



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

IBE



entuzjaści
edukacji

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Monitorowanie losów absolwentów uczelni
z wykorzystaniem danych administracyjnych Zakładu
Ubezpieczeń Społecznych

Narzędzie przygotowujące eksporty danych z rejestrów USOS na użytek badań losów absolwentów z wykorzystaniem rejestrów uczelni i ZUS – dokumentacja techniczna i raport metodologiczny

Raport przygotowany przez Pracownię Ewaluacji Jakości Kształcenia na Uniwersytecie Warszawskim w ramach projektu systemowego Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego, współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego, realizowanego przez Instytut Badań Edukacyjnych.

Warszawa, 21 Kwietnia 2015

Autorzy:
dr Mikołaj Jasiński
dr Tomasz Zając

Wydawca:
Instytut Badań Edukacyjnych
ul. Górczewska 8
01-180 Warszawa
tel. (22) 241 71 00; www.ibe.edu.pl

© Copyright by: Instytut Badań Edukacyjnych, Warszawa, kwiecień 2015

Publikacja opracowana w ramach projektu systemowego „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego”, współfinansowanego przez Unię Europejską ze środków Europejskiego Funduszu Społecznego, realizowanego przez Instytut Badań Edukacyjnych.

Spis Treści

1. Wprowadzenie	4
2. Eksportowane zmienne oraz opis algorytmów ich obliczania	5
3. Bezpieczeństwo i poufność danych	11
4. Wnioski metodologiczne i rekomendacje	11
4.1. Jakość danych w USOS	11
4.2. Rozszerzenie zawartości zbiorów	12
4.3. Prowadzenie badań reprezentacyjnych	12

1. Wprowadzenie

Stosowane na Uniwersytecie Warszawskim i na większości największych polskich uczelni systemy informatyczne (USOS oraz IRK) zostały zaprojektowane z myślą o bieżącej obsłudze procesów dydaktycznych. Miały służyć sprawnemu wprowadzaniu i przeglądaniu różnorodnych typów informacji dotyczących pojedynczych studentów. Systemy te nie były jednak projektowane z myślą o konstruowaniu zestawień i raportów. Wszelkie analizy wymagają eksportu danych i ich niekiedy długotrwałej obróbki (zob. „Konceptualizacja problemów badawczych”).

W ramach projektu *Monitorowanie losów absolwentów uczelni wyższych z wykorzystaniem danych administracyjnych Zakładu Ubezpieczeń Społecznych* informacje z systemów informatycznych pozyskiwane były w postaci niemal „surowej” tj. zbior danych zorganizowany był w sposób dość bliski strukturze danych w rejestrach uczelni. Rozwiązanie takie miało jedną podstawową zaletę - pozwalało na pełniejszą kontrolę jakości danych (zidentyfikowanie występujących błędów, ocenę kompletności zbieranych informacji itd.) i w efekcie wyselekcjonowanie informacji, które mogą zostać wykorzystane w badaniach. Istotną wadą takiego podejścia do eksportu informacji z rejestrów uczelni jest konieczność długotrwałej i pracochłonnej obróbki danych.

Do sprawnego prowadzenia regularnych badań dotyczących ścieżek edukacyjnych konieczne jest zautomatyzowanie procesu przygotowania do analiz danych pochodzących z rejestrów informatycznych uczelni. W ramach przedstawionej przez Uniwersytet Warszawski oferty opisaliśmy dwa możliwe rozwiązania. Pierwszy wariant zakładał wbudowanie narzędzia służącemu eksportom w Uniwersytecki System Obsługi Studiów. Cały proces przygotowania danych odbywałby się „wewnątrz” systemów informatycznych uczelni. Procedura eksportu danych byłaby jedną z opcji dostępnych dla uprawnionych użytkowników. Użytkownik otrzymywałby przetworzone dane, zawierające już obliczone wartości indeksów itd. Drugi wariant przewidywał, że tylko część procesów związanych z przygotowaniem baz danych przebiegałaby w obrębie USOS. Z USOS eksportowane byłyby dane przetworzone w niewielkim stopniu, ale zorganizowane wedle ściśle określonego schematu. Tego typu rozwiązania istnieją już w ramach USOS. Dalsza obróbka przebiegałaby z wykorzystaniem dodatkowego oprogramowania, które przetwarzałoby dane pochodzące z systemów informatycznych uczelni do postaci, w jakiej wykorzystywane są w ramach badań nad ścieżkami edukacyjnymi i karierami zawodowymi.

W ramach realizacji projektu badawczego *Monitorowanie losów absolwentów uczelni z wykorzystaniem danych administracyjnych Zakładu Ubezpieczeń Społecznych. Część 3: Stworzenie narzędzi służących do przetwarzania i analizy informacji pochodzących z rejestrów administracyjnych wielu uczelni i ZUS na potrzeby badań losów edukacyjno-zawodowych absolwentów szkół wyższych* udało się zrealizować wariant pierwszy, który już na etapie składania oferty wydawał się bardziej atrakcyjny. Narzędzie zostało wbudowane w USOS i jest dystrybuowane wraz z systemem przez MUCI, które jest odpowiedzialne za rozwój systemu. Oznacza to, że każda uczelnia, która zainstaluje USOS będzie mogła skorzystać z formularza *Dane do badań losów absolwentów* w ramach modułu „sprawozdawczość”. Znaczna część uczelni Polskich, w tym duża część największych z nich, będzie mogła przygotować dane według tego samego schematu.

Przedstawione rozwiązanie informatyczne zostało przygotowane przede wszystkim z myślą o uczelniach korzystających z USOS. Zamieszczony poniżej opis algorytmów przygotowania zmiennych pozwala jednak na przygotowanie zbiorów danych o identycznej strukturze przez uczelnie wykorzystujące inne rozwiązania informatyczne do obsługi dydaktyki.

W poniższym opracowaniu opisujemy przede wszystkim zawartości eksportowanych danych wraz z charakterystyką zmiennych i książką kodową oraz algorytmy wykorzystywane do ich wyznaczenia. Zestaw eksportowanych zmiennych został zmodyfikowany w stosunku do zbioru wykorzystywanego w badaniu *Monitorowanie losów absolwentów uczelni wyższych z wykorzystaniem danych administracyjnych Zakładu Ubezpieczeń Społecznych*. Jest to wynikiem badań prowadzonych na Uniwersytecie Warszawskim i Uniwersytecie Śląskim w ramach *Studium współzależności procesów*

dydaktycznych oraz funkcjonowania elektronicznego systemu obsługi studiów. Jedynym ograniczeniem przy wykorzystaniu wyeksportowanych danych jest jakość informacji gromadzonych w rejestrach administracyjnych uczelni, o czym piszemy w części poświęconej rekomendacjom i wnioskom metodologicznym. Działanie narzędzia została przetestowane na przykładzie danych z Uniwersytetu Warszawskiego oraz Uniwersytetu Śląskiego. Sprawdzana była zgodność wyeksportowanych zmiennych z zapisami w rejestrach uczelni. W trakcie testów zostały wygenerowane pliki, które w postaci zanonimizowanej zostały załączone do raportu. Posłużyły one także do przygotowania przykładowych raportów automatycznych.

2. Eksportowane zmienne oraz opis algorytmów ich obliczania

W USOS dane dotyczące studentów oraz zdarzeń zachodzących w trakcie studiów są gromadzone w szeregu powiązanych ze sobą tabel. Osobno zbierane są informacje na temat programów studiów odbywanych przez studentów, osobno na temat kolejnych etapów studiów i osobno na temat ocen uzyskiwanych z egzaminów. Zadaniem narzędzia eksportującego dane jest przetworzenie informacji zawartych w rejestrach uczelni tak, aby powstała prostokątna tablica opisująca przebieg studiów, co pozwoli znacznie usprawnić zarówno proces analizy danych jak i łączenia ich z informacjami pochodzącymi z innych źródeł.

Jednostką obserwacji w wyeksportowanym zbiorze danych jest osoba wpisana na program studiów, przy czym eksportowane są wyłącznie dane osób posiadających numer PESEL. Narzędzie daje możliwość wyeksportowania informacji na temat kierunków studiów pierwszego stopnia, drugiego stopnia oraz jednolitych magisterskich. Oznacza to, że osoba studiująca na więcej niż jednym programie studiów będzie występowała w więcej niż jednym wierszu. Odrębne kody programów nadawane są studiom na tym samym kierunku, ale w różnych formach (trybach) oraz studiom na różnych stopniach. Osoba rozpoczynająca studia równoległe na kolejnym kierunku będzie występować w wyeksportowanym zbiorze dwukrotnie. Podobnie będzie w przypadku studentów, którzy zmieniają tryb studiów np. z niestacjonarnych wieczorowych przeniosą się na stacjonarne oraz osób, które po ukończeniu studiów pierwszego stopnia kontynuowały naukę na studiach drugiego stopnia. Jako dwie obserwacje wystąpią także osoby, które przerwały studia na danym programie i wznowiły studia na programie o zmienionym kodzie (np. kod studiów stacjonarnych pierwszego stopnia na socjologii na UW zmienił się z DZ-SC na S1-SC). Przyjęte rozwiązanie pozwala na prowadzenie analiz osobno dla każdego kierunku studiów przyjmując jako jednostkę obserwacji „osobo-kierunek”, jednocześnie nie wyklucza to przetworzenia danych w taki sposób, aby możliwe było uwzględnienie w analizie jednoczesnych studiów na więcej niż jednym kierunku.

Dla każdej z obserwacji zostanie przygotowany zestaw zmiennych. Poniżej znajduje się wykaz zmiennych wraz z opisem algorytmów ich wyznaczenia oraz klucze kodowe stanowiące książkę kodową (słowniki dla niektórych zmiennych np. kierunków studiów są definiowane przez uczelnie; poniższy dokument zawiera te klucze kodowe, które są wspólne dla wszystkich uczelni mogących korzystać z narzędzia):

- PESEL – numer PESEL osoby, pozwalający na identyfikację badanych w przypadku łączenia informacji pochodzących z różnych rejestrów.
- PLEC - płeć osoby
 - K – kobieta
 - M – mężczyzna
- ROKUR - rok urodzenia osoby w postaci RRRR

- NPELUSOS - trwała niepełnosprawność nabyta przed lub w trakcie studiów (typ_waznosc = T i od_kiedy <= planowana data ukończenia studiów)
 - 0 – nie
 - 1 – tak
- SZKOLA - rodzaj szkoły średniej (szkoła średnia podpięta do osoby, jedna z LO, LP, TE)
 - LO – Liceum Ogólnokształcące
 - LP – Liceum Profilowane
 - TE - Technikum
- KODSZK - kod pocztowy szkoły średniej
- KIERUNEK - kod programu studiów stosowany przez uczelnię
- POLON - kod POL-on programu studiów
- TRYB - tryb studiów:
 - 1 – stacjonarne;
 - 2 – wieczorowe;
 - 3 – zaoczne
- STOPIEN - stopień studiów:
 - 1 - I stopień;
 - 2 - II stopień;
 - 3 - jednolite mgr
- CZASSTUD - czas trwania studiów wyrażony w semestrach
- JEDN - kod jednostki administracyjna programu studiów nadany przez uczelnię
- JEDN2 – nadany przez uczelnię kod jednostki, na której osoba uzyskała pierwszy dyplom. Zazwyczaj będzie się pokrywała z jednostką administracyjną.
- ROKSTART - rok rozpoczęcia studiów na programie (w ogóle - nie instancji). Student może być wpisywany na program wielokrotnie, np. jeśli został skreślony i ponownie rozpoczął studia na danym programie studiów. Na potrzeby badania brana jest pod uwagę data pierwszego wpisania na program studiów, co pozwoli m.in. uwzględnić przerwy w studiowaniu przy obliczaniu opóźnienia w ukończeniu studiów.
- DATSTART - miesiąc rozpoczęcia studiów na programie (w ogóle - nie instancji).
- ROKABS - rok zakończenia ostatniego zaliczonego etapu na programie (rok z ostatniej daty zakończenia etapu o statusie A – zaliczony automatycznie, Z – zaliczony ręcznie lub S –

zaliczony warunek). Brak danych oznacza, że student nie zaliczył żadnego etapu w ramach programu.

- DATABS - miesiąc zakończenia ostatniego zaliczonego etapu na programie (miesiąc z ostatniej daty zakończenia etapu o statusie A – zaliczony automatycznie, Z – zaliczony ręcznie lub S – zaliczony warunek). Brak danych oznacza, że student nie zaliczył żadnego etapu w ramach programu.
- ROKDYP - rok uzyskania pierwszego dyplomu na programie (dla programów o statusie DYP rok uzyskania pierwszego dyplomu na danym programie)
- DATDYP - miesiąc uzyskania pierwszego dyplomu na programie (dla programów o statusie DYP miesiąc uzyskania pierwszego dyplomu na danym programie)
- SKR - czy skreślony (czy program ostatni - maksymalne ID - jest skreślony). Tylko dla osób, które aktualnie są skreślone. Jeśli student był skreślony, a następnie wznowił studia i uzyskał dyplom, to zmienna przyjmie wartość 0.
- ROKSKR - rok ostatniego skreślenia (jeśli SKR=1)
- DATSKR - miesiąc ostatniego skreślenia (jeśli SKR=1)
- SEMSTART - semestr od którego student rozpoczął studia (najmniejszy numer semestru z historii danej osoby na programie). Niektórzy studenci przenoszą się z innych programów studiów lub nawet z innych uczelni i nie muszą rozpoczynać studiów od samego początku. Dla osób takich w systemie brakuje zapisów dotyczących początkowych semestrów lub lat studiów. Rozpoczęcie studiów nie od pierwszego etapu musi zostać uwzględnione przy ustalaniu terminowości ukończenia studiów.
- DTERM1 - czy dyplom w terminie z uwzględnieniem tego, że ktoś miał urlop lub, że nie zaczął studiów od pierwszego etapu

(dla programów o statusie DYP

data rozpoczęcia := minimum z "data od" cykli etapów osoby

przedłużenie := różnica pomiędzy datą zakończenia ostatniego etapu i maksimum z data zakon dla cykli przy etapach osoby o ile ta różnica dodatnia

długość studiów (w semestrach) := (długość programu w semestrach - pierwszy semestr osoby + 1) + ile semestrów urlopu

wyliczone ukończenie terminowe := data rozpoczęcia + 6*długość studiów miesięcy + przedłużenie

jeśli dat pierwszego dyplomu na programie < wyliczone ukończenie terminowe to 1

w przeciwnym przypadku 0)

- DTERM2 - Dyplom w terminie w stosunku do ostatniego etapu. Istotne jest tylko to, czy student uzyskał dyplom na koniec ostatniego etapu studiów bez uwzględniania ewentualnych wcześniejszych opóźnień w studiowaniu.

dla programów o statusie DYP

jeśli data zakończenia ostatniego etapu <= data pierwszego dyplomu to 1

w przeciwnym przypadku 0

- D_A_PLAN - opóźnienie / przyspieszenie w ukończeniu studiów w miesiącach

Sufit z różnicy pomiędzy datą ukończenia studiów i wyliczonym ukończeniem studiów

liczby ujemne to przyspieszenie

patrz DTERM1:

dla DTERM1 = 1 liczba miesięcy pomiędzy maksimum z daty zakończenia dla cykli przy etapach osoby, a datą egzaminu dyplomowego

dla DTERM1 = 0 liczba miesięcy pomiędzy wyliczonym terminem ukończenia, a datą egzaminu dyplomowego

- ZAL1 - Czy pierwszy rok studiów zaliczony?

1 - zaliczony,

0 - niezaliczony,

null – osoba nie była nigdy wpisana na dany rok;

Jeśli na programie obowiązuje rozliczenie semestralne, to brany jest pod uwagę semestr parzysty (letni). Wpisanie tylko na semestr nieparzysty (zimowy) traktowane jest jako rozpoczęcie studiów na danym roku.

- ZAL2 - Czy drugi rok studiów zaliczony?

Procedura wyznaczania analogiczna jak w przypadku pierwszego roku.

- ZAL3 - Czy trzeci rok studiów zaliczony?

Procedura wyznaczania analogiczna jak w przypadku pierwszego roku.

- ZAL4 - Czy czwarty rok studiów zaliczony?

Procedura wyznaczania analogiczna jak w przypadku pierwszego roku.

- ZAL5 - Czy piąty rok studiów zaliczony?

Procedura wyznaczania analogiczna jak w przypadku pierwszego roku.

- ZAL6 - Czy szósty rok studiów zaliczony?

Procedura wyznaczania analogiczna jak w przypadku pierwszego roku.

Zmienne opisujące trajektorie studiowania na Uniwersytecie Warszawskim:

Tylko dla studiów II stopnia (zakończenie studiów na wcześniejszym programie musi nastąpić przed rozpoczęciem studiów na danym programie):

- UCZ – czy student przed rozpoczęciem studiów na danym programie ukończył studia I stopnia na tej samej uczelni (czy dla studenta istnieją studia o statusie DYP ukończone przed rozpoczęciem danego programu)
- UCZROK - czy student przed rozpoczęciem studiów na danym programie, w roku rozpoczęcia studiów na danym programie ukończył studia I stopnia na tej samej uczelni (czy dla studenta istnieją studia o statusie DYP ukończone w roku rozpoczęcia danego programu)
- UCZJ - czy student przed rozpoczęciem studiów na danym programie ukończył studia I stopnia w tej samej jednostce dydaktycznej (czy dla studenta istnieją studia o statusie DYP ukończone przed rozpoczęciem danego programu, które administrowane są przez tę samą jednostkę, co dany program studiów)
- UCZJROK - czy student przed rozpoczęciem studiów na danym programie ukończył studia I stopnia w tej samej jednostce dydaktycznej w roku rozpoczęcia studiów na danym programie (czy dla studenta istnieją studia o statusie DYP ukończone w tym samym roku, przed rozpoczęciem danego programu, które administrowane są przez tę samą jednostkę, co dany program studiów)

Tylko dla studiów I stopnia. Rozpoczęcie studiów na kolejnym programie musi nastąpić po ukończeniu studiów I stopnia.

- KONTUCZ - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia na tej samej uczelni (czy dla studenta istnieje program studiów II stopnia rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZS - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia stacjonarnych w innej jednostce dydaktycznej tej samej uczelni (czy dla studenta istnieje program studiów II stopnia stacjonarnych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZW - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia niestacjonarnych wieczorowych w innej jednostce dydaktycznej tej samej uczelni (czy dla studenta istnieje program studiów II stopnia niestacjonarnych wieczorowych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZZ - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia niestacjonarnych zaocznych w innej jednostce dydaktycznej tej samej uczelni (czy dla studenta istnieje program studiów II stopnia niestacjonarnych zaocznych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZJS - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia stacjonarnych w tej samej jednostce dydaktycznej (czy dla studenta istnieje program studiów II stopnia stacjonarnych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZJW - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia niestacjonarnych wieczorowych w tej samej jednostce dydaktycznej (czy dla studenta istnieje program studiów II stopnia niestacjonarnych wieczorowych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)
- KONTUCZJZ - czy po zakończeniu danych studiów student kontynuował naukę na studiach II stopnia niestacjonarnych zaocznych w tej samej jednostce dydaktycznej (czy dla studenta istnieje program studiów II stopnia niestacjonarnych zaocznych w innej jednostce rozpoczynający się po zakończeniu danego programu studiów)

- ERASMUS – Czy student kiedykolwiek w trakcie studiów na danej uczelni (na dowolnym programie) wyjechał na wymianę w ramach programu ERASMUS (czy osoba kiedykolwiek była na wyjeździe - istnieje wpis w tabeli dz_wyjazdy)
- STYPEND - Czy osoba dostawała stypendium naukowe w ramach tego programu (w dowolnym momencie). Brane są pod uwagę wyłącznie stypendia o kodach POL-on STS03, STS04.
- LPOWT - liczba powtarzanych etapów studiów (liczba etapów studiów na których osoba była więcej niż raz)
- LWAR - liczba zaliczonych warunkowo etapów studiów (liczba etapów o statusie S – zaliczony warunek)

Dane dla pierwszego dyplomu pobierane z certyfikatu:

- OCNADYP - ocena na dyplomie
- DYPEGZOC - ocena z egzaminu dyplomowego
- DYPOC - ocena z pracy dyplomowej
- STUDWYN - ogólny wynik studiów (funkcja oceny z egzaminu dyplomowego, pracy dyplomowej oraz średniej ze studiów).
- ESTUDOC - średnia ocen ze studiów
- LDYP - liczba dyplomów uzyskanych w ramach programu studiów
- EROK1 - średnia ocen z przedmiotów podpiętych do pierwszego roku studiów uwzględniane są wszystkie instancje danego programu.

Jeśli przedmiot jest podpięty pod więcej niż jedną instancję danego programu to przedmiot liczony jest tylko w jednej - pierwszej - decyduje kolejność programów wg ID

- EROK2 - średnia ocen z przedmiotów podpiętych do drugiego roku studiów uwzględniane są wszystkie instancje danego programu.
- EROK3 - średnia ocen z przedmiotów podpiętych do trzeciego roku studiów uwzględniane są wszystkie instancje danego programu.
- EROK4 - średnia ocen z przedmiotów podpiętych do czwartego roku studiów uwzględniane są wszystkie instancje danego programu.
- EROK5 - średnia ocen z przedmiotów podpiętych do piątego roku studiów uwzględniane są wszystkie instancje danego programu.

Wyniki studenta zrelatywizowane do osiągnięć pozostałych studentów. Ranga zdającego to: $(\text{liczba osób z gorszym wynikiem} + 0.5 * (\text{liczby osób o tym samym wyniku} - 1)) / (\text{liczba zdających} - 1)$ (uwzględniane są tylko przedmioty gdzie było więcej niż jeden zdający)

- ERANGA1 - średnia ranga z przedmiotów podpiętych do pierwszego roku studiów (wybór przedmiotów tak jak w przypadku obliczania średniej ocen)

- ERANGA2 - średnia ranga z przedmiotów podpiętych do drugiego roku studiów (wybór przedmiotów tak jak w przypadku obliczania średniej ocen)
- ERANGA3 - średnia ranga z przedmiotów podpiętych do trzeciego roku studiów (wybór przedmiotów tak jak w przypadku obliczania średniej ocen)
- ERANGA4 - średnia ranga z przedmiotów podpiętych do czwartego roku studiów (wybór przedmiotów tak jak w przypadku obliczania średniej ocen)
- ERANGA5 - średnia ranga z przedmiotów podpiętych do piątego roku studiów (wybór przedmiotów tak jak w przypadku obliczania średniej ocen)
- RANGSTUD - średnia ranga z przedmiotów podpiętych do wszystkich lat studiów.

- ECTS1 - suma punktów ECTS z przedmiotów podpiętych do pierwszego roku (wybór przedmiotów tak jak w przypadku liczenia średniej)
- ECTS2 - suma punktów ECTS z przedmiotów podpiętych do drugiego roku (wybór przedmiotów tak jak w przypadku liczenia średniej)
- ECTS3 - suma punktów ECTS z przedmiotów podpiętych do trzeciego roku (wybór przedmiotów tak jak w przypadku liczenia średniej)
- ECTS4 - suma punktów ECTS z przedmiotów podpiętych do czwartego roku (wybór przedmiotów tak jak w przypadku liczenia średniej)
- ECTS5 - suma punktów ECTS z przedmiotów podpiętych do piątego roku (wybór przedmiotów tak jak w przypadku liczenia średniej)

3. Bezpieczeństwo i poufność danych

Ponieważ narzędzie jest wbudowane w Uniwersyteckie System Obsługi Studiów, to skorzystanie z niego wymaga posiadać konta w systemie. Wiąże się to z koniecznością uzyskania uprawnień do przetwarzania danych gromadzonych w rejestrach uczelni. Dzięki temu prezentowane narzędzie spełnia wszystkie restrykcyjne wymogi odnośnie przetwarzania danych osobowych i prywatności badanych, które spełnia USOS.

4. Wnioski metodologiczne i rekomendacje

4.1. Jakość danych w USOS

Podstawową rekomendacją związaną z prezentowanym narzędziem służącym do eksportu informacji z USOS jest zalecenie budowania kultury wykorzystania rejestrów oraz wynikającej z niej dbałości o jakość wprowadzanych danych. Podobne zalecenia sformułowane były już na etapie raportu ze *Studium współzależności procesów dydaktycznych oraz funkcjonowania elektronicznego systemu obsługi studiów*. Bez odpowiedniej jakości danych w rejestrach uczelni nie może powstać zbiór opisujący faktyczne ścieżki edukacyjne studentów na uczelniach objętych badaniem.

Opisane powyżej narzędzie jest przygotowane przy założeniu, że dane wprowadzane są do systemu prawidłowo. Oznacza to np. że zakłada się, że każdy student powinien zostać rozliczony, że nie powinny zdarzać się sytuacje, że etap studiów jest „otwarty”, mimo że upłynęło już wiele lat od momentu, kiedy powinien się zakończyć.

Nieprawidłowości we wprowadzaniu do USOS informacji o przebiegu studiów odbijają się na jakości wyeksportowanych danych. Jest to szczególnie istotne ze względu na charakter eksportowanych zmiennych. Znaczna część z nich nie jest tylko wartością wyeksportowaną z określonego pola w bazie danych uczelni, ale stanowi efekt obliczeń prowadzonych na podstawie bardzo wielu informacji pochodzących z USOS (np. rangi obliczone na podstawie ocen uzyskanych na podstawie poszczególnych lat). Dane nie są w żaden sposób „czyszczone” z ewentualnych błędów powstałych przy wprowadzaniu do systemu. Część nieprawidłowości można wykryć analizując niespójności w wyeksportowanych danych, ale wykrycie błędów w przypadku zmiennych może być opartych o informacje z wielu pól może być niemożliwe.

Przed skorzystaniem z narzędzia należy więc przeanalizować stan posiadanych przez uczelnię rejestrów. Konieczne jest upewnienie się, jakie typy informacji i od jak dawna są gromadzone w systemie uczelnianym w sposób pozwalający na ich wykorzystanie w badaniach.

Przedstawione narzędzie do eksportu danych z USOS może być także elementem budowania kultury korzystania z rejestrów. Przede wszystkim świadomość, iż dane będą potem analizowane, powinna powodować, że osoby odpowiadające za ich wprowadzanie będą robić to dokładniej. Dodatkowo przygotowane eksporty mogą posłużyć do wykrycia części błędów w rejestrach uczelni.

4.2. Rozszerzenie zawartości zbiorów

Obecnie eksportowane są wyłącznie informacje dotyczące przebiegu studiów. Do pełnego opisu ścieżki kształcenia brakuje danych z okresu rekrutacji, czyli przede wszystkim liczby punktów rekrutacyjnych oraz w przypadku studiów pierwszego stopnia i jednolitych magisterskich wyników maturalnych. Dodatkowo warto wykorzystać znajdującą się w danych rekrutacyjnych informację o pochodzeniu geograficznym kandydata. Wszystkie wspomniane informacje zbierane są w rejestrach systemu Internetowej Rejestracji Kandydatów (IRK). Niestety, ze względów technicznych nie jest możliwe włączenie informacji pochodzących z rejestrów IRK do eksportu. System IRK posiada osobną bazę danych, która co roku tworzona jest od nowa. Wydaje się, że najprostszym rozwiązaniem pozwalającym na dodanie do eksportowanego zbioru danych z okresu rekrutacji jest wcześniejsze ich przekopiowanie do USOS. Co roku po zakończeniu rekrutacji część informacji dotyczących przyjętych kandydatów jest importowana do USOS. W ramach takiego importu mogłyby być kopiowane dodatkowe informacje jak liczba punktów rekrutacyjnych, czy wyniki maturalne. Dodanie informacji do USOS pozwoli na względnie łatwe wyeksportowanie ich wraz z pozostałymi zmiennymi opisującymi przebieg studiów.

4.3. Prowadzenie badań reprezentacyjnych

Przygotowane narzędzie zostało stworzone przede wszystkim z myślą o prowadzeniu badań wyczerpujących – dotyczących całej populacji studentów i absolwentów danej uczelni. Przygotowany zbiór danych może jednak posłużyć jako operat w badaniu reprezentacyjnym. Dzięki temu, iż zestaw zmiennych opisuje pełną ścieżkę kształcenia, a jednocześnie pozwala odzwierciedlić strukturę kształcenia (jednostki, kierunki studiów i ich cechy), możliwe jest losowanie prób według złożonych schematów doboru na różnych poziomach agregacji.