

Wielki Zderzacz Hadronów (LHC) i inne duże eksperymenty z dziedziny fizyki cząstek elementarnych w przeszłości, obecnie i w przyszłości, stanowią ogromne inwestycje publiczne w naukę podstawową. Jednakże, podczas gdy mechanizmy analizy i publikacji ich danych są wystarczająco dobrze zdefiniowanych dla standardowych pomiarów takich jak odkrycie bozonu Higgsa w 2012 roku (oraz bozonów W, Z czy wcześniej gluonów), to obecnie nie jest jasne jak takie mechanizmy powinny wyglądać w przypadku badania bardziej subtelnych zjawisk.

W ciągu 10 lat od odkrycia bozonu Higgsa, w LHC opublikowano wiele analiz sprawdzających granice Modelu Standardowego (SM) - ugruntowanego, ale przypuszczalnie niekompletnego modelu fizyki cząstek. Każdy taki do tej pory wykonany pomiar statystycznie wyklucza pewne uproszczone modele fizyki poza SM. Jednak takie modele nie są a priori bardziej prawdopodobne niż modele bardziej złożone. Te ostatnie charakteryzują się mieszaniną nowych mechanizmów zaczerpniętych z modeli uproszczonych, ale z mniejszą ich intensywnością, zatem nie ograniczają się do pojedynczego mechanizmu. Badanie takich złożonych modeli wymaga zmiany sposobu interpretacji wyników LHC. Nacisk musi zostać przeniesiony na łączenie pomiarów z wielu różnych typów analiz i tworzenie meta-analiz. Tylko takie globalne podejście może optymalnie wykorzystać dane LHC.

Proponowany projekt będzie stanowił krok w kierunku zbudowania infrastruktury potrzebnej do dokonania tej zmiany. Ułatwi on eksperymentom dostarczenie szybkich, możliwych do ponownego wykorzystania algorytmów analizy danych poprzez udoskonalenie języka specyficznego dla dziedziny fizyki cząstek oraz przygotowanie zestawów narzędzi do analizy ich danych. Narzędzia te zostaną połączone z dostępnymi publicznymi repozytoriami danych badawczych i metadanych. Zaprojektowane zostaną wspólne interfejsy dla kontroli wykonanych analiz w wielu zestawach narzędziowych, a także dla statystycznego łączenia tysięcy pomiarów, których kombinacje mogą dostarczyć najmocniejszego naukowego potwierdzenia jednej z teorii poza SM. W momencie rozpoczęcia trzeciego dużego cyklu zbierania danych przez LHC nadszedł czas, by wprowadzić te mechanizmy i rozwiązania, tak by spuścizna LHC była publicznie zachowana do ponownego wykorzystania w przyszłości.

Celem projektu jest zwiększenie zakresu, w jakim ogólnodostępne dane analiz z eksperymentów fizyki cząstek elementarnych (w sensie ogólnym, ale w szczególności wyniki zbiorcze, takie jak te wykorzystywane w publikacjach i wnioskowaniu statystycznym, natomiast nie obejmuje nieprzetworzonych przypadków zderzeń) mogą być łączone i ponownie wykorzystywane do testowania teorii Nowej Fizyki. Testy te, prowadzone zarówno przez teoretyków, jak i eksperymentatorów, mogą również wykraczać poza fizykę cząstek i łączyć się z astrofizyką i kosmologią, fizyką jądrową i bezpośrednimi poszukiwaniami ciemnej materii. Wartość łączenia informacji pochodzących z różnych indywidualnych analiz stała się jasna już na początku programu LHC. Wczesne dane eksperymentalne okazały się kluczowe dla poprawy modeli fizyki SM. Natomiast ogromna objętość danych, co raz większa złożoność modeli i zwiększona precyzja pomiarów w LHC wymaga kontynuowania tego podejścia w sposób bardziej systematyczny i skalowalny, otwarty dla całej społeczności i obejmujący wykorzystanie referencyjnych zbiorów danych w celu zapewnienia ich przydatności w dalekiej przyszłości.

Czas na ten krok jest odpowiedni, ponieważ kluczowe technologie (tworzenie i śledzenie DOI, RESTful Web APIs, kontrola wersji z ciągłą integracją, konteneryzacja) w ostatnich 5 latach stały się dojrzałymi rozwiązaniami. Fizyka cząstek elementarnych ma już ustanowione otwarte repozytoria danych i publikacji, ale kluczowe jest ich połączenie z uniwersalnymi łatwo skalowalnymi algorytmami analizy, a także uregulowanie zasad dostarczania danych i zaangażowanie eksperymentów w zasady FAIR dla otwartych danych naukowych. Poszczególni fizycy są na ogół entuzjastycznie nastawieni do takich pomysłów, czego dowodem jest przyjęcie polityki otwartych danych w laboratoriach zajmujących się fizyką cząstek elementarnych oraz zachowanie pełnego oprogramowania LHC. Potrzebny jest jednak wyraźny, finansowany wysiłek, by wyeliminować bariery techniczne i sprawić, by te pożądane działania były bardziej dostępne i opłacalne.