Przykład funkcjonalności aplikacji analitycznej – parametry aktualizowania indeksu



Bibliografia

1. Adam, U. A., & K., K. (2021). *Driving forces behind the management of Institutional Repositories: Qualitative evidences*. Malaysian Journal of Library & Information Science, 26(3), 33–56. Retrieved from <https://sare.um.edu.my/index.php/MJLIS/article/view/24444>
2. Aguirre, S., & Rodríguez, A. (2017). *Automation of a Business Process Using Robotic Process Automation (RPA): A Case Study*. WEA.
3. Antasari, I. W. (2022). *Influence of Offline and Online Access Skills on Motivation to Access Institutional Repositories*. Library Philosophy and Practice, 2022(Feb).
4. CoreTrustSeal Trustworthy Data Repositories Requirements, <https://www.coretrustseal.org/why-certification/requirements/>, data dostępu: 12.01.2022 r.
5. Crow, R. (2002). *The case for institutional repositories: a SPARC position paper*. ARL Bimonthly Report 223.
6. Davidson, E., Edwards, R., Jamieson, L., & Weller, S. (2019). *Big data, qualitative style: a breadth-and-depth method for working with large amounts of secondary qualitative data*. Quality & Quantity, 53(1), 363–376.
7. Del Valle, C. (2019). *A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle*. In *Advanced Information Systems Engineering: 31st International Conference, CAiSE 2019, Rome, Italy, June 3–7, 2019, Proceedings* (Vol. 11483, p. 446). Springer.
8. Desai, P. (2020). *Robotic process automation: RPA Pre-requisite and pivotal points*. Special issue: IAISCT(SS4). 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 446–451.
9. Enríquez, J.G., Jiménez-Ramírez, A., Dominguez-Mayo, F.J., & García-García, J.A. (2020). *Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study*. IEEE Access, 8, 39113–39129.
10. Gałuszka, E. (2017). *Repozytorium instytucjonalne jako narzędzie komunikacji naukowej i promocji nauki*. Nowa Biblioteka. Usługi, Technologie Informacyjne i Media, 27(4), 7–17.
11. Haddow, G., & White, H. (Eds.). (2021). *Assessment as Information Practice: Evaluating Collections and Services*. Routledge.
12. INCOSE 2010 *Systems Engineering Handbook (SEBoK)* v.3. s. 3.4/20
13. ISO 10006:2017 *Quality management systems – Guidelines for quality management in Project*.
14. ISO 10014:2006 *Quality management – Guidelines for realizing financial and economic benefits*.
15. ISO 10015:2019 *Guidelines for competence management and people development*.
16. ISO 56000:2020 *Innovation management – Fundamentals and vocabulary*.

17. ISO 9000:2015 *Quality management systems – Fundamentals and vocabulary*.
18. ISO/IEC 15288:2008 *Systems and software engineering – System life cycle processes*.
19. Janiak, M., & Próchnicka, M. (2017). *Otwarte repozytorium Uniwersytetu Jagiellońskiego jako element systemu zarządzania informacją instytucjonalną: wyniki badań przeprowadzonych wśród przedstawicieli środowiska akademickiego uczelni*. Kraków: Biblioteka Jagiellońska.
20. Jimenez-Ramirez A., Reijers H.A., Barba I., Del Valle C. (2019) *A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle*. In: Giorgini P., Weber B. (eds) *Advanced Information Systems Engineering*. CAiSE 2019. Lecture Notes in Computer Science, vol 11483. Springer, Cham.
21. Judy Chuan-Chuan Lin (2007) *Online stickiness: its antecedents and effect on purchasing intention*, Behaviour & Information Technology, 26:6, 507–516, DOI: 10.1080/01449290600740843.
22. Knoth, P., Anastasiou, L., & Pearce, S. (2014). *My repository is being aggregated: a blessing or a curse?* In: Open Repositories 2014 (OR2014), 9–13 Jun 2014, Helsinki, Finland.
23. Kotnis M., *Modele zarządzania ryzykiem w warunkach niepewności*. Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 802 Finanse, Rynki Finansowe, Ubezpieczenia nr 65 (2014), s. 669–678.
24. Lewandowski, T., Starczewski, M. (2014). *Po czym rozpoznać dobre repozytorium?* Biblioteka i Edukacja, 6, 1–16.
25. *Międzynarodowe Ramowe Praktyki Zawodowe (IPPF) – Praktyczny przewodnik Zarządzanie ciągłością działania*. Instytut Audytorów Wewnętrznych, 2014.
26. Mikołajuk L., *Nowoczesny system zarządzania wiedzą na przykładzie funkcjonowania repozytorium instytucjonalnego*, [w:] Piestrzyński T., Jerzyk-Wojtecka J. (red.), *Biblioteka w społeczeństwie wiedzy. Informacja, edukacja, profesjonalizm*. Konferencja naukowa Biblioteki Uniwersytetu Łódzkiego. Materiały konferencyjne, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015, s. 262–269, doi: 10.18778/8088-191-4.20.
27. Mikołajuk, L. (2015). *Repozytorium instytucjonalne jako istotny element promocji nauki*. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
28. Nahotko, M. (2021). *Zastosowanie metadanych w cyklu życia danych badawczych*. In P. Korycińska (Ed.), *Horyzonty informacji* (Vol. 2, pp. 8–20). Uniwersytet Jagielloński, Biblioteka Jagiellońska. <https://ruj.uj.edu.pl/xmlui/handle/item/287474>
29. Nascimento, E.D., Nguyen-Duc, A., Sundbø, I., & Conte, T.U. (2020). *Software engineering for artificial intelligence and machine learning software: A systematic literature review*. ArXiv, abs/2011.03751.

30. Neale, B., Bishop, L. (2012). *The Timescapes Archive: a stakeholder approach to archiving qualitative longitudinal data*. *Qualitative Research*, 12, 53–65.
31. NICE RPA team with Steve Kaelble: *Robotic Process Automation for Dummies®*, NICE Special Edition, John Wiley & Sons, 2018.
32. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), <http://www.prisma-statement.org>, data dostępu 15.01.2022 r.
33. Poskrobko, B. (2009). *Nauka i naukowość w ekonomii i zarządzaniu*. Optimum, *Studia Ekonomiczne* Nr 4 (44).
34. *Research Intelligence. Research Metrics Guidebook*. Elsevier, March 2019.
35. *Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study*, 10.1109/ACCESS.2020.2974934.
36. Sobczak A. (2022). *Robotic Process Automation as a Digital Transformation Tool for Increasing Organizational Resilience in Polish Enterprises*. *Sustainability*. 14(3):1333. <https://doi.org/10.3390/su14031333>.
37. Szafrąński, L. (2020). *Repozytoria instytucjonalne na świecie: próba charakterystyki*. In A. Podrazik, J. Stanek-Kapcia, M. Urbaniec (Eds.), *Biblioteki uczelniane wobec środowiska akademickiego: nowe obszary działania: materiały Jubileuszowej Konferencji Biblioteki Głównej AGH z okazji obchodów 100-lecia Akademii Górniczo-Hutniczej im. Stanisława Staszica*, Kraków 12–13 września 2019 r. (s. 35–46). Wydawnictwo AGH.
38. Tarrant, D., O'Steen, B., Brody, T., Hitchcock, S., Jefferies, N., & Carr, L. (2009). *Using OAI-ORE to Transform Digital Repositories into Interoperable Storage and Services Applications*. *Code4Lib Journal*.
39. *The Forrester Wave™: Robotic Process Automation*, Q4 2019, October 23, 2019.
40. Ujournunna, J. C., & Agbawudikeizu, E. C. (2021). *Management of digital records for knowledge preservation in academic institutions for a sustainable development in education sector*. *Law, Business and Sustainability Herald*, 1(3), 5–9.
41. *WEB of Knowledge XML User Guide*. Thomson Reuters, May 9, 2013.
42. Wewerka, J., & Reichert, M. (2021). *Robotic process automation-a systematic mapping study and classification framework*. *Enterprise Information Systems*, 1–38.
43. *Wspólna Metoda Oceny. Doskonalenie organizacji poprzez samoocenę*. Kancelaria Prezesa Rady Ministrów Warszawa, 2008.
44. Xiao, Y., Watson, M. (2019). *Guidance on Conducting a Systematic Literature Review*. *Journal of Planning Education and Research*, 39(1), 93–112.
45. Xie, Y., Cruz, L., Heck, P., Rellermeyer, J. S. (2021). *Systematic Mapping Study on the Machine Learning Lifecycle*. In L. O'Conner (Ed.), *Proceedings – 2021 IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI*, WAIN 2021 (pp. 70–73). [9474380] (Proceedings–2021

- IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI, WAIN 2021).
46. Yatskiv, S., Voytyuk, I., Yatskiv, N., Kushnir, O., Trufanova, Y., & Panasyuk, V. (2019). *Improved Method of Software Automation Testing Based on the Robotic Process Automation Technology*. 2019 9th International Conference on Advanced Computer Information Technologies (ACIT), 293–296. DOI: 10.1109/ACITT.2019.8780038.
 47. Zdonek, I., Hysa, B. (2017). *Analiza publikacji z obszaru nauk o zarządzaniu pod względem stosowanych metod badawczych*. Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska.
 48. Zdonek, I., Hysa, B., & Zdonek, D. (2016). *Publikacje przeglądowe w naukach o zarządzaniu – istota i tendencje*. Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska.

Spis rysunków

Rys. 1. Nośniki dokumentów elektronicznych z zawartością danych.....	14
Rys. 2. Bazy danych w badaniach bibliograficznych.....	15
Rys. 3. Systematyczny przegląd literatury	16
Rys. 4. Przykład zastosowania metodyki doboru publikacji w artykule naukowym	17
Rys. 5. Indywidualne i wielokrotne badania bibliograficzne na dostawcach danych – schemat poglądowy badań bibliograficznych	18
Rys. 6. Badania bibliograficzne prowadzone indywidualnie w grupach lub zespołach specjalistów	19
na dostawcach danych – schemat poglądowy badań bibliograficznych.....	19
Rys. 7. Przykład publikacyjnego procesu selekcji w badaniach bibliograficznych.....	19
Rys. 8. Zautomatyzowane i manualne pozyskanie, aktualizacja zasobów oraz analiza danych wtórnych przez użytkowników z wykorzystaniem RID na etapie badań bibliograficznych	31
Rys. 9. Cele pracy w kontekście prowadzenia badań bibliograficznych w ocenie stanu wiedzy naukowej, badawczej, praktycznej.....	33
Rys. 10. Uszczegółowiony kontekst wewnętrzny i zewnętrzny prowadzenia badań bibliograficznych w skali makro.....	34
Rys. 11. Relacje asocjacyjne pomiędzy terminami związanymi z cyfrowymi aktywami intelektualnymi w organizacji.....	35
Rys. 12. Grupy interesariuszy z różnych sektorów gospodarki zainteresowane cyfrowymi aktywami intelektualnymi	36
Rys. 13. Przedmiot badań w ujęciu hierarchicznym w podziale cyklu życia na procesy, działania i zadania	37
Rys. 14. Algorytmiczne przedstawienie modelu badawczego w monografii.....	40

Rys. 15. Warstwa opisowa modelu badawczego w monografii	41
Rys. 16. Warstwa aplikacyjna modelu badawczego w monografii	41
Rys. 1.1. Kluczowe elementy składowe procesu w metodyce zautomatyzowanego prowadzenia badań bibliograficznych	42
Rys. 1.2. Kluczowe elementy składowe procesu w metodyce zautomatyzowanego prowadzenia badań bibliograficznych	43
Rys. 1.3. Uogólniona metodyka badania bibliograficznego: a) procesy w metodyce badania bibliograficznego w cyklu życia, b) schemat poglądowy badań bibliograficznych.....	44
Rys. 1.4. Uogólniona metodyka badań bibliograficznych specjalisty dla wielu osiągnięć twórczych.....	45
Rys. 1.5. Metodyka zautomatyzowanego prowadzenia badań bibliograficznych na zasobach Repozytorium Instytucjonalnego Danych.....	46
Rys. 1.6. Proces pozyskania, aktualizacji zasobów RID w chmurze cyfrowej ...	48
Rys. 1.7. Obiekty i elementy graficzne wykorzystane w schematach algorytmów procesów	48
Rys. 1.8. Schemat pierwszego procesu dla opracowania zbioru założeń badań bibliograficznych	49
Rys. 1.9. Specyfikacja grup dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych	50
Rys. 1.10. Specyfikacja grup zbiorów danych, dokumentów źródłowych i stowarzyszonych zawierających aktywa intelektualne	51
Rys. 1.11. Specyfikacja cyfrowych aktywów osobowych	52
Rys. 1.12. Schemat pozyskiwania i opracowania słów kluczowych lub wyrażen w postaci kwerend	53
Rys. 1.13. Specyfikacja grup źródeł danych i zbiorów danych w eksploracji i ekstrakcji dokumentów źródłowych.....	54
Rys. 1.14. Typowa architektura klient–serwer w dostępie klienta do danych na serwerze.....	55
Rys. 1.15. Algorytm procesu opracowania zbioru założeń badań bibliograficznych dla założeń wstępnych.....	57
Rys. 1.16. Ogólny schemat procesu pozyskania zasobów cyfrowych	58
Rys. 1.17. Szczegółowy schemat procesu pozyskania zasobów cyfrowych	60
Rys. 1.18. Proces pozyskania zasobów cyfrowych w metodyce badań bibliograficznych na zasobach RID.....	61
Rys. 1.19. Algorytm procesu aktualizacji w procesie pozyskania zasobów cyfrowych.....	62
Rys. 1.20. Algorytm procesu pozyskania zasobów cyfrowych.....	62
Rys. 1.21. Algorytm procedury P1. <i>Utworzenie bazy dokumentów PDF w chmurze obliczeniowej</i>	63
Rys. 1.22. Algorytm procedury P2. <i>Translacja automatyczna i manualna dokumentów</i>	64
Rys. 1.23. Algorytm procedury P3. <i>Tworzenie dokumentu PDF z plików</i>	65

Rys. 1.24. Algorytm procedury P4. <i>Tworzenie indeksów</i>	66
Rys. 1.25. Algorytm procedury P5. <i>Sprawdzenie zabezpieczenia dokumentu PDF przed edycją lub zaznaczaniem treści</i>	67
Rys. 1.26. Algorytm procedury P6. <i>Indeksowanie</i>	68
Rys. 1.27. Algorytm procedury P7. <i>Archiwizacja raportu</i>	69
Rys. 1.28. Algorytm podprocesu 2f. <i>Tworzenie indeksów danych w chmurze obliczeniowej</i>	70
Rys. 1.29. Szczegółowy schemat procesu wykonania zapytania bibliograficznego i wyszukiwanie	72
Rys. 1.30. Algorytm podprocesu 3a. <i>Wyszukiwanie, ekstrakcja informacji w indeksach danych</i>	73
Rys. 1.31. Algorytm podprocesu 3a. <i>Wyszukiwanie wyrazów lub wyrażeń z zaawansowanymi filtrami</i>	74
Rys. 1.32. Algorytm podprocesu 3a. <i>Ekstrakcja informacji z bazy danych</i>	75
Rys. 1.33. Algorytm procesu analizy ilościowej i jakościowej danych	76
Rys. 1.34. Algorytm procesu poddania pełnotekstowej analizie treści dokumentów	77
Rys. 1.35. Algorytm procesu oceny wyników z konsumpcją zebranych informacji	78
Rys. 1.36. Proces weryfikacji i walidacji w metodyce zautomatyzowanego prowadzenia badań bibliograficznych	80
Rys. 1.37. Proces weryfikacji w metodyce zautomatyzowanego prowadzenia badań bibliograficznych	81
Rys. 1.38. Proces weryfikacji i walidacji badań bibliograficznych	82
Rys. 1.39. Algorytm procedury P8. <i>Walidacja poszerzona RID</i>	83
Rys. 1.40. Algorytm procedury P9. <i>Walidacja satysfakcji użytkowników w RID</i>	84
Rys. 1.41. Algorytm procedury P10. <i>Walidacja produktu cyfrowego RID</i>	85
Rys. 1.42. Algorytm procedury P11. <i>Walidacja badań bibliograficznych na zasobach RID</i>	85
Rys. 1.43. Algorytm procedury P12. <i>Walidacja zdefiniowanych ról w RID</i>	86
Rys. 2.1. Stany i procesy w cyklu życia produktu cyfrowego–RID	88
Rys. 2.2. Cykl życia produktu cyfrowego–programowalnego robota cyfrowego RID w cyklu życia Repozytorium Instytucjonalnego Danych	91
Rys. 2.3. Cykl życia danych RID w cyklu życia Repozytorium Instytucjonalnego Danych	92
Rys. 2.4. Cykl życia badań bibliograficznych na zasobach RID w cyklu życia Repozytorium Instytucjonalnego Danych	93
Rys. 2.5. Proces wsparcia i pielęgnacji w cyklu życia RID	94
Rys. 2.6. Cztery kategorie doskonalenia produktów cyfrowych w procesie pielęgnacji w cyklu życia RID	96
Rys. 2.7. Syntetyczne ujęcie cykli życia w cyklu życia produktu cyfrowego – Repozytorium Instytucjonalnego Danych	97

Rys. 2.8. Cykl życia badań bibliograficznych w podejściu tradycyjnym oraz podejściu zrobotyzowanym	100
Rys. 2.9. Podejście tradycyjne w badaniach bibliograficznych	102
Rys. 2.10. Podejście zautomatyzowane w badaniach bibliograficznych.....	104
Rys. 2.11. Zarządzanie ciągłością działania w procesie użytkowania Repozytorium Instytucjonalnego Danych	107
Rys. 2.12. Modelowe zapewnianie ciągłości działania RID w prowadzeniu badań bibliograficznych	108
Rys. 2.13. Możliwości aktualizacji zasobów RID z wykorzystaniem robotyzacji procesów: a) za pomocą interfejsu użytkownika, b) komunikacji bezpośredniej przez API.....	109
Rys. 2.14. Procesy zarządzania, podstawowe i pomocnicze w systemie zarządzania jakością	120
Rys. 2.15. Elementy składowe w ciągłym doskonaleniu jakości produktów cyfrowych, zasobów i dostępu do RID w chmurze cyfrowej.....	122
Rys. 2.16. Ciągłe doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez użytkownika	123
Rys. 2.17. Rozwój, doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez użytkownika	123
Rys. 2.18. Ciągłe doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez koordynatora i operatora programowalnego robota cyfrowego.....	124
Rys. 2.19. Rozwój, doskonalenie jakości zasobów i dostępu do RID w chmurze cyfrowej przez koordynatora i operatora programowalnego robota cyfrowego.....	124