

MACIEJ STROIŃSKI*, JAN WĘGLARZ**

Znaczenie e-infrastruktury dla nauki

1. Wstęp

Współczesna nauka ma wymiar globalny, i o ile globalizacja w znaczeniu ogólnym może budzić kontrowersje, o tyle w nauce jest ona zjawiskiem jednoznacznie pozytywnym.

Naukowcy powinni bowiem mieć równe szanse, co oznacza, że powinni mieć jednakowy dostęp do tych samych informacji i możliwości ich przetwarzania oraz do laboratoriów. Uzyskane wyniki pracy naukowej winny być powszechnie dostępne, gdy tylko ich autorzy zdecydują się na ich ujawnienie. Łatwość kojarzenia i wspólnej pracy zespołów międzynarodowych i interdyscyplinarnych winna powodować przyspieszenie uzyskania istotnie nowych wyników.

Warunkiem koniecznym praktycznej realizacji takiej wizji światowej społeczności naukowej jest powszechna dostępność odpowiedniej infrastruktury informatycznej nauki. Z tego powodu rozwój tej infrastruktury w wielu krajach jest obejmowany specjalnymi programami. W ich efekcie powstaje wielopoziomowa architektura narzędzi i metod informacyjno-komunikacyjnych (ang. *Information and Communication Technology, ICT*), w ramach której powstają aplikacje i usługi dla obsługi zarówno całego środowiska naukowego, jak i poszczególnych dyscyplin nauki.

W artykule omówiono budowę europejskiej infrastruktury informatycznej nauki, nazywanej też e-infrastrukturą (ang. *e-infrastructure*), ze szczególnym uwzględnieniem wkładu polskiego środowiska naukowego. Znaczenie tej infrastruktury dla wybranych dyscyplin nauki zilustrowano przykładami z obszaru radioastronomii, fizyki wysokich energii i medycyny. Wskazano także kierunki dalszego rozwoju infrastruktury informatycznej nauki, jak i jej możliwego wykorzystania przez inne dyscypliny nauki.

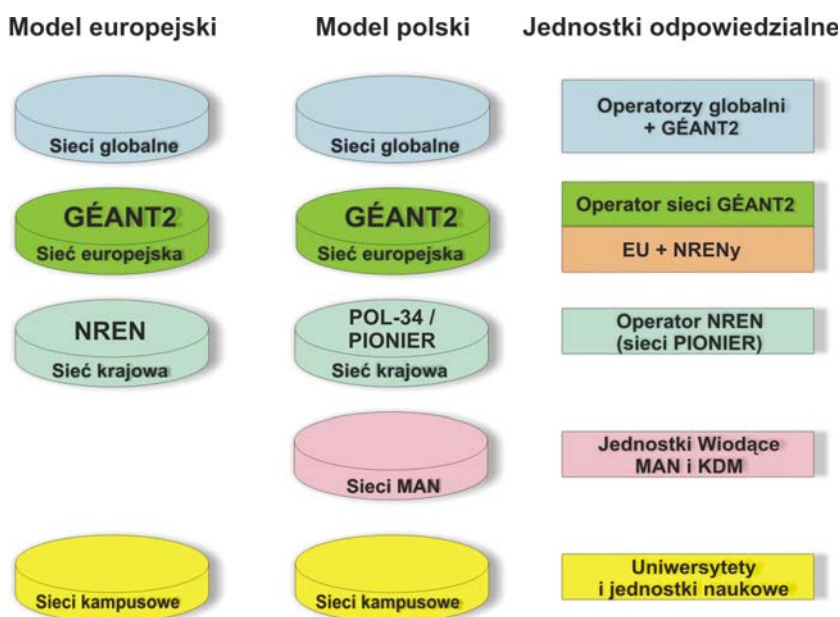
2. Infrastruktura informatyczna nauki

Duże znaczenie infrastruktury informatycznej nauki spowodowało, że w wielu krajach powstały specjalne programy związane z jej rozwojem. Budowę infrastruktury informatycznej rozpoczęto u nas 15 lat temu, realizując naukowe sieci światłowodowe w głównych ośrodkach akademickich, tzw. sieci MAN (ang. *Metropolitan Area Network*). Obecnie jest ich 21. W sieciach tych istnieje infrastruktura dla wytwarzania usług. Tym różni

* dr inż. Maciej Stroiński, Poznańskie Centrum Superkomputerowo-Sieciowe, Instytut Informatyki Politechniki Poznańskiej

** prof. dr hab. inż. Jan Węglarz, członek rzeczywisty PAN, Instytut Informatyki Politechniki Poznańskiej, Poznańskie Centrum Superkomputerowo-Sieciowe

się model polski od modelu europejskiego (ryc. 1), w którym sieci kampusowe dołączone są bezpośrednio do sieci krajowej (tzw. NREN: *National Research and Education Network*). Powstało także 5 centrów komputerów dużej mocy.



Ryc. 1. Modele infrastruktury informatycznej nauki

W efekcie doświadczeń polskiego środowiska naukowego w dziedzinie budowy sieci MAN w głównych ośrodkach akademickich oraz sieci krajowej POL-34 [1], powstał w 1999 r. projekt PIONIER, zaakceptowany przez KBN do realizacji w roku 2000. Zakres tego programu jest dobrze zdefiniowany jego pełną nazwą: „PIONIER: Polski Internet Optyczny. Zaawansowane Aplikacje, Usługi i Technologie dla Społeczeństwa Informatycznego” [2]. Jednym z ważnych składników tego programu była budowa ogólnopolskiej światłowodowej sieci PIONIER. W 2001 r. w USA, *National Science Foundation* ogłosiła realizację podobnego programu o nazwie *Cyberinfrastructure* [3]. Równocześnie w Europie, opierając się efektach 5. Programu Ramowego i rozpoczętego w 2002 roku 6. Programu Ramowego, zdefiniowano w 2003 r. koncepcję zintegrowanej infrastruktury informatycznej dla nauki i edukacji pod nazwą *e-Infrastructure* [4], która obejmowała dwie warstwy. Szerokopasmowe sieci nowej generacji tworzyły warstwę pierwszą, a rozproszone systemy dostępu do zasobów obliczeniowych, pamięci i archiwizacji oraz instrumentów badawczych tworzyły warstwę drugą. Zintegrowane w warstwie drugiej metody i narzędzia tworzą złożony system rozproszony, który nazywamy gridem. Z tego powodu, mówiąc o e-infrastrukturze, obejmujemy zarówno infrastrukturę sieciową, jak i gridową (ryc. 2).



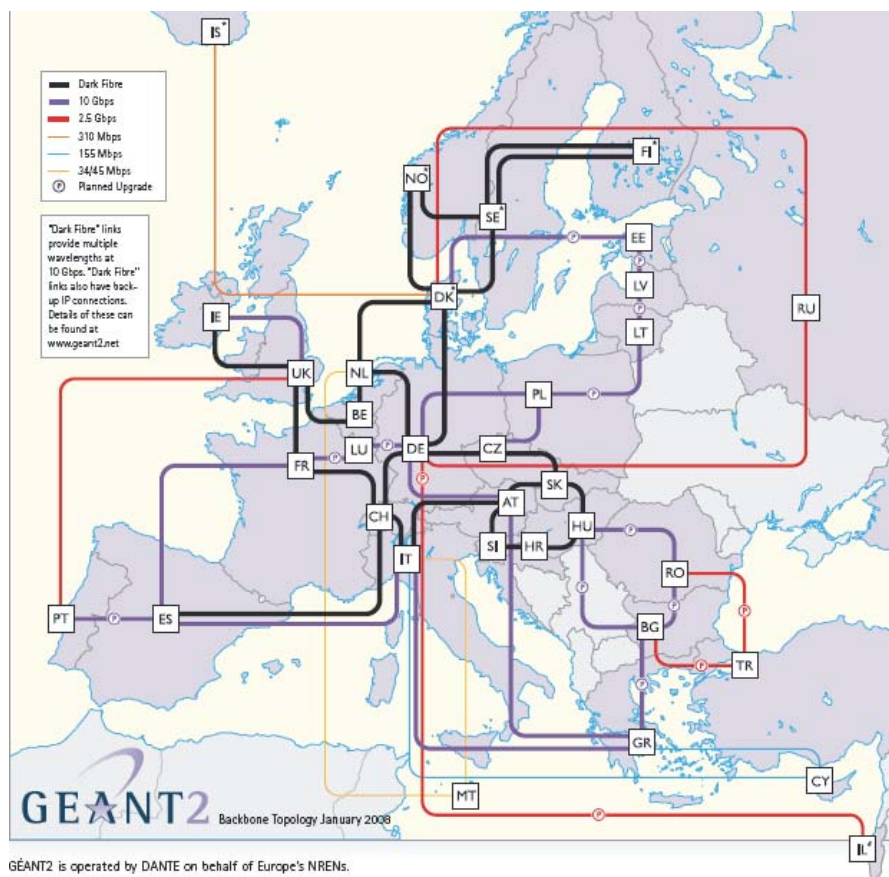
Ryc. 2. Europejska i polska e-infrastruktura

Koncepcja e-infrastruktury, rozbudowana w ramach 6. Programu Ramowego, potwierdziła swoją użyteczność dla wielu dziedzin nauki, które bez jej wykorzystania nie miałyby szans rozwoju. Zdobyte doświadczenia pozwoliły zdefiniować, jako jeden z kierunków prac w 7. Programie Ramowym, dodanie do e-infrastruktury nowej warstwy – warstwy danych naukowych (ang. *Scientific Data Layer*) [5]. Warstwa ta wyłącza, z ogólnie zdefiniowanego obszaru warstwy aplikacji, funkcje wspólne dla wielu zastosowań umieszczane powyżej warstwy gridowej. Program PIONIER oraz prace polskich zespołów naukowych w projektach 5. i 6. Programu Ramowego w obszarze sieci i gridów, wniosły partnerski wkład w budowę europejskiej infrastruktury informatycznej nauki, która stanowi bazę dla europejskiej przestrzeni badawczej. W dalszej części tego rozdziału omówimy poszczególne składniki tej e-infrastruktury.

2.1. Infrastruktura sieciowa

Obecnie infrastruktura sieciowa składa się z sieci paneuropejskiej GÉANT2 i powiązanych z nią narodowych sieci naukowych, głównie optycznych, takich jak np. PIONIER, CESNET, NORDUnet. Początkiem dzisiejszej infrastruktury sieciowej w Europie była sieć GÉANT, powstała w ramach 5. Programu Ramowego jako następcza sieci europejskiej TEN-155. Sieć GÉANT bazowała na szybkich kanałach komunikacyjnych w technologii SDH, wydierzawionych od operatorów telekomunikacyjnych. Sieć rozpoczęła

swoje działanie w roku 2001. W Polsce w tym czasie działała sieć POL-34, budowana według identycznych zasad, przy czym maksymalną przepustowością było 622 Mb/s. Do budowy tej sieci zastosowano tradycyjne technologie telekomunikacyjne SDH i ATM, powyżej których pracował protokół IP. Rozwój technologii światłowodowych spowodował pojawienie się bardzo wydajnych i niezawodnych technologii transmisyjnych DWDM (ang. *Dense Wavelength Division Multiplexing*), które wymusiły rewolucyjne przemiany w architekturze systemów komunikacyjnych. Duża przepustowość i niezawodność kanałów optycznych (tzw. lambda) nie wymagała stosowania dodatkowych protokołów zapewniających zwiększenie bezpieczeństwa i niezawodności transmisji, co umożliwiło zastosowanie w sieciach optycznych powyżej technologii DWDM prostej metody ramkowania pakietów (np. Ethernet), dzięki czemu protokół IP „zbliżył się” do światłowodu. Doprowadziło to do bardzo znaczącego uproszczenia zarządzania i tym samym potaniaenia transmisji. Uzyskiwane w tej transmisji popularne szybkości kanałów optycznych to 10 Gb/s.



Ryc. 3. Sieć GÉANT2 (źródło [6])

Ponadto technologia przełączalnego Ethernetu stosowana w tych sieciach, została wzbogacona o protokół MPLS (ang. *MultiProtocol Label Switching*) oraz VPLS (ang. *Virtual Private LAN Service*), co stwarza możliwość dynamicznego tworzenia w sieci fizycznej kanałów i sieci wirtualnych o różnych konfiguracjach. Co więcej, za pośrednictwem niedawno zdefiniowanego protokołu GMPLS (ang. *Generalized Multiprotocol Label Switching*) możliwe jest sterowanie kanałami transmisyjnymi na różnych poziomach architektury sieci, od światłowodu do kanałów wirtualnych. Dostępność kanałów szerokopasmowych przyczyniła się znacząco do rozwoju w takich sieciach rozproszonych systemów przetwarzania danych i wykorzystywania tych kanałów w różnych eksperymentach naukowych. Opisane technologie zastosowane zostały między innymi w sieci GÉANT2, która powstała w ramach 6. Programu Ramowego w 2004 roku (ryc. 3).

W Polsce prace wokół następcy sieci POL-34 rozpoczęły się w 1999 r. od budowy eksperymentalnej sieci optycznej na światłowodach uzyskanych od innych operatorów [7]. Prace te kontynuowane były w ramach projektów 5. i 6. Programu Ramowego. Obejmowały one problematykę: jakości usług sieciowych i badania nowych technologii (projekty SEQUIN [8], ATRIUM [9], 6NET[10], MUPBED[11]); usługi zarządzania pasmem na żądanie, architektury systemów pomiarowych i obsługi zdarzeń w sieciach nowych generacji (ng) (projekt GN2 [12, 13]) oraz zarządzanie sieciami ng (projekt EMANICS [14]). W projektach tych współuczestniczył zespół Poznańskiego Centrum Superkomputerowo-Sieciowego (PCSS). Równocześnie od 2001 r. rozpoczęto realizację ogólnopolskiej światłowodowej sieci PIONIER, jednej z najnowocześniejszych sieci na świecie (ryc. 4).



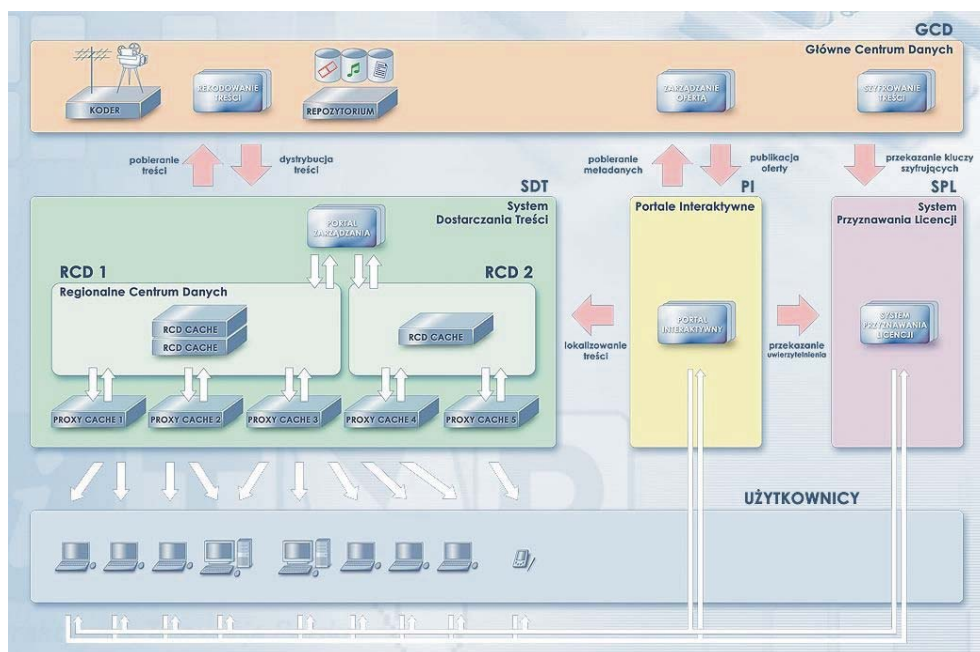
Ryc. 4. Sieć PIONIER

Sieć PIONIER, łącząca 21 MANów i 5 centrów Komputerów Dużej Mocy, uruchomiona w roku 2003, obejmuje 5200 km własnych linii światłowodowych, z czego 1200 km jest w trakcie budowy. Sieć bazuje na takich technologiach, jak: DWDM (do 32 λ 10 Gb/s), MPLS, VPLS, IPv6 i oferuje usługi BoD (ang. *Bandwidth on Demand*) oraz wymaganą QoS (ang. *Quality of Service*).

Z siecią PIONIER zintegrowane są także zaawansowane usługi transmisji strumieniowej. Realizujący te usługi system SDT [15, 16] stanowi podstawę do budowy zaawansowanych systemów użytkowych, jak np. Platforma iTVP [17].

Platformę iTVP (interaktywna telewizja bazująca na protokole IP) tworzy kilka systemów współpracujących ze sobą i realizujących zadania w rozdzielnych obszarach funkcjonalnych (ryc. 5). Są to m.in.:

- Główne Centrum Danych (GCD);
- System Dostarczania Treści (SDT);
- System Przyznawania Licencji (SPL);
- Portal Interaktywny.



Ryc. 5. Architektura Platformy iTVP

Zasadniczy składnik platformy – System Dostarczania Treści – jest systemem odpowiedzialnym za dystrybucję materiałów audiowizualnych pochodzących z systemów GCD oraz ich dostarczenie do odbiorców. Realizuje usługi transmisji treści w trybach: „na

żywo”, „na żywo z przesunięciem czasowym”, „na żądanie” oraz „na zamówienie”. Jest on systemem hierarchicznym dwupoziomowym, osadzonym w sieci PIONIER, opartym o strukturę Regionalnych Centrów Danych (RCD, I poziom dystrybucji) oraz urządzeń proxy/cache (II poziom dystrybucji) – przyłączonych do odpowiednich RCD i zlokalizowanych blisko użytkowników końcowych (w sieciach operatorów Internetu ISP). Zadaniem systemu SDT jest zapewnienie skalowalnego i efektywnego dostępu do treści multimedialnej dla dużej liczby odbiorców. System SDT implementuje mechanizmy dystrybucji treści, wymiatania oraz replikacji i migracji zasobów cyfrowych. System ten zapewnia mechanizmy lokalizowania treści na rzecz wybranego użytkownika oraz jej udostępniania z wysoką jakością. Ponadto wyposażony został w mechanizmy monitorowania stanu, co pozwala dokonywać równoważenia obciążenia w systemie.

Sieć PIONIER stanowi podstawę do tworzenia zaawansowanych usług dla nauki. Równocześnie posiada ona bezpośrednie połączenia z sieciami naukowymi krajów sąsiednich, co stanowi obecnie w Europie ważny paradygmat budowania kolejnej sieci paneuropejskiej GÉANT3 w ramach 7. Programu Ramowego. Bezpośrednie połączenia światłowodowe pomiędzy sieciami naukowymi tzw. CBF (ang. *Cross-Border Fiber*) zostały zaproponowane przez polskie środowisko naukowe [18, 19]. Polskie podejście do budowania krajowych sieci optycznych, połączonych z sieciami krajów sąsiednich, i tworzenie tzw. e-Regionów [20] stanowiło bazę dla realizacji studyjnego projektu europejskiego PORTA OPTICA [21] ukierunkowanego na włączenie do sieci europejskiej sieci naukowych krajów dawnej Europy Wschodniej i niektórych państw zakaukaskich.

Polskie inicjatywy w zakresie budowania światłowodowych sieci lokalnych i kampusowych, naukowych sieci miejskich czy krajowej sieci PIONIER umożliwiają realizację idei światłowodu „od końca do końca”. Zapewnia to środowisku naukowemu praktycznie nieograniczony, szerokopasmowy dostęp do przepustowości sieci i zasobów do tej sieci podłączonych. Drugą grupą technologii związanych z rozwojem komunikacji w środowisku akademickim są technologie łączności bezprzewodowej (np. WiFi). Gwałtowny rozwój tych technologii doprowadził do ich szerokiej dostępności, w związku z czym naukowiec nie musi być „przywiązany” do stacjonarnego komputera i może w dowolnym czasie i miejscu, w obszarze oddziaływania technologii WiFi, połączyć się ze swoją macierzystą siecią i kontynuować pracę lub nadzorować badania. Stworzony został system ułatwień dla światowego (europejskiego) roamingu i dostępu do sieci o nazwie Eduroam [22]. Powszechność stacji dostępowych WiFi umożliwia realizację połączeń w bardzo wielu obiektach prywatnych, publicznych, centrach miast itp. Jeśli do tych wymienionych doda się jeszcze możliwość łączności za pośrednictwem transmisji przez telefon komórkowy, to w efekcie otrzymujemy możliwość komunikacji „zawsze i wszędzie”. Lawinowy wzrost liczby połączeń światłowodowych oraz połączeń transmisji mobilnych staje się podstawowym paradygmatem rozwoju zaawansowanych systemów usługowych.

2.2. Infrastruktura gridowa

Rozproszone zasoby obliczeniowe, pamięci, urządzenia badawcze, źródła danych (np. sensory) stanowią bardzo ważny składnik infrastruktury informatycznej nauki. Szybkie sieci komputerowe pozwalają zbudować efektywną sieć połączeń tych rozproszonych składników. Powstaje jednak bardzo trudny problem współużytkowania tych zasobów przez wiele środowisk naukowych, projektów itp., z równoczesnym zapewnieniem adaptowania się tej infrastruktury do ich wymagań. Ponadto problemem jest kwestia dostosowania systemów do użytkowania przez szerokie gremium nieinformatyków. Aby rozwiązać te problemy, opracowano metody i narzędzia tworzące tzw. architekturę gridową [23]. Umożliwiają one wirtualizację zasobów z możliwością ich dynamicznego przydzielania wirtualnym organizacjom badawczym, zapewniając równocześnie łatwy i efektywny dostęp do tego środowiska [24]. Ponadto wymagane są takie cechy, jak: autonomiczność, niezawodność i bezpieczeństwo. Powstało wiele różnych systemów gridowych (ang. *Grid Middleware*), np. Globus Toolkit [25], Condor [26], Unicore [27]. Wszystkie te systemy oferują podobne usługi gridowe, chociaż zastosowane w nich technologie i standardy są różne. Mimo bardzo wysokiej złożoności tych systemów, ich funkcjonalność jest jeszcze ograniczona. Z tego względu bardzo duże znaczenie mają prace badawczo-rozwojowe w zakresie tej tematyki, prowadzone w 5. i 6. Programie Ramowym, a także w krajowych projektach celowych. Do najbardziej znanych rozwiązań należą architektury gridowe powstałe w projektach GridLab i CrossGrid.

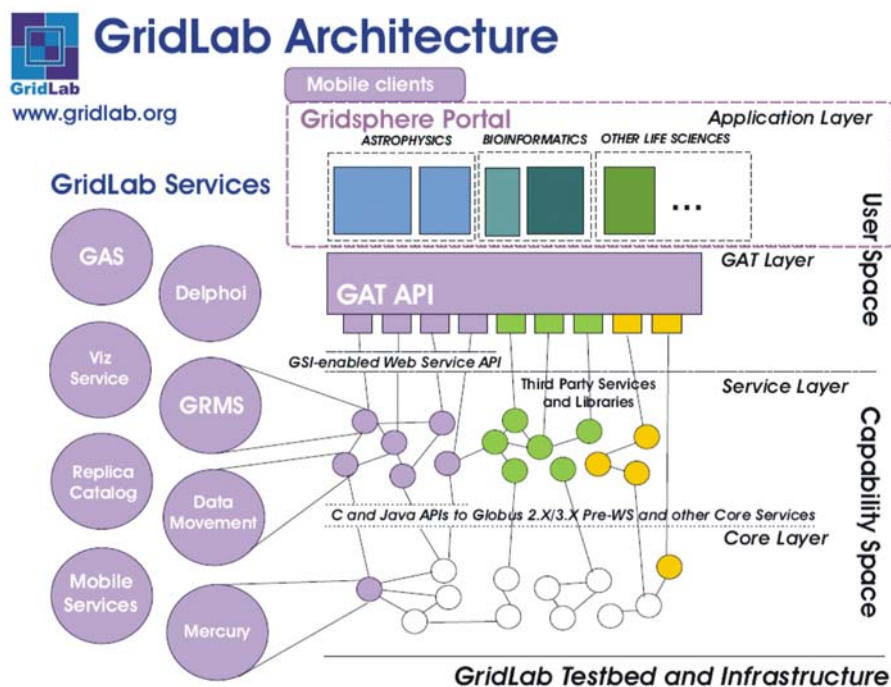
2.2.1. Architektura GridLab

Architektura GridLab (ryc. 6) powstała w ramach projektu koordynowanego przez Poznańskie Centrum Superkomputerowo-Sieciowe. Projekt GridLab (*Grid Application Toolkit and Testbed*) był jednym z największych projektów dotyczących rozwoju środowiska gridowego w Europie i przyczynił się znacząco do jego rozwoju na świecie [28].

Technologia gridowa, której poświęcony był projekt Gridlab, to model przetwarzania, pozwalający na maksymalne wykorzystanie rozproszonych, połączonych siecią zasobów obliczeniowych, włączając zarówno superkomputery, jak i małe systemy obliczeniowe, takie jak na przykład palmtopy. Sieć łącząca te systemy może być siecią optyczną, dowolną siecią w technologii ATM, siecią GbE, bądź nawet siecią bezprzewodową.

W projekcie nacisk położono na narzędzia umożliwiające rozwój aplikacji dla gridu, oparty o uniwersalne API do ich tworzenia: GAT – *Grid Application Toolkit* [29]. GAT przetestowano dla aplikacji Cactus [30], która jest otwartym systemem powszechnie używanym w symulacjach wielkiej skali oraz dla aplikacji TRIANA [31], wykorzystywanej przez astrofizyków w badaniach dotyczących fal grawitacyjnych. GAT umożliwia użytkownikom oraz twórcom aplikacji efektywne wykorzystanie zasobów gridu w rzeczywistym środowisku w bardzo prosty i wydajny sposób. Fakt, iż skupiono się na specyficznych, bardzo dobrze sprecyzowanych i powszechnie wykorzystywanych aplikacjach,

pozwolił na natychmiastowe utworzenie działających aplikacji dedykowanych dla gridów oraz zdobycie ogromnie przydatnego doświadczenia, które z kolei pozwoliło na opracowanie ogólniejszych rozwiązań, mających zastosowanie również w innych aplikacjach. GAT był podstawą standardu SAGA (*Wimpli API for Grid Applications*) opracowanego przez Open Grid Forum.



Ryc. 6. Architektura GridLab (źródło [28])

Drugim istotnym nurtem w projekcie GridLab było zaprojektowanie i zaimplementowanie zestawu zaawansowanych usług gridowych, z których mógłby korzystać użytkownik oraz działająca w gridzie aplikacja (przez GAT).

Najważniejszym osiągnięciem w tym zakresie była implementacja systemu do zarządzania zasobami – GRMS – umożliwiającego dynamiczny przydział zasobów do aplikacji w zależności od aktualnego stanu zasobów. System ten, działający jako niezależny serwis, w sposób automatyczny odkrywa dostępne węzły obliczeniowe i wykorzystując wielokryterialne algorytmy przedziela zadanie do zasobu. Interfejsy GRMSa dostępne były zarówno dla użytkownika, jak i dla aplikacji, co umożliwiło realizację zaawansowanych, dynamicznych scenariuszy.

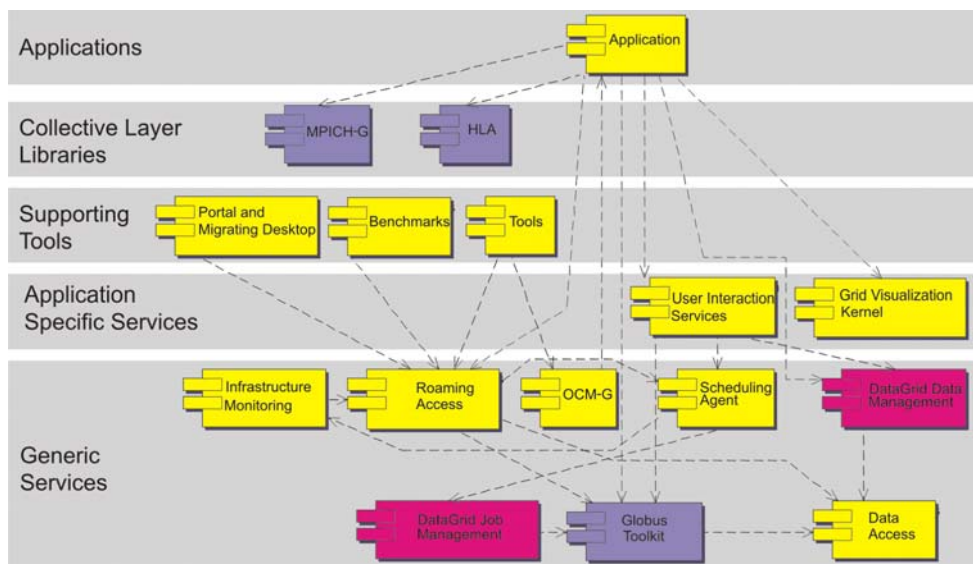
Pozostałe składniki warstwy pośredniej GridLab, to: system zarządzania danymi (*Data Management i Replica Catalog*), serwisy wspomagające użytkownika mobilnego

(*Mobile Services*), system monitorowania (*Mercury*), gridowy system informacyjny (*iGrid*). Cała infrastruktura programowa objęta była zunifikowaną infrastrukturą bezpieczeństwa, opartą na szeroko stosowanych standardach. Dodatkowym elementem był system autoryzacji GAS (*GridLab Authorization Service*), z którego korzystały wszystkie pozostałe komponenty.

Ważnym elementem projektu było też zaprojektowanie i przetestowanie infrastruktury oraz aplikacji gridowych w środowisku testowym europejskim oraz transatlantycznym. Interoperabilność między środowiskami pilotowymi została zapewniona przez współpracę ze środowiskiem amerykańskim i koreańskim. Testowanie środowiska było zapewnione zarówno przez członków projektu, jak i przez różne środowiska naukowe, takie jak np. Europejska Sieć Astrofizyczna (*European Astrophysics Network*) oraz wiele projektów interdyscyplinarnych finansowanych przez NSF.

2.2.2. Architektura CrossGrid

Architektura CrossGrid (ryc. 7) powstała w ramach projektu koordynowanego przez Cyfronet AGH [32]. Projekt CrossGrid (*Development of Grid Environment for Interactive Applications*), realizowany w latach 2002-2005, zajmował się trzema głównymi obszarami: zaawansowane narzędzia dla warstwy pośredniej środowiska gridowego, aplikacje interaktywne oraz budowa instalacji pilotowej heterogenicznego środowiska rozproszonego.



Ryc. 7. Architektura CrossGrid (źródło [33])

Prace projektowe rozpoczęły się od rozwoju nowych narzędzi dla środowiska gridowego, służących przede wszystkim do przeprowadzania obliczeń interaktywnych, wyko-

rzystujących duże wolumeny danych. Zastosowania te obejmowały między innymi symulacje i wizualizacje wspomagające pracę chirurgów, systemy wspomagania decyzji dla zespołów kryzysowych, rozproszoną analizę danych dla fizyki wysokich energii, prognozę pogody w połączeniu z symulacjami dotyczącymi zanieczyszczeń powietrza. Złożona metodologia, w ogólności architektura aplikacji, środowisko programowania i nowo wytworzone usługi były przedmiotem walidacji i gruntownego testowania w środowisku pilotowym CrossGrid, ze szczególnym naciskiem na prostotę obsługi środowiska pracy użytkownika.

Realizacja projektu przebiegała w ścisłej współpracy z DataGrid (koordynator CERN) w celu wykorzystania uzyskanych wyników i doświadczenia. Realizacja projektu przebiegała w ścisłej współpracy z DataGrid (koordynator CERN) w celu wykorzystania uzyskanych wyników i doświadczenia. Elementy warstwy pośredniej projektu DataGrid, jak zarządzanie danymi (*DataGrid Data Management*), zarządzanie zadaniami (DataGrid Job Management), połączone zostały systemem Globus Toolkit.

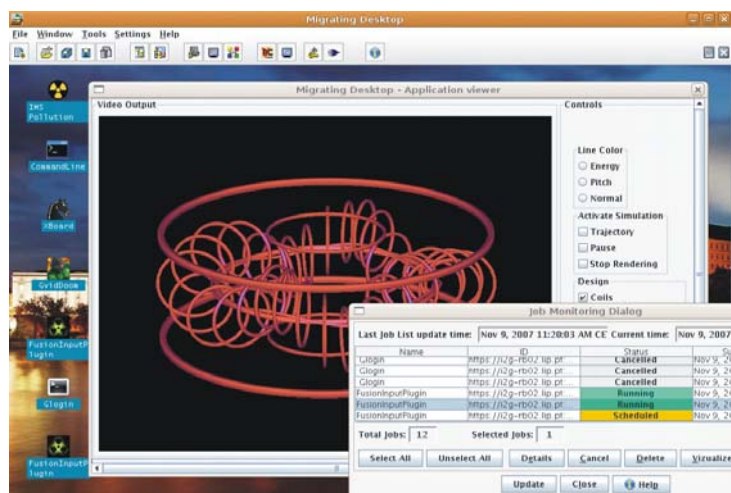
Jednym z ważnych narzędzi, które powstały w ramach prac rozwojowych w warstwie wspierającej aplikacje (ang. *Supporting Tools*), był migrujący desktop (ang. *Migrating Desktop* – MD), system pełniący rolę przyjaznego interfejsu dostępu aplikacji do środowiska rozproszonych zasobów obliczeniowych oraz danych [34]. MD jest intuicyjnym środowiskiem pracy użytkownika Gridu. Środowisko to znacznie ułatwia pracę ze złożoną infrastrukturą, dając dostęp do zaawansowanych technologii. System MD to graficzny interfejs użytkownika, umożliwiający kontrolę aplikacji, monitorowanie infrastruktury oraz aktywnych zadań, a także przeprowadzanie operacji na danych. Oprogramowanie to (zaimplementowane w języku Java) pośredniczy między użytkownikiem a usługami sieciowymi i umożliwia interaktywną pracę ze złożonymi aplikacjami. Działanie MD jest wspierane usługami z warstwy Generic Services (oznaczonych czerwonymi wypustkami na ryc. 7).

Główne cechy i zalety MD to:

- możliwość prostej integracji z aplikacjami gridowymi,
- prostota definiowania, zlecenia i monitorowania zadań,
- wsparcie dla zadań wsadowych i interaktywnych,
- wsparcie dla zadań sekwencyjnych i równoległych (Open MPI – w klastrze, PACX-MPI – między klastrami),
- intuicyjne mechanizmy do zarządzania danymi gridowymi (ftp, gridFTP, LFC, srm v2.2),
- niezależność od platformy sprzętowej (Java) i lokalizacji (migrujące środowisko),
- możliwość zdalnej, zaawansowanej wizualizacji i sterowania aplikacją.

Warto wymienić przykłady kilku aplikacji z różnych dziedzin nauki i o różnych charakterystykach zintegrowanych z MD, np. Visualization of Plasma Particles in Fusion De-

vices (ryc. 8), GAMESS, Natural Language Processing, Ultra Sound Computer Tomography, IMS Model Suite, Analysis of Cosmic Microwave Background.



Ryc. 8. Przykład integracji migrującego desktopu z aplikacją syntezy jądrowej (źródło [35])

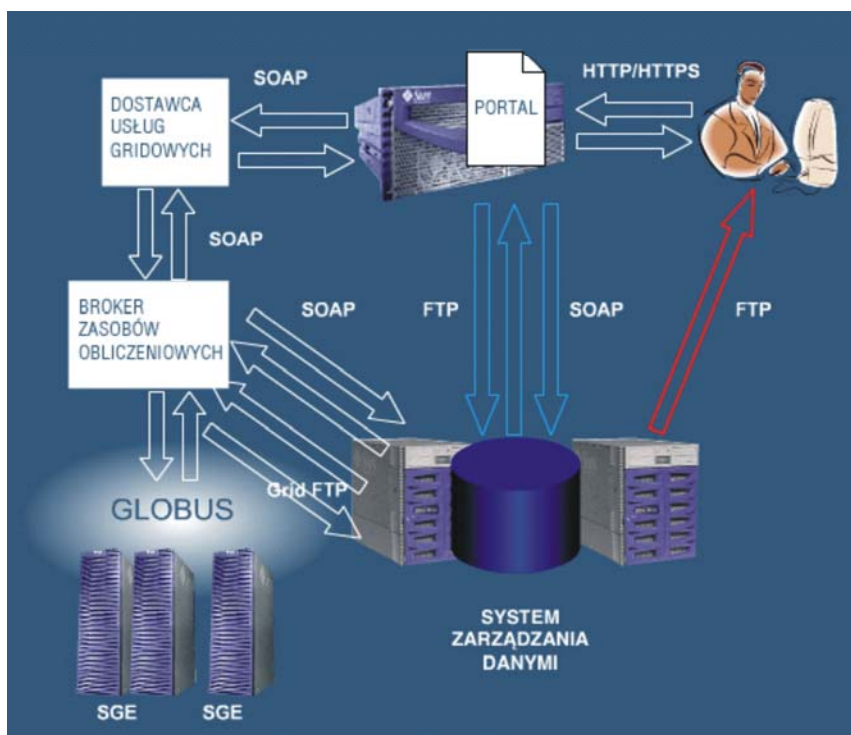
Migrujący Desktop, zrealizowany w PCSS, jest przykładem produktu, który po wytworzeniu w projekcie CrossGrid, znalazł szersze zastosowanie w innych wdrożeniach produkcyjnych (Irlandia, Hiszpania – w ramach narodowych inicjatyw gridowych), programach międzynarodowych, jak NorduGrid. MD znalazł również swoje zastosowania w projektach europejskich 6. i 7. Programu Ramowego: int.eu.grid, BalticGrid, Euforia oraz BalticGrid-II.

2.2.3. Polskie projekty gridowe

Prace badawczo-rozwojowe w obszarze gridów prowadzone były także w krajowych projektach celowych: PROGRESS, SGIGrid i Clusterix.

Projekt celowy „Stworzenie środowiska dostępu do usług obliczeniowych realizowanych przez klastr komputerów SUN” (PROGRESS) był realizowany wspólnie z firmą Sun Microsystems. Projekt miał na celu przygotowanie środowiska, w którym w wygodny sposób można prowadzić obliczenia i eksperymenty naukowe z wykorzystaniem zasobów komputerowych i sieciowych, połączonych w infrastrukturę gridową. W celu realizacji tak postawionego zadania opracowano i zaimplementowano architekturę środowiska gridowo-portalowego oraz zbudowano pilotową instalację klastra w środowisku sieci optycznej PIONIER [36-39]. Efektem projektu jest pakiet otwartych narzędzi do tworzenia środowisk dostępowych do usług i zasobów gridowych. Obejmują one: broker systemu zarządzania zasobami obliczeniowymi, dostawcę usług gridowych, system zarządzania danymi, środowisko zarządzania portalem (ryc. 9). Narzędzia te dopełniały

rozwiązania firmowe Sun Microsystems i były częścią oferty, którą ta firma prezentowała dla środowisk naukowych na świecie. Prace badawczo-rozwojowe przeprowadzone w ramach projektu PROGRESS były prezentowane na licznych konferencjach i wystawach, w tym na kolejnych edycjach wystawy Supercomputing w USA w latach 2002-2004.



Ryc. 9. Struktura środowiska gridowo-portalowego PROGRESS (źródło [40])

W ramach wdrożenia środowiska gridowo-portalowego PROGRESS w sieci PIONIER udostępnione zostały aplikacje obliczeniowe z zakresu bioinformatyki: asemblacja łańcuchów DNA, przewidywanie struktur drugorzędowych aminokwasów oraz budowa drzew filogenetycznych. Dodatkowo uruchomiono aplikację obliczeniową Gaussian 98. Środowisko narzędziowe PROGRESS stanowi obecnie podstawę budowy portalu telekonsultacji medycznych Telemedycyna Wielkopolska.

Projekt celowy PROGRESS był realizowany przez Poznańskie Centrum Superkomputerowo-Sieciowe, we współpracy z Akademickim Centrum Komputerowym AGH Cyfronet w Krakowie oraz Centrum Komputerowym Politechniki Łódzkiej. Prace prowadzone w ramach projektu zostały uznane za niezwykle interesujące przez Sun Microsystems. Zostało to potwierdzone podpisaniem w 2002 r. przez Sun Microsystems i PCSS umowy o utworzeniu Centrum Doskonałości (SUN Center of Excellence, CoE) w zakresie sieci nowych generacji, przetwarzania w gridach i portalu [41].

Kolejnym przykładem realizacji wspólnych projektów z prądującymi na rynku firmami IT był projekt SGGrid (Obliczenia wielkiej skali i wizualizacja do zastosowań w wirtualnym laboratorium z użyciem klastra SGI). Przedmiotem projektu realizowanego z firmami Silicon Graphics (USA) oraz ATM SA (Warszawa) było opracowanie koncepcji i implementacja nowoczesnych, szerokopasmowych usług zdalnego udostępnienia kosztownych urządzeń laboratoryjnych, centrum zapasowego oraz zdalnej wizualizacji danych.

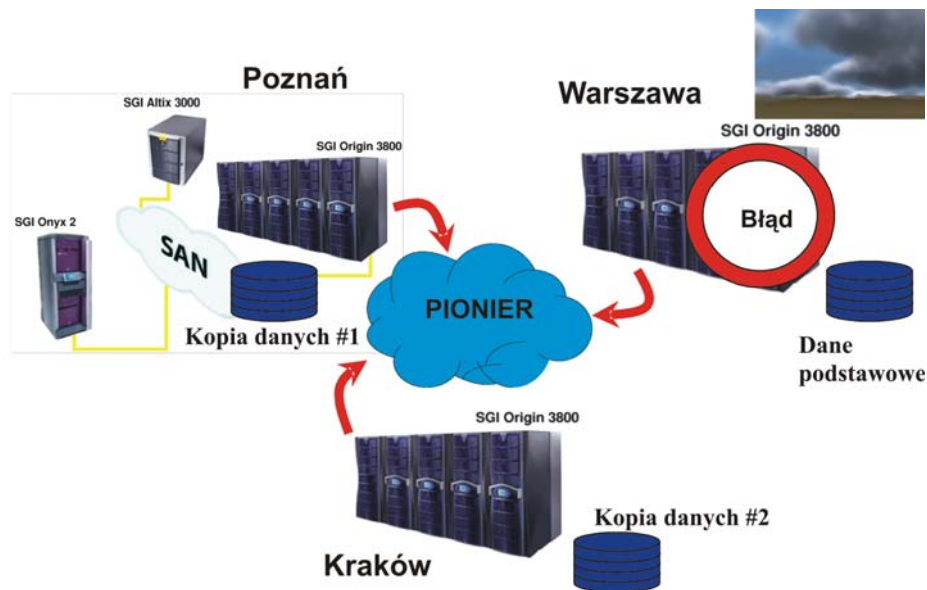
W ramach SGGrid opracowana została szczegółowa koncepcja i zrealizowana implementacja nowoczesnego, ogólnopolskiego środowiska (klastra) dla obliczeń dużej skali oraz dla zaawansowanej wizualizacji, z docelowym przeznaczeniem do zdalnej obsługi kosztownych urządzeń laboratoryjnych (laboratorium wirtualne) oraz systemu zapasowego centrum obliczeniowego, potrafiącego przejąć w czasie rzeczywistym określone zadania z możliwością wykonania ich w ściśle określonym czasie. Najważniejszym od strony naukowej wkładem tego projektu było oprogramowanie warstwy pośredniej, przez co każdy użytkownik klastra uzyskał obraz jednolitego systemu i przezroczysty dostęp do zasobów udostępnianych w obrębie klastra, w tym także do istniejących obecnie i rozbudowywanych w przyszłości systemów przechowywania danych. Niezwykle istotnym elementem projektu było opracowanie koncepcji i implementacja Wirtualnego Laboratorium Spektrometrii Magnetycznego Rezonansu Jądrowego, które potraktowane zostało jako rozwiązanie pilotowe i wzorcowe dla laboratoriów wirtualnych z innych dziedzin. Platformą sprzętową proponowanej instalacji pilotowej, niezbędnej do wykonania prac badawczo-rozwojowych, był klaster komputerów SGI z procesorami MIPS oraz Intel, przeznaczony do intensywnych obliczeń oraz zaawansowanej wizualizacji.

Problemem szczególnej wagi zarówno ze względów badawczo-rozwojowych, jak również ekonomiczno-społecznych jest zapasowe centrum obliczeniowe. Specyfika niektórych zadań wymusza zakończenie obliczeń w ściśle określonym czasie. Problemy, które mogą pojawić się w trakcie obliczeń, to wydłużenie czasu wykonania zadania lub awaria systemów obliczeniowych przed lub w trakcie obliczeń. Celem prac było utworzenie systemu pozwalającego na zachowanie głównego kryterium, jakim jest czas zakończenia zadania. Zapasowe centrum obliczeniowe wykorzystane zostało w fazie pilotowej do zwiększenia niezawodności obliczeń przez Instytut Meteorologii i Gospodarki Wodnej w Warszawie (ryc. 10).

Projekt SGGrid realizowany był przez ACK Cyfronet AGH w Krakowie i PCSS we współpracy z TASK (Trójmiejska Akademicka Sieć Komputerowa), WCSS (Wrocławskie Centrum Sieciowo-Superkomputerowe), Centrum Komputerowym Politechniki Łódzkiej, IMGW oraz firmami ATM SA i Silicon Graphics.

Wiele prac zrealizowanych w projekcie doczekało się wdrożenia w kraju i na świecie. Przykładem jest system punktów kontrolnych dla procesorów o architekturze IA-64

(Itanium), testowany przez ośrodek obliczeniowy NASA AMES (USA), wdrożony w kilku lokalizacjach przez Silicon Graphics (Japonia) oraz wykorzystany do realizacji prac w kolejnych projektach międzynarodowych 6. Programu Ramowego UE, np. CoreGRID.



Ryc. 10. Przykład dwóch instancji zapasowego centrum obliczeń dla różnych numerycznych modeli klimatycznych (kopia danych #1 – COSMO, kopia danych #2 – ALADIN)

Projekt celowy Clusterix (Krajowy Klaster Linuksowy) uruchomiony został przy znaczącym wsparciu firmy Intel. Celem projektu było opracowanie narzędzi oraz mechanizmów umożliwiających wdrożenie produkcyjnego środowiska typu Grid utworzonego z lokalnych klastrów PC, zlokalizowanych w wielu niezależnych ośrodkach, geograficznie odległych [42]. Instalacja szkieletowa zbudowana została z lokalnych klastrów PC o architekturze 64-bitowej, połączonych dedykowanymi kanałami na bazie krajowej sieci optycznej PIONIER. Cele szczegółowe obejmowały między innymi przygotowanie systemu umożliwiającego zarządzanie i udostępnianie zasobów, obsługę dynamicznie zmieniającej się wielkości i konfiguracji infrastruktury sprzętowej (z możliwością czasowego dołączania innych klastrów PC i tym samym skalowania infrastruktury), jak również wykorzystania usług oferowanych przez inne projekty powstałe w ramach prac skojarzonych z programem PIONIER, m.in. PROGRESS oraz SGIgrid.

Jednym z głównych założeń było wykorzystanie oraz udostępnienie (po zakończeniu prac badawczych) tylko i wyłącznie oprogramowania typu Open Source, co jest warunkiem koniecznym rozpowszechnienia idei budowy dużego, rozproszonego systemu w skali kraju. Zasadniczy nacisk w projekcie został położony na wykorzystanie rodziny protokołów IP wersji 6 oraz dodatkowej funkcjonalności dostarczanej w ramach tej wer-

sji, np. zwiększonej niezawodności, gwarancji jakości usługi oraz większej możliwości określania wymaganego pasma.

Wśród wartości dodanych, będących wynikiem projektu, wymienić należy:

- utworzenie oprogramowania zdolnego do zarządzania strukturą klastra o dynamicznie zmieniającej się konfiguracji (zmienna liczba węzłów, użytkowników, udostępnianych usług),
- nowa jakość usługi i aplikacji budowanych w oparciu o rodzinę protokołów IPv6,
- integracja i wykorzystanie dostępnych już usług, dostarczanych w ramach innych projektów badawczych bądź celowych (np. składowanie danych, zdalna wizualizacja, moc obliczeniowa dla aplikacji o dużym poziomie skalowalności),
- uwzględnienie lokalnych polityk administracji i zarządzania infrastrukturą w ramach niezależnych domen,
- zintegrowany system interfejsu użytkownika końcowego oraz administratora,
- udostępnienie produkcyjnej infrastruktury Grid,
- zapewnienie odpowiedniego poziomu bezpieczeństwa w heterogenicznym środowisku rozproszonym.

Rozliczanie wykorzystanych zasobów i zdefiniowanie polityki dostępu do całego klastra stanowiły integralny element prac badawczo-rozwojowych projektu Clusterix. W ramach projektu zbudowana została instalacja pilotowa złożona z 12 klastrów lokalnych, połączonych dedykowanymi kanałami opartymi na sieci PIONIER. Rdzeń instalacji pilotowej został wyposażony w węzły PC o architekturze 64-bitowej (typu Intel IA-64) zarządzanych systemem operacyjnym Linux.

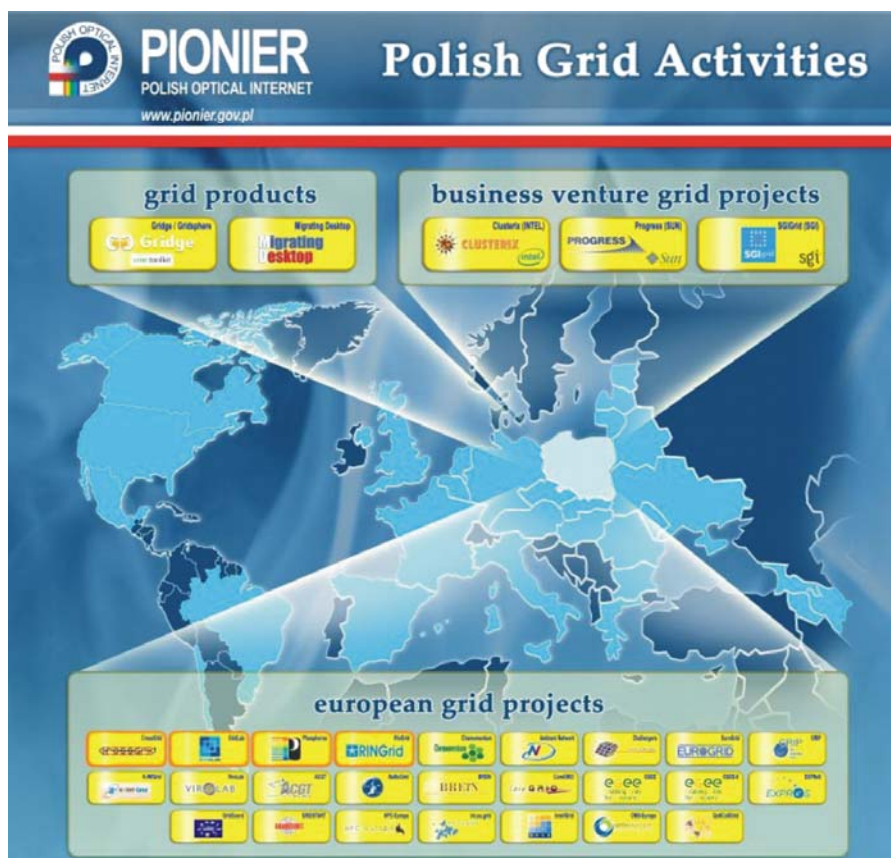
Infrastruktura Clusterix wykorzystywana była zarówno do przetwarzania w technologii HTC (ang. *High Throughput Computing*), jak i uruchamiania aplikacji rozproszonych dużej skali, wymagających równoległego wykorzystania zasobów jednego lub więcej klastrów lokalnych. Wśród aplikacji uruchomionych w środowisku rozproszonym wymienić należy:

- oprogramowanie do modelowania zjawisk termomechanicznych w krzepnących odlewach, wykorzystujące metodę elementów skończonych (MES),
- oprogramowanie do symulacji przepływu transonicznego oraz projektowania zaawansowanych końcówek skrzydeł,
- aplikacja symulacji wielkiej skali przepływu krwi w mikrokapilarach, za pomocą heterogenicznego modelu cząstek (wykorzystującego dysypatywną metodę cząstek wraz z modelem siatkowej reprezentacji krwinek i innych komponentów),
- oprogramowanie do wizualizacji złożonych struktur, wykorzystujące metody rozpoznawania obrazów,
- aplikacje służące przewidywaniu struktury białek z sekwencji aminokwasowej oraz symulacje procesu zwijania białek,

- oprogramowanie umożliwiające badanie właściwości molekularnych układów biologicznych o znaczeniu chemoterapeutycznym,
- oprogramowanie do projektowania układów elektroniki molekularnej oraz symulacji mechanosyntezy,
- pakiet GAMESS (ang. *The General Atomic and Molecular Electronic Structure System*).

2.2.4. Udział w europejskich projektach gridowych

Polskie zespoły uczestniczyły także w bardzo wielu europejskich projektach gridowych w 5. i 6. Programie Ramowym (ryc. 11).



Ryc. 11. Udział polskich zespołów w projektach gridowych

Na rysunku tym czerwoną obwódką objęte są projekty koordynowane przez polskie zespoły naukowe i opisane wcześniej. Ponadto rysunek zawiera także dwa produkty (Gridge [43], Migrating Desktop [44]) powstałe w ramach tych projektów, które były szeroko używane w innych projektach europejskich i światowych.

Podsumowując krótką dyskusję na temat e-infrastruktury, należy zauważyć, że powstała w Europie i w Polsce infrastruktura gridowa, w powiązaniu z infrastrukturą sieciową, dała podstawę do opracowania wielu aplikacji w różnych obszarach, np. radioastronomii, fizyce wysokich energii, medycynie.

2.3. Infrastruktura warstwy danych naukowych

Ogólna idea powiązania z e-infrastrukturą dodatkowej warstwy, tzw. warstwy danych naukowych, jest związana z potrzebą udostępnienia metod i narzędzi wspomagania całości procesów badawczych. Wynika z tego potrzeba zintegrowania z e-infrastrukturą takich systemów, jak np. wirtualne laboratoria, biblioteki cyfrowe, repozytoria, bazy danych. To otoczenie badań naukowych powinno być dostępne dla użytkowników w podobny sposób, jak usługi dostępu do sieci i gridów. Prace dotyczące aplikacji wspomagających prace naukowe prowadzone były już w 6. Programie Ramowym. Również w Polsce, w ramach programu PIONIER i projektu celowego SGIGrid, powstały odpowiednio: biblioteka cyfrowa dLibra i wirtualne laboratorium. Warstwa danych naukowych w 7. Programie Ramowym nie została jeszcze precyzyjnie zdefiniowana. Program rozpoczął się w roku 2007, a trwać będzie do roku 2013, a więc w kolejnych jego fazach podstawowe tezy będą uszczegóławiane. Dlatego w dalszej części tego punktu omówimy krótko dwa pilotowe rozwiązania i wdrożenia: biblioteki cyfrowej i wirtualnego laboratorium.

2.3.1 Biblioteka cyfrowa

Biblioteka cyfrowa dLibra powstała w ramach własnych prac badawczo-rozwojowych prowadzonych przez PCSS od 1999 roku i od momentu pierwszego wdrożenia w ramach Wielkopolskiej Biblioteki Cyfrowej (rok 2002) stanowi główny składnik warstwy danych naukowych w sieci PIONIER.

Oprogramowanie dLibra służy do budowy profesjonalnych repozytoriów obiektów cyfrowych. Repozytoria oparte o oprogramowanie dLibra mogą pełnić rolę systemów takich, jak: biblioteki cyfrowe, instytucjonalne repozytoria dokumentów, cyfrowe archiwa itp.

System dLibra pozwala na przechowywanie obiektów cyfrowych w dowolnym formacie (np.: wieloplikowe strony HTML, dokumenty PDF, nagrania audio i wideo itd.), a każdy z przechowywanych obiektów opisany może być za pomocą zdefiniowanego w ramach danego repozytorium zestawu metadanych. Przechowywane obiekty mogą być udostępniane czytelnikom w Internecie poprzez dedykowane strony WWW repozytorium. Repozytoria dLibra mogą także współpracować z innymi systemami internetowymi przy użyciu otwartych standardów komunikacyjnych, takich jak protokół OAI-PMH czy kanały RSS. Jednym z najważniejszych zadań oprogramowania dLibra jest przechowywanie i udostępnianie obiektów cyfrowych. W celu realizacji tego zadania oprogramowanie dLibra zostało wyposażone m.in. w następujące cechy:

- Hierarchiczna struktura repozytorium – struktura katalogów i obiektów tworzona przez redaktorów, i tylko dla nich dostępna. Jest to struktura, w której umieszczają oni obiekty cyfrowe.
- Kolekcje – są zbiorami obiektów, które są widoczne dla użytkowników stron WWW repozytorium. Wprowadzają one z reguły podział tematyczny zasobów biblioteki. Struktura kolekcji jest również hierarchiczna – kolekcje mogą być rozdzielane na podkolekcje, np. o węższym zakresie tematycznym. Każdy obiekt cyfrowy umieszczony w repozytorium może być przypisany do jednej lub wielu kolekcji.
- Grupowanie obiektów cyfrowych – oprogramowanie dLibra pozwala na tworzenie grup obiektów cyfrowych, które mają pewne wspólne cechy. Grupy ułatwiają zarówno zarządzanie znajdującymi się w niej obiektami (np. umożliwiają modyfikację pewnych wspólnych właściwości wszystkich obiektów w grupie, takich jak przypisanie do kolekcji czy elementy opisu), jak i prezentację powiązanych ze sobą obiektów.
- Planowane obiekty cyfrowe – w oprogramowaniu dLibra można tworzyć planowane obiekty cyfrowe. Obiekty te pozwalają informować użytkowników o planach umieszczenia obiektów w repozytorium. Obiekty planowane posiadają tylko metadane (nie posiadają treści).
- Wersjonowanie treści – oprogramowanie dLibra umożliwia aktualizowanie treści obiektów cyfrowych przy równoczesnym zachowaniu ich wcześniejszych wersji. W rezultacie w systemie dLibra znajdują się wszystkie kolejne wersje obiektu cyfrowego.

Obecnie w sieci PIONIER uruchomionych jest 18 publicznie dostępnych bibliotek cyfrowych, których zasoby są dostępne poprzez otwarte protokoły (ryc. 12). Część z tych bibliotek to biblioteki regionalne współtworzone zazwyczaj przez instytucje naukowe i kulturalne danego regionu. Największe z nich to Wielkopolska Biblioteka Cyfrowa i Kujawsko-Pomorska Biblioteka Cyfrowa. Pozostałe to biblioteki cyfrowe o charakterze instytucjonalnym, w których zasoby udostępniane są przez jedną instytucję, jednakże wspomagają prace naukowe prowadzone w różnych ośrodkach w Polsce. Do największych w tej grupie należą Cyfrowa Biblioteka Narodowa „Polona” oraz Wirtualny Księgozbiór Federacji FIDES. Wszystkie biblioteki cyfrowe udostępniają ponad 130 000 cyfrowych publikacji.

Dzięki usłudze wyszukiwania rozproszonego cały zasób zebrany w ramach krajowej sieci bibliotek cyfrowych opartej na sieci PIONIER dostępny jest dla każdego z użytkowników i dla każdej z bibliotek. W 2006 r. udostępniono bowiem mechanizm wymiany metadanych pomiędzy rozproszonymi bibliotekami cyfrowymi i opartych na tym rozwiązaniu zaawansowanych funkcji, takich jak wyszukiwanie rozproszone czy dynamiczne wirtualne kolekcje [45]. Kolejnym krokiem było udostępnienie w czerwcu 2007 r. serwisu Federacji Bibliotek Cyfrowych. Serwis ten, opierając się na otwartych standardach, umożliwia przeszukiwanie zasobów rozproszonych bibliotek cyfrowych, tworzenie trwa-

łych referencji do poszczególnych obiektów cyfrowych oraz dostęp do raportów zawierających plany digitalizacji poszczególnych bibliotek. Dzięki FBC możliwe jest również automatyczne wykrywanie duplikatów między bibliotekami cyfrowymi oraz zapobieganie dublowaniu digitalizacji już na etapie jej planowania w rozproszonych ośrodkach. Najnowsze funkcje FBC są związane z sieciowym profilem użytkownika, pozwalającym użytkownikom polskich bibliotek cyfrowych korzystać z jednej tożsamości we wszystkich bibliotekach Federacji.



Ryc. 12. Instalacje bibliotek cyfrowych w sieci PIONIER

Dostępne w tej chwili zasoby platformy bibliotek cyfrowych są środowiskiem dla pracy badawczej humanistów (w szczególności historyków i literaturoznawców), jak również są niezbędną infrastrukturą do prowadzenia prac badawczo-rozwojowych w takich przekrojowych obszarach, jak: sieci semantyczne i zaawansowana analiza treści, rozproszone systemy usług informacyjnych, architektura zorientowana na zasoby.

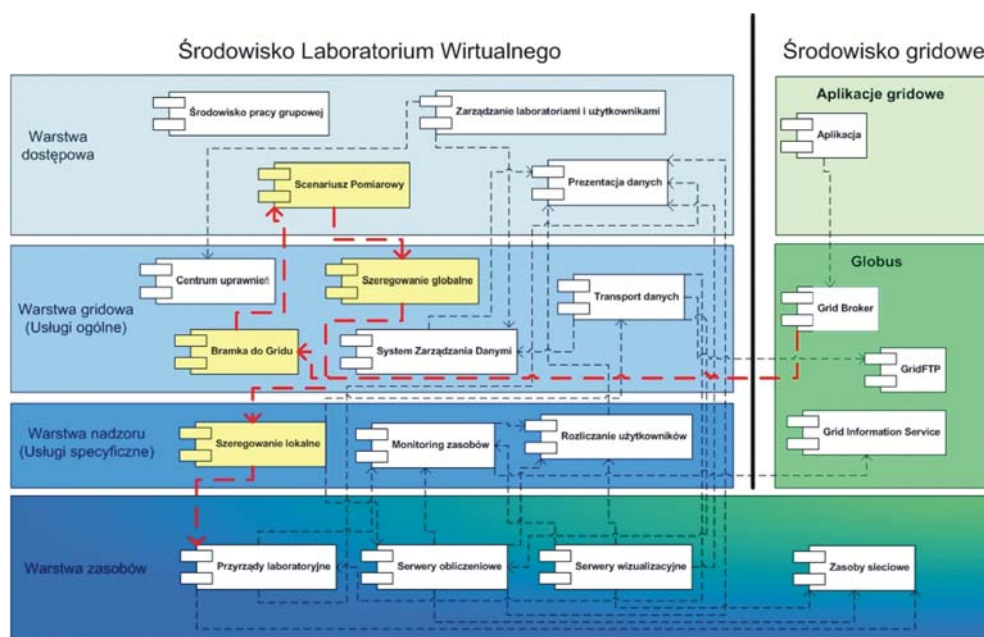
W tym celu wymagany jest również rozwój samej biblioteki cyfrowej w kierunku opracowania mechanizmów usług atomowych [46]. Opracowano następujące usługi: usługę przechowywania obiektów cyfrowych, usługę przechowywania metadanych, usługę kompozycji obiektów cyfrowych i tworzenia powiązań między nimi oraz usługę wyszukiwania obiektów cyfrowych. Kolejny etap prac polegał na zdefiniowaniu protokołów komunikacyjnych dla poszczególnych usług atomowych oraz mechanizmu odkrywania, kompozycji i monitorowania tych usług. Cele te zostały zrealizowane w architekturze zorientowanej na zasoby (ang. *Resource Oriented Architecture*, ROA) z wykorzystaniem podejścia REST (ang. *Representational State Transfer*) oraz ontologii OWL-S. W efekcie uzyskano prototypowy system atomowych usług dla rozproszonych bibliotek cyfrowych.

Dodatковым efektem powstania warstwy danych naukowych, realizowanej funkcjonalnie w środowisku biblioteki cyfrowej, jest możliwość wykorzystania tych narzędzi do budowy kolejnych aplikacji naukowych. Opracowanie i uruchomienie w sieci PIONIER zbioru usług atomowych dla zaawansowanego zarządzania treścią cyfrową i dostępem do niej wraz z usługami opartymi na metadanych pozwoliło na budowę zaawansowanych aplikacji. Poza wspomnianą powyżej Federacją Bibliotek Cyfrowych dla nauk humanistycznych, powstały również aplikacje: Medyczna Biblioteka Cyfrowa dla Telekonsultacji czy Biblioteka Naukowa dla Wirtualnego Laboratorium [47].

2.3.2 Wirtualne Laboratorium

Wirtualne laboratorium powstało w wyniku prac realizowanych w dwóch projektach: celowym SGIgrid i badawczym pt.: Budowa uniwersalnej architektury dla Laboratorium Wirtualnego. Laboratorium Wirtualne (VLab) jest systemem umożliwiającym zdalny dostęp do przyrządów naukowych. Przyrządy te ze względu na swoją wysoką cenę są unikatowe i na ich zakup stać tylko największe ośrodki badawcze. Przykładami takich urządzeń są spektrometry rezonansu magnetycznego (NMR) i radioteleskopy. W czasie prac badawczych został wypracowany model uniwersalnej architektury Laboratorium Wirtualnego, który uwzględnia heterogeniczność przyrządów, pochodzących z różnych dziedzin nauki i tym samym posiadających swoją specyfikę działania, zróżnicowany dostęp oraz poziom automatyzacji realizacji zlecenia [48]. Proponowany system Laboratorium Wirtualnego ma budowę modułową, która umożliwia szybkie dostosowanie go do nowych zastosowań oraz w łatwy sposób pozwala rozszerzać możliwości systemu (poprzez dodanie nowych modułów), zmieniać już istniejące moduły na inne, o większych możliwościach. Aby to umożliwić, konieczne jest zdefiniowanie ujednoliconego interfejsu dostępu i sposobu komunikacji między poszczególnymi modułami. Różny sposób sterowania przyrządami laboratoryjnymi pociąga za sobą konieczność budowy systemu modułowego z rozróżnieniem na moduły uniwersalne oraz specyficzne. Moduły uniwersalne możemy stosować w każdym typie laboratorium, ich działanie jest niezależne od obsługiwanych przyrządów laboratoryjnych. Moduły specyficzne, dostosowane do przy-

rządów stosowanych w laboratorium, „znają” ich specyfikę i muszą być implementowane w znacznej części od nowa podczas adaptacji nowego profilu laboratorium. W architekturze wyróżniono cztery warstwy (ryc. 13).



Ryc. 13. Architektura Laboratorium Wirtualnego (źródło [49])

Warstwa dostępową zawiera:

- dynamiczne scenariusze pomiarowe – zapewniające realizację i sterowanie ciągiem operacji (eksperymentów) zleconych przez użytkownika,
- zarządzanie laboratoriami i użytkownikami – narzędzia do tworzenia nowych profili laboratoriów, nowych użytkowników, zarządzania prawami dostępu do przyrządów, bibliotek cyfrowych i mediów komunikacyjnych,
- prezentację danych – prezentacja wyników eksperymentów zarówno tych zgromadzonych w systemie zarządzania danymi, jak i tych zmieniających się, wyświetlanych na bieżąco, wyświetlanie danych rozliczeniowych,
- mechanizmy komunikacji pomiędzy użytkownikami – korzystanie z narzędzi do porozumiewania się.

Warstwa gridowa (usługi ogólne) zawiera:

- Centrum Certyfikacji – generowanie certyfikatów i sprawdzanie ich ważności, przechowywanie i kontrola praw dostępu (integralną częścią są narzędzia zapewniające autoryzację i uwierzytelnianie użytkowników podczas logowania),

- szeregowanie globalne – wybór przyrządu laboratoryjnego lub serwera obliczeniowego do uruchomienia zleconego zadania,
- system zarządzania danymi – przechowywanie wyników eksperymentów i obliczeń, publikacji elektronicznych,
- transport danych – przesłanie danych na maszynę docelową, kontrola transmisji,
- bramkę do Gridu – komunikacja z brokerem gridowym, odbiór zadań z Gridu (spoza systemu VLab) i przesłanie zadania do wykonania na przyrządzie.

Warstwa nadzoru definiuje specyficzne usługi zależne od wybranych urządzeń laboratoryjnych, stanowiąc tym samym interfejs między warstwami ogólnego zastosowania a infrastrukturą specyficzną dla danego eksperymentu. Warstwa ta obejmuje między innymi system szeregowania zadań w urządzeniu, monitorowanie zasobów oraz rozliczanie użytkowników i wykorzystanych zasobów.

Warstwa zasobów obejmuje przyrządy laboratoryjne i ich oprogramowanie sterujące, systemy obliczeniowe, wizualizacji oraz danych.

W efekcie powstał system umożliwiający szybką adaptację do obsługi urządzeń o odmiennej charakterystyce oraz zapewnienia często zróżnicowanych wymagań użytkowników, począwszy od prostych urządzeń sensorowych (czujniki temperatury) do skomplikowanych urządzeń (radioteleskopy). Dla tego zastosowania VLab pełni rolę pomostu pomiędzy tradycyjną nauką eksperymentalną a nową dziedziną, która w pełni wykorzystuje infrastrukturę europejską w postaci szybkich sieci dostępowych o zdefiniowanej jakości usługi (SLA – *Service Level Agreement*, QoS) – GÉANT2, rozproszonych zasobów obliczeniowych w postaci gridów (EGEE2/EGEE3) i instalacji HPC (PRACE) z usługami dodanymi, jak interaktywność w środowisku rozproszonym (np. projekty europejskie: int.eu.grid, BalticGrid).

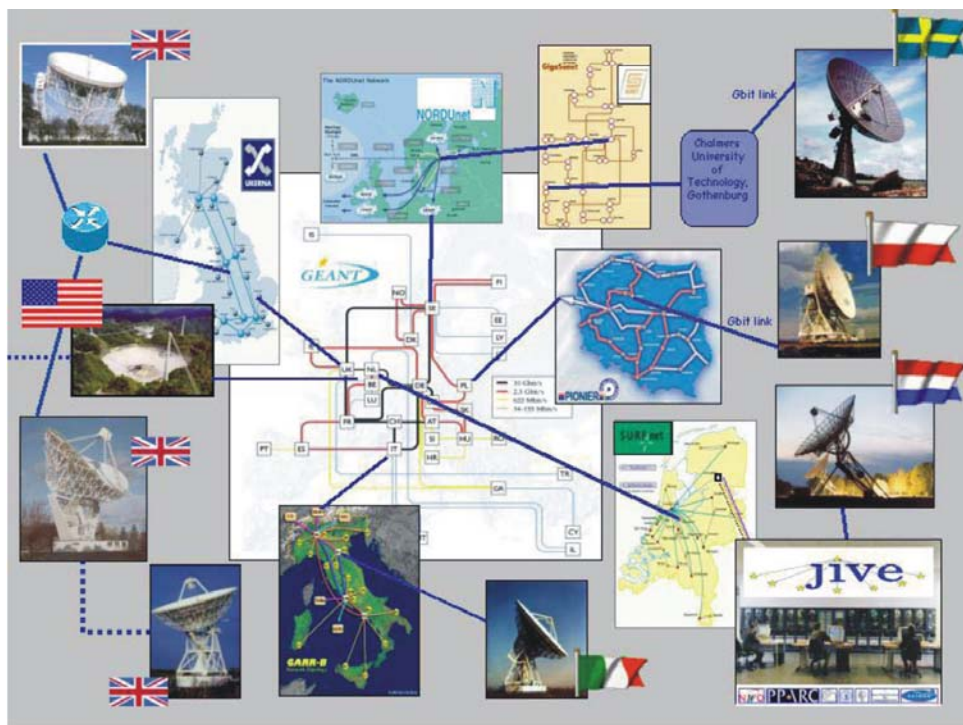
Kontynuacja prac i szerokie zastosowanie Laboratorium Wirtualnego na poziomie europejskim realizowane jest w projektach RINGRID [50], EXPReS [51] oraz DORII [52].

3. Integracja e-infrastruktury z różnymi obszarami badań naukowych

3.1. Radioastronomia

Tradycyjna radioastronomia wykorzystywała radioteleskopy pracujące oddzielnie i oddalone od siebie o tysiące kilometrów, które obserwowały ten sam fragment nieba (technika VLBI). Zebrane dane były próbkowane cyfrowo i przechowywane lokalnie na taśmach magnetycznych i dyskach. Taśmy te były następnie przesyłane do centralnego punktu (JIVE w Holandii), gdzie były korelowane za pomocą korelatora sprzętowego. Sumaryczny strumień danych trafiający do procesora centralnego to 10-100 TB na

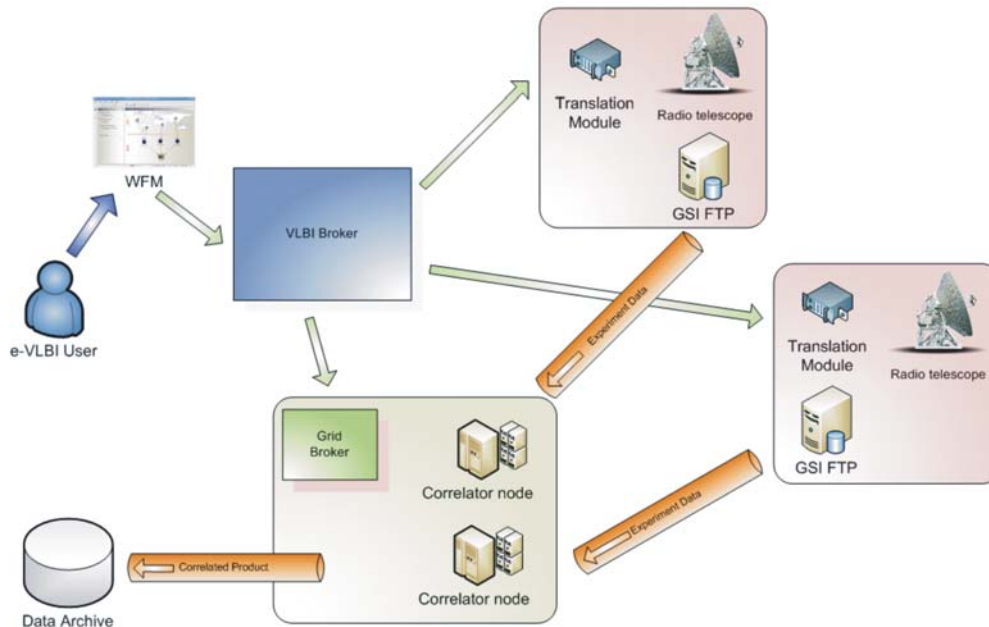
obserwowany sygnał. Po wstępnym przetworzeniu dane są redukowane do 10-100 GB. W efekcie uzyskiwane są bardzo wysokiej rozdzielczości obrazy continuum i linii spektralnych kosmicznych źródeł radiowych. Cykl zajmował ok. 6 miesięcy. Próba korelacji w czasie rzeczywistym została podjęta w ramach programu EXPReS. Uzyskano produkcyjną usługę e-VLBI, umożliwiającą niezawodne podłączenie radioteleskopów do komputera centralnego za pośrednictwem szybkich łączy optycznych (1 Gb/s). Konfigurację pilotową przedstawiono na ryc. 14.



Ryc. 14. Konfiguracja pilotowa e-VLBI (źródło [53])

W niedalekiej przyszłości wymagania przepustowości dla jednego radioteleskopu osiągną 10 Gb/s. Warto zauważyć, że jakościowa zmiana badań została uzyskana przez zastosowanie e-infrastruktury na poziomie sieciowym.

W projekcie EXPReS podjęto próbę wykorzystania także infrastruktury gridowej do zbudowania rozproszonego korelatora, pozbawionego wad korelatora sprzętowego. System obejmuje sieć serwerów plików skojarzonych z radioteleskopami. Zbierane dane, dzielone na paczki, są wysyłane do węzłów rozproszonego korelatora i przetwarzane w czasie rzeczywistym w środowisku gridowym. Architektura systemu wraz z przepływem danych przedstawiona jest na ryc. 15.



Ryc. 15. Architektura rozproszonego korelatora (źródło [54])

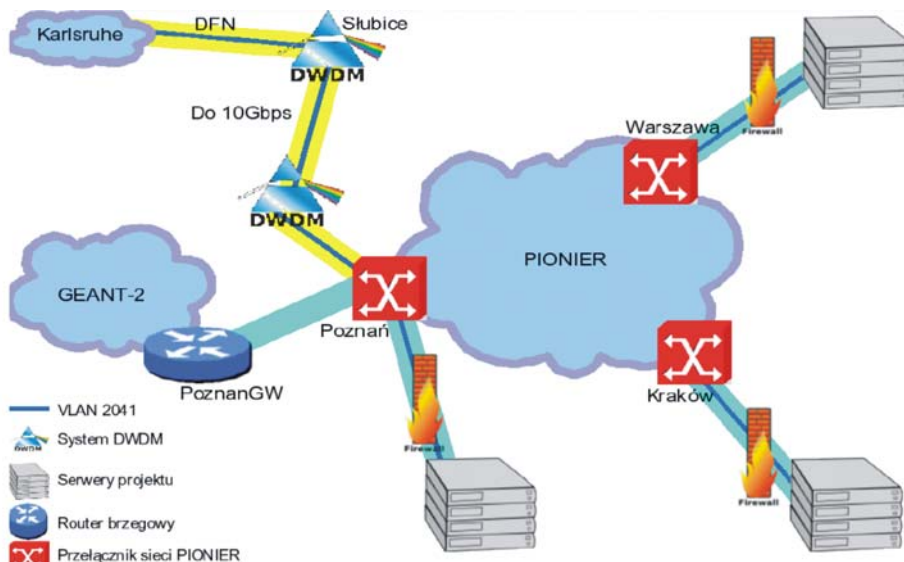
3.2. Fizyka wysokich energii

„Światowa społeczność fizyków niecierpliwie czeka na uruchomienie najpotężniejszego w historii urządzenia do badania cząstek elementarnych” napisał Graham P. Collins [55].

Tym urządzeniem jest Wielki Zderzacz Hadronów (LHC – ang. *Large Hadron Collider*), którego budowa w CERN zakończy się w tym roku. Rozpędzone przeciwbieżne paczki protonów będą zderzały się, produkując nowe cząstki, które muszą zostać wykryte przez detektory ATLAS, ALICE, CMS, LHCb, rejestrujące te zdarzenia. Dane z detektora produkowane są z szybkością 0,5 PB/s i następnie redukowane zgodnie z predefiniowanymi kryteriami. Do przechowywania i przetwarzania tych ogromnych ilości danych stosuje się wielopoziomą infrastrukturę gridową. Poziom „0” zlokalizowany jest w CERN, natomiast poziom 1 w 11 dużych ośrodkach obliczeniowych na świecie. Do wielu dołączone są ośrodki regionalne poziomu 2. W Polsce są trzy takie ośrodki: w Krakowie (Cyfronet), Poznaniu (PCSS) i w Warszawie (ICM). Są one podłączone dedykowanymi kanałami 1 Gb/s przez sieci naukowe PIONIER i DFN do centrum Tier1 w Karlsruhe (ryc. 16).

Polski udział w przygotowaniu LHC jest znacznie szerszy i został opisany w artykule J. P. Nassalskiego [56]. Współpraca fizyków z IFJ PAN w Krakowie i IPJ z Uniwersytetu Warszawskiego z informatykami z tych trzech ośrodków skutkowało udziałem i koordy-

nacją w dużym gridowym projekcie CrossGrid (patrz rozdz. 2.2), a następnie w trzech kolejnych projektach EGEE (*The Enabling Grids for E-science*).



Ryc. 16. Grid fizyki wysokich energii

Infrastruktura sieciowa i gridowa umożliwiła w tych projektach zbudowanie produkcyjnego systemu przechowywania danych i obliczeń w środowisku rozproszonych w świecie klastrów. Bez takiej sieci obliczeniowej nie byłoby możliwe prowadzenie badań naukowych związanych z LHC. Przyszłościowe wymagania na przepustowość sieci dla tego zastosowania definiowane są w perspektywie do 2015 r. w zakresie od 100 Gb/s do 1 Tb/s [57].

3.3. Medycyna

Medycyna jest jedną z dziedzin, w których e-infrastruktura ma do spełnienia ważną rolę. Wyniki badań specjalistycznych z urządzeń diagnostycznych są bowiem przechowywane i dostępne w wersjach cyfrowych. W ogólności mają one charakter multimedialny i wymagają dużej przepustowości sieci.

Przykładem może być eksperyment telepatomorfologiczny, przeprowadzony podczas konferencji w 2004 r. [58]. W tym celu zestawiono dedykowane połączenie gigabitowe o przepustowości 1 Gb/s ze szpitalem w Łodzi (ryc. 17). Warto w tym miejscu zauważyć, że w sieci PIONIER, za pośrednictwem MANów, szerokopasmowo łączonych jest 86 szpitali, w tym wszystkie szpitale kliniczne. Część przyszłych aplikacji i usług telemedycznych jest silnie uwarunkowana możliwością podłączenia szpitali do sieci szerokopasmowych.

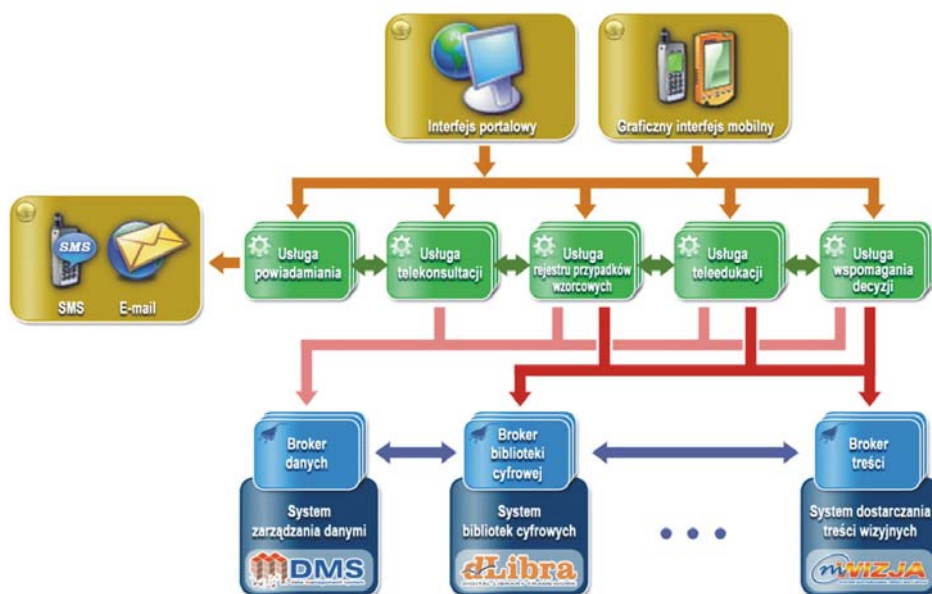
Przykładem bardzo zaawansowanego systemu wymagającego takich połączeń jest portal telemedyczny, uruchomiony w wersji pilotowej pod nazwą Telemedycyna Wielkopolska [59-61].



Ryc. 17. Dedykowane połączenie Poznań-Łódź dla eksperymentu telepatomorfologicznego

Główną funkcją tego portalu w jego obecnej wersji są telekonsultacje z obszaru chirurgii urazowej. Scenariusz telekonsultacji obejmuje pełen cykl procedury konsultacyjnej wspartej multimedialnymi danymi pacjenta i umożliwia realizację telekonsultacji w trybie ciągłym. Telekonsultacje są wspierane systemem klasyfikacji przypadków urazowych oraz rejestrem przypadków wzorcowych. Portal integruje szereg narzędzi informatycznych, w szczególności opisaną wcześniej bibliotekę cyfrową dLibra. Kolejne planowane moduły będą związane z funkcjami edukacyjnymi (zdalne nauczanie, transmisja pola operacyjnego), wspomaganie decyzji diagnostycznych systemami eksperckimi. Architektura systemu przedstawiona jest na ryc. 18. Wielkim wyzwaniem dla medycyny jest tworzenie modeli medycznych reprezentujących choroby, czyli biotypów. Mogą one być wykorzystane do symulacji związków pomiędzy typami organizmów pacjentów, czynnikami chorób i możliwym procesem leczenia. Próby badań podejmowane w tym obszarze wymagają dostępności do bardzo dużych zasobów przetwarzania i przecho-

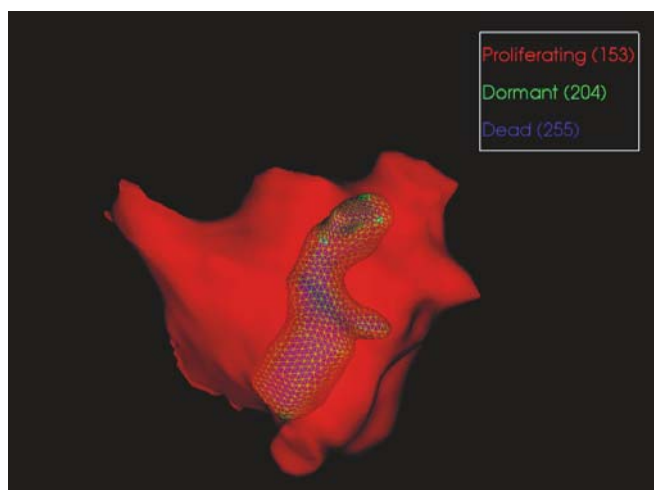
wywnania danych. Z tego powodu postęp w tych badaniach można uzyskać przez użycie infrastruktury gridowej.



Ryc. 18. Architektura systemu Telemedycyna Wielkopolska

Przykładem zastosowania architektury gridowej może być projekt ACGT (*Advancing Clinico Genomic Trial on Cancer*) [62], skupiający się na badaniach, których celem jest dostarczenie nowej broni do walki z rakiem. Realizacja tego zadania polega na stworzeniu narzędzi umożliwiających dokładniejszą identyfikację choroby oraz wspomaganie lekarza w procesie leczenia pacjenta. W projekcie zamierza się wykorzystać możliwości, jakie daje zaawansowane środowisko gridowe do analizy i porównywania danych klinicznych i genetycznych z rezultatami badań zgromadzonymi w bazach danych. Ma to na celu szybkie diagnozowanie, a także definiowanie dokładnej terapii dla każdego pacjenta. Projekt ACGT rozpoczął się w lutym 2006 r. i ma się zakończyć w styczniu 2010 r., a bierze w nim udział 25 ośrodków naukowych z Europy oraz Japonii. PCSS odpowiedzialny jest za dostarczenie podstawowej platformy gridowej dla projektu. Zastosowanie technologii gridowej w projekcie ma istotne znaczenie. Z jednej strony dostarcza ona uniwersalnych mechanizmów współdzielenia rozproszonych zasobów przez różnorodne grupy naukowców z Europy i świata, przez tworzenie wirtualnych organizacji. Z drugiej strony zwiększa możliwości wykorzystania aplikacji intensywnych obliczeniowo, co powoduje znaczny wzrost dokładności obliczeń lub istotne zmniejszenie czasu oczekiwania na wyniki. Pierwszy wspomniany powyżej aspekt, implementowany

jest w projekcie ACGT w postaci zestawu usług wysokiego poziomu zorientowanego na specyficzne problemy biomedyczne. Chodzi tu głównie o dostarczenie środowiska dla naukowców i lekarzy, umożliwiającego skorzystanie w prosty sposób z istniejących zasobów bazodanowych oraz prostych mechanizmów przetwarzania tych danych w postaci przepływu akcji (*workflow*), w którym dane wyjściowe jednego algorytmu są wejściem dla innego. Warstwa ta osadzona jest na uniwersalnej platformie gridowej, zbudowanej w oparciu o szeroko stosowane technologie gridowe Globus Toolkit, wspomagane przez zaawansowane mechanizmy wyższego poziomu, odpowiedzialne za zarządzanie zasobami obliczeniowymi, zarządzanie danymi, czy też zapewniające uniwersalne mechanizmy autoryzacji (Grid Toolkit).



Ryc. 19. Wizualizacja wyników symulacji rozwoju nowotworu na podstawie danych wyjściowych aplikacji Oncosimulator (źródło [62])

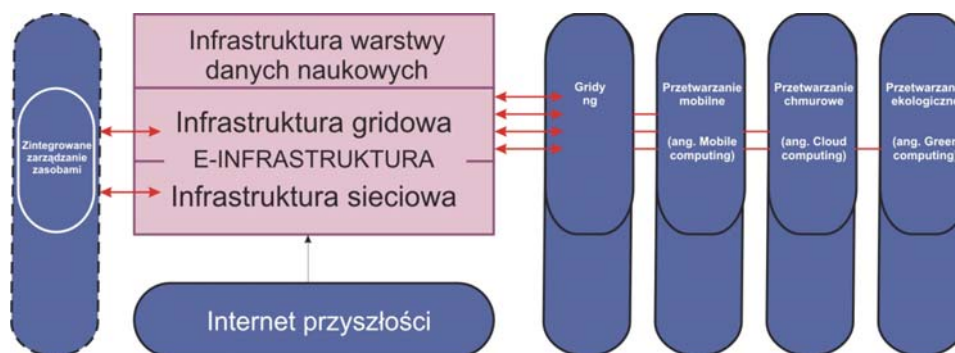
Dla badań obliczeniowych najbardziej istotna w ACGT jest aplikacja służąca do symulacji rozwoju guza rakowego (Oncosimulator) (ryc. 19), w zależności od zastosowanej formy leczenia – radioterapia, chemioterapia. Ponadto istotną rolę odgrywa tu profil genetyczny pacjenta i próby porównywania tego profilu z podobnymi przypadkami opisanymi w bazach danych. Znalezienie zależności między szeroko pojętym profilem biochemicznym chorego a jego odpowiedzią na terapię, to wielkie wyzwanie stojące przed środowiskami onkologicznymi. Tego typu aplikacja może stanowić dla lekarza nieocenione narzędzie do walki z nowotworami. Aplikacje wspomagane przez technologie gridowe nigdy nie zastąpią lekarza, ale mogą stać się istotnym czynnikiem wspomagającym go w procesie podejmowania decyzji w zakresie leczenia pacjenta. Rosnąca rola tego typu aplikacji będzie możliwa dzięki współdzieleniu coraz większej ilości danych biomedycznych o pacjentach (przy zachowaniu wszelkich norm w zakresie ochrony danych

osobowych) oraz dzięki zwiększaniu skali precyzji i wielkości obliczeń, co mogą zapewnić stosowane technologie gridowe.

Projekt ma środowisko pilotowe stworzone w oparciu o maszyny udostępnione przez ośrodki biorące udział w projekcie. Na tych zasobach została zainstalowana podstawowa infrastruktura gridowa w postaci serwisów pakietu Globus Toolkit oraz podstawowych elementów stworzonego w PCSS Gridge Toolkit. Elementami tymi są: system zarządzania zasobami GRMS, system zarządzania danymi GDMS oraz system autoryzacji gridowej GAS służący również jako komponent do definiowania wirtualnych organizacji.

4. Kierunki badań w zakresie e-infrastruktury

Rozwój technologii sieci szerokopasmowych oraz sieci mobilnych stwarza szansę budowania połączeń szerokopasmowych „od końca do końca” i zapewnienia dostępności użytkowników (lub zasobów) w trybie „zawsze i wszędzie”. Skutkuje to nowymi wymaganiami w stosunku do rozproszonych systemów usługowych. Systemy usługowe nowej generacji powinny być ukierunkowane na użytkownika (ang. *user driven*) i powinny posiadać dużą skalowalność (do setek tysięcy użytkowników). Konstrukcja takich systemów wymaga jednak intensywnych badań w różnych obszarach (ryc. 20).



Ryc. 20. Obszary badań w zakresie e-infrastruktury

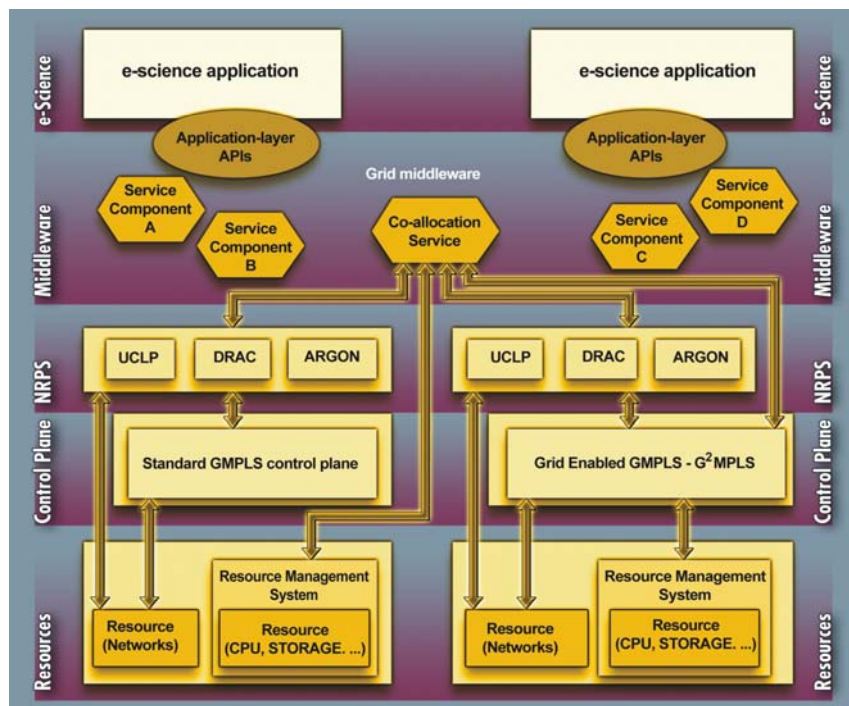
Stawiane są pytania o nową architekturę sieci wobec ograniczeń dzisiejszego Internetu, zapewniającą zwiększone bezpieczeństwo, łatwość zarządzania, realizację nowych modeli ekonomicznych dla Internetu przyszłości. Równocześnie masowe pojawienie się w sieciach w niedalekiej przyszłości takich urządzeń, jak: sensory, systemy wbudowane, systemy elektroniki konsumenckiej, wymagać będzie nowych metod integracji różnych, obszarowo i technologicznie, sieci oraz zarządzania nimi.

Nowych koncepcji można się spodziewać w zakresie gridów. W 7. Programie Ramowym formułuje się zadania dotyczące zwiększenia ich niezawodności i autonomiczności, bezpieczeństwa i skalowalności.

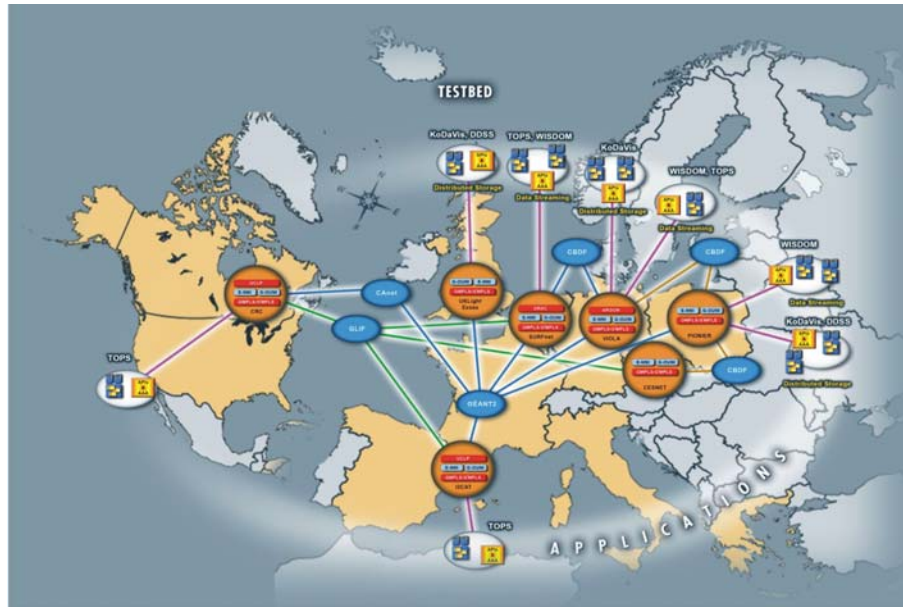
Gridy nowej generacji mają także w znaczącym stopniu wykorzystywać usługi wspomagane wiedzą [63]. Nowe przemysłowo-biznesowe spojrzenie na gridy definiuje potrzebę badań w obszarach tzw. przetwarzania mobilnego (ang. *mobile computing*), „chmury” przetwarzającej (ang. *cloud computing*), obliczeń ekologicznych (tj. ekologicznych pod względem parametrów zasilania i chłodzenia).

Interesującym i ważnym kierunkiem badań są próby zintegrowania zarządzania zasobami e-infrastruktury, tj. zasobami sieciowymi i gridowymi, a w przyszłości także zasobami warstwy danych naukowych.

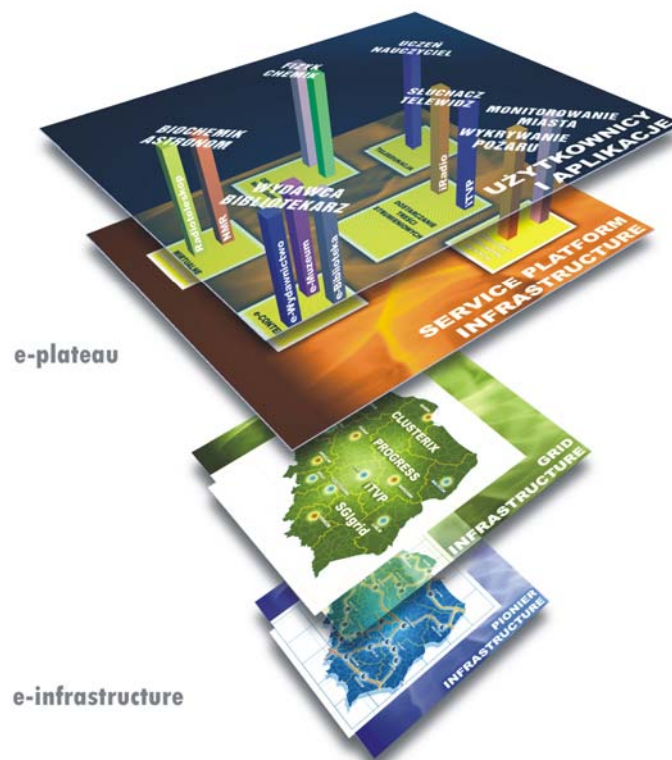
Jednym z projektów europejskich poświęconych integracji sieci optycznych oraz zasobów obliczeniowych jest projekt PHOSPHORUS [64]. Głównym celem tego projektu jest integracja zasobów obliczeniowych i sieciowych tak, aby były one widoczne i rezerwowane przez użytkowników w zunifikowany sposób w jednym systemie gridowym. Oferowane rozwiązania nie tylko umożliwiają realizację takich usług w rozproszonym środowisku w wielu domenach administracyjnych, ale dynamicznie dostosowują się do zmieniających warunków w rozległej infrastrukturze. Rozwiązywane są problemy dostarczenia usług na żądanie oraz alokacji zasobów w wymaganym czasie w środowisku wielu domen, pracujących w różnych technologiach sieciowych. Rozwiązanie takie wymusza ścisłą współpracę naukowców, operatorów telekomunikacyjnych oraz producentów sprzętu.



Ryc. 21. Architektura Phosphorus (źródło [64])



Ryc. 22. Połączenie testowe Phosphorus



Ryc. 23. Infrastruktura warstwy naukowych danych i usług

W ramach osiągniętych wyników uzyskano wysoki stopień integracji pomiędzy aplikacjami, oprogramowaniem gridowym oraz sieciami optycznymi w oparciu o trzy warstwy architektury całego systemu (ryc. 21): warstwę usługową (ang. *service plane*), warstwę systemów NRPS (ang. *Network Resource Provisioning System*) i warstwę sterującą (ang. *control plane*) oraz zaproponowano wspólne mechanizmy uwierzytelniania i autoryzacji we wszystkich wyżej wymienionych warstwach. Dla weryfikacji rezultatów projektu zbudowano środowisko testowe obejmujące obszar Europy oraz połączenia do Kanady i USA. Infrastruktura ta, wykorzystując sieci GÉANT2, GLIF czy połączenia CBDF (*Cross Border Dark Fiber*), łączy lokalne sieci pilotowe partnerów projektu (ryc. 22). Uzyskane wyniki są przedmiotem standaryzacji w Open Grid Forum [65].

W koncepcji PIONIER2 [66] zaproponowano, aby rozproszone systemy obsługi dużej populacji użytkowników przy ogromnej liczbie źródeł informacji nazywać platformami usługowymi. Platformy te oferują usługi dedykowane dla określonej „wirtualnej” społeczności. Tworzą one infrastrukturę warstwy naukowych danych i usług (ryc. 23).

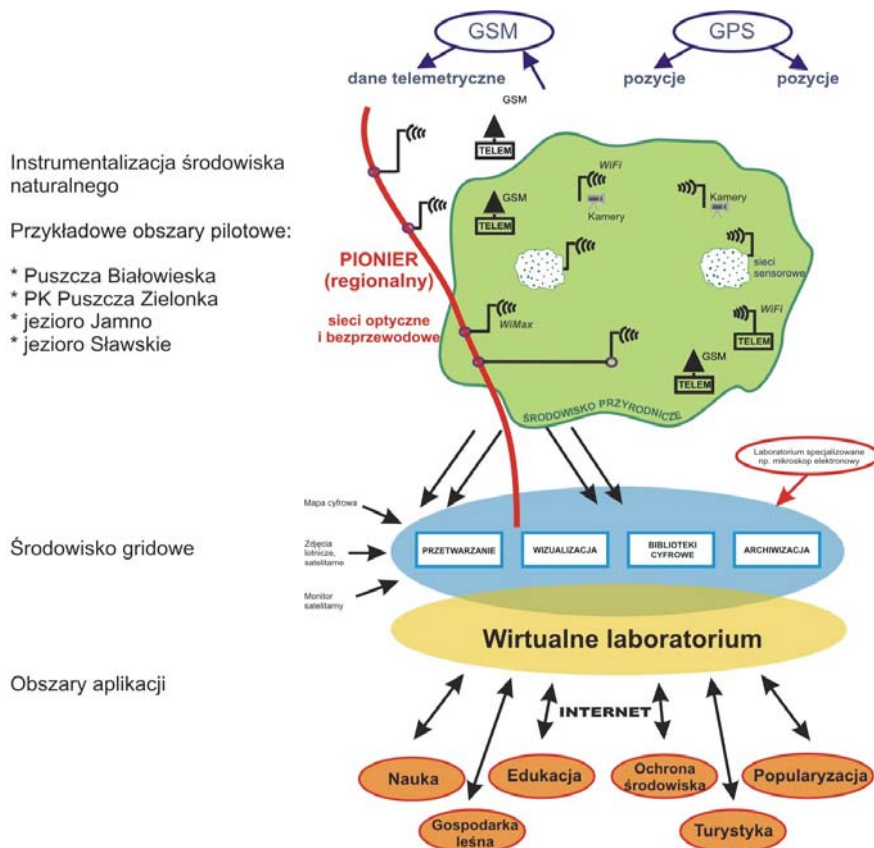
Przy projektowaniu platform zastosowane zostanie podejście „od góry do dołu”, przy założonym sterowaniu przez użytkowników i wiedzę (dane). Architektura platform powinna zapewniać jej dużą niezawodność, autonomiczność, bezpieczeństwo, skalowalność i możliwość adaptacji do różnych modeli obsługi. Realizacja platformy powinna przebiegać przy założeniach maksymalnej równoległości i wirtualizacji, a każdy z jej elementów może być usługą/zasobem. Planowana realizacja programu w latach 2008-2013 zapewnia możliwość harmonizacji prac badawczo-rozwojowych z biegnącym równoległe 7. Programem Ramowym w obszarze ICT.

Rozwój e-infrastruktury wraz z rozwojem technologii ICT sprawiają, że powstają nowe, ekscytujące możliwości badawcze w innych, odległych od ICT dziedzinach, np. w badaniach środowiska naturalnego. Przykładowo przedstawiona na ryc. 24 koncepcja zbudowania wirtualnego laboratorium środowiska naturalnego integruje:

- dostępną aparaturę eksperymentalną (stanowiącą źródło danych wejściowych poddawanych dalszej obróbce, jak sieci sensorowe, odbiorniki GPS, kamery, urządzenia telemetryczne itp.),
- infrastrukturę obliczeniową, składowania danych i wizualizacji,
- sieci komunikacyjne (sieć krajowa i sieci miejskie w dostępie do zasobów obliczeniowych, sieci mobilne i bezprzewodowe w dostępie do sensorów),
- wymagane oprogramowanie tworzące aplikacje integrujące dwa środowiska.

Laboratorium takie może być wykorzystane w wielu zastosowaniach, np. nauce i edukacji, ochronie środowiska, turystyce.

Powyższy przykład ilustruje tezę, że w niedalekiej przyszłości wiele dziedzin nauki, np. nauki przyrodnicze, humanistyczne, może zmienić swoje oblicze dzięki szerokiemu wkomponowaniu w ich procesy badawcze technologii ICT.



Ryc. 24. Wirtualne laboratorium środowiska naturalnego

5. Podsumowanie

Infrastruktura informatyczna nauki, zwana e-infrastrukturą, ma bardzo istotne znaczenie dla prowadzenia badań we wszystkich dyscyplinach nauki. Niektóre z dyscyplin naukowych, np. fizyka obliczeniowa, chemia obliczeniowa, radioastronomia, fizyka wysokich energii, nie mogłyby być dzisiaj rozwijane, gdyby nie integracja z e-infrastrukturą. Integracja ta może występować na dowolnym poziomie e-infrastruktury. Im ten poziom jest wyższy, tym powstałe aplikacje i usługi są bardziej kompleksowe i lepiej dopasowane do specyficznych wymagań zespołów naukowych.

Aby sprostać tym wymaganiom, w zakresie e-infrastruktury muszą być prowadzone intensywne prace naukowo-badawcze i badawczo-rozwojowe.

Integracja e-infrastruktury ze specyficznymi dyscyplinami nauki stanowi, ze względu na wymóg interdyscyplinarnego podejścia, bardzo trudną sztukę. Dotyczy to w szczególności dyscyplin dotychczas raczej odpornych na stosowanie narzędzi ICT, takich jak nauki przyrodnicze czy humanistyczne.

Podejmowanie przez naukę wielkich wyzwań jest ściśle związane z wielkim wyzwaniem integracji globalnie rozumianej e-infrastruktury ze wszystkimi dyscyplinami naukowymi, której celem jest stworzenie pełnej wizji e-nauki.

Literatura

- [1] M. Nakonieczny, S. Starzak, M. Stroiński, J. Węglarz *Polish scientific broadband network: POL-34*, Computer Networks 30, 1998, str. 1669-1676.
- [2] J. Rychlewski, J. Węglarz, S. Starzak et al. *PIONIER: Polish Optical Internet*, ISTHmus2000, 11-13 kwietnia 2000, Poznań, str. 19-28.
- [3] D.E. Atkins, K.K. Droegemeir, S.I. Feldman et al. Report on the *National Science Foundations Blue-Ribbon Advisory Panel on Cyberinfrastructure „Revolutionizing Science and Engineering Through Cyberinfrastructure”*, January 2003.
- [4] *e-Infrastructure – Building blocks for the European research area*, Information Society and Media DG, December 2004.
- [5] http://ec.europa.eu/research/infrastructures/pdf/fp7-brochure_en.pdf
- [6] www.geant2.net
- [7] A. Binczewski, N. Meyer, J. Nabrzyski et al. *First experiences with the Polish Optical Internet*, Computer Networks 37 (2001) 747-759.
- [8] M. Campanella, M. Przybylski et al. *Multidomain End to End IP QoS and SLA – The Sequin Approach and First Experimental Results on a Pan-European Scale*, Springer 2003 Proceedings 2003 XV, str. 759.
- [9] W. Barbaix, A. Binczewski, M. Przybylski et al. *Testing ATRIUM with Demanding Applications*, TNC 2003, Zagrzeb.
- [10] A. Binczewski, B. Gajda, W. Procyk et al. *Management of IPv6 Networks with IPv4/IPv6 SNMP Gateway*, TNC 2003, Zagrzeb.
- [11] <http://www.ist-mupbed.org/>
- [12] A. Hanemann, J.W. Boote, E.L. Boyd et al. *PerfSONAR: A Service Oriented Architecture for Multi-Domain Network Monitoring*, [w:] Proceedings of the Third International Conference on Service Oriented Computing, Springer Verlag, LNCS 3826, str. 241-254, ACM Sigsoft and Sigweb, Amsterdam, December, 2005.
- [13] J.W. Boote, E.L. Boyd, J. Durand et al. *Towards Multi-Domain Monitoring for the European Research Networks*, [w:] Selected Papers from the TERENA Networking Conference, TERENA, 2005; also published in Computational Methods in Science and Technology, vol. 11(2), str. 91-100, Poznań, Poland, 2005.
- [14] <http://emanics.org>
- [15] M. Czyrnek, E. Kuśmierek, C. Mazurek, M. Stroiński *Large-Scale Multimedia Content Delivery over Optical Networks for Interactive TV Services*, The International Journal of Future Generation Computer Systems, Elsevier, Vol 22/8 (2006), str. 1018-1024.
- [16] E. Kuśmierek, M. Czyrnek, C. Mazurek, M. Stroiński *iTVP: Largescale content distribution for live and on-demand video services* in Multimedia Computing and Networking SPIEIS&T Electronic Imaging, R. Zimmermann and C. Griwodz, (red.), 2007, article CID 6504-8
- [17] R. Buyya, M. Pathan, A. Vakali, red. *Content Delivery Networks*, Springer, 2008, roz. 13: *CDN for Live and On-Demand Video Services over IP*.
- [18] J. Nabrzyski, S. Starzak, M. Stroiński, J. Węglarz *Concept of the Academic and Research European Multicolour Network*, ISTHmus2000, 11-13 kwietnia 2000, Poznań, str. 31-38.

- [19] A. Binczewski, M. Stroiński, J. Węglarz *Sieć nowej generacji dla elektronicznej Europy*, materiały Konferencji „Polski Internet Optyczny: Technologie, Usługi i Aplikacje PIONIER 2001”, 24-27 kwietnia 2001 r., Poznań, str. 371-382.
- [20] N. Meyer, M. Stroiński, J. Węglarz *Optical Network for Euro-Regional Cooperation*, materiały konferencyjne Polish-German Symposium Science Research Education, SRE' 2000, 28-29 września 2000, Zielona Góra.
- [21] *Optical Networks development in Eastern Europe and Southern Caucasus regions*, Parliament Magazine, 2007.
- [22] www.eduroam.org
- [23] I. Foster, C. Kesselman *The Grid: Blueprint for a New Computing Infrastructure*, Second Edition, Elsevier 2004.
- [24] J. Nabrzyski, J. Schopf, J. Węglarz (red.) *Grid Resource Management*, Kluwer Academic Publishers, Boston-Dordrecht-London, 2003.
- [25] *Globus Toolkit Version 4: Software for Service-oriented Systems*. I. Foster. IFIP International Conference on Network and Parallel Computing, Springer LNCS 3779, pp 2-13, 2006.
- [26] M.J. Litzkow, M. Livny, M.W. Mutka *Condor – A Hunter of Idle Workstations*, ICDCS 1988: 104-111.
- [27] D. Breuer, E. Dietmar, D. Mallmann et al. (2004) *Scientific Computing with UNICORE*, NIC Symposium 2004: February 2004, Forschungszentrum Jülich; proceedings D. Wolf et al. (red.), Jülich, Forschungszentrum, 2004, (NIC series; 20), str. 429-440.
- [28] I. Foster *The grid: Computing without Bounds*. Scientific American, March 2003.
- [29] A. Gabrielle, K. Davis, T. Goodale et al. *The Grid Application Toolkit: Towards Generic and Easy Application Programming Interfaces for the Grid*, Proceedings of the IEEE, 93(3), str. 534-550, (2005).
- [30] G. Allen, T. Dramlitsch, I. Foster et al. *Supporting Efficient Execution in Heterogeneous Distributed Computing Environments with Cactus and Globus*, Winning Paper for Gordon Bell Prize (Special Category), Supercomputing 2001 (Rev. version: August 2001).
- [31] I. Taylor, M. Shields, I. Wang, O. Rana *Triana Applications within Grid Computing and Peer to Peer Environments*, Journal of Grid Computing, 1(2): 199-217, 2003.
- [32] M. Bubak, J. Marco, H. Marten *CrossGrid – development of grid environment for interactive applications*, mat. konferencyjne PIONIER 2002, Poznań, kwiecień 2002, str. 97-112.
- [33] M. Bubak, M. Malawski, K. Zając *The CrossGrid Architecture: Applications, Tools, and Grid Services*, *Grid Computing*, First European Grids Conference, Santiago de Compostela, Spain, February 2003, LNCS 2970, Springer, str. 299-308.
- [34] M. Kosiedowski, M. Kupczyk, R. Lichwala et al. *Mobile work environment for Grid Users. Grid Applications' Framework*, Proceedings of the ICCS '03, Saint Petersburg 2003, Rosja.
- [35] M. Owsiak, B. Palak, M. Płóciennik *Graphical framework for Grid interactive and parallel applications*, Journal of Computing and Informatics, Bratislava, Slovakia, April 2008.
- [36] M. Bogdanski, M. Kosiedowski, C. Mazurek, M. Wolniewicz *GRID SERVICE PROVIDER: How to improve flexibility of grid user interfaces?*, Lecture Notes in Computer Science 2657: Proceedings of the International Conference on Computing Science ICCS 2003, Springer-Verlag, Berlin Heidelberg New York, 2003, str. 255-263.
- [37] M. Kosiedowski, K. Kurowski, C. Mazurek et al. *Workflow applications in GridLab and PROGRESS projects*, Concurrency and Computation: Practice and Experience, Vol. 18, Issue 10, John Wiley & Sons, Ltd., Chichester, 2006, str. 1141-1154.
- [38] M. Bogdanski, M. Kosiedowski, C. Mazurek et al. *Flexibility and user-friendliness of Grid*

- portals: the PROGRESS approach*, Concurrency and Computation: Practice and Experience, Vol. 19, Issue 6, John Wiley & Sons, Chichester, 2007, str. 827-838.
- [39] M. Bogdanski, M. Kosiedowski, C. Mazurek et al. *Plugging grids into computing portals: the PROGRESS grid resource broker plug-in mechanism*, Concurrency and Computation: Practice and Experience, Vol. 19, Issue 12, John Wiley & Sons, Chichester, 2007, str. 1653-1661.
- [40] M. Kosiedowski, C. Mazurek, M. Stroiński, *PROGRESS Access Environment to Computational Services Performed by Cluster of Sun Systems*, Proceedings of the 2nd Cracow Grid Workshop, Cracow, 2002, str. 45-56.
- [41] <http://www.sun.com/smi/Press/sunflash/2002-11/sunflash.20021118.6.html>
- [42] R. Wyrzykowski, N. Meyer, M. Stroiński *Concept and Implementation of CLUSTERIX: National Cluster of Linux Systems*, Springer LNCS 3470, red. P.M.A. Sloot et al. *Advances in Grid Computing – EGC 2005*, European Grid Conference, Amsterdam, February 14-16, 2005.
- [43] J. Pukacki, M. Kosiedowski, R. Mikołajczak et al. *Programming Grid Applications with Gridge*, Computational Methods in Science and Technology 12(1), 47-68 (2006).
- [44] M. Płóciennik, M. Owsiak, E. Fernández et al. *Int.eu.grid project approach on supporting interactive applications in grid environment* INGRID 2007, Springer, 2008 (w druku).
- [45] C. Mazurek, M. Stroiński, M. Werla, J. Węglarz *Metadata harvesting in regional digital libraries in PIONIER Network*, Campus-Wide Information Systems, Emerald, 2006, vol. 23, no 4, str. 241-253.
- [46] C. Mazurek, T. Parkoła, M. Werla *Atomowe usługi w środowisku rozproszonych bibliotek cyfrowych*, materiały konferencji „Technologie przetwarzania danych”, Wydawnictwo Politechniki Poznańskiej, Poznań, 2007, str. 446-457.
- [47] M. Kosiedowski, C. Mazurek, M. Stroiński et al. *Federating digital library services for advanced applications in science and education*, Computational Methods in Science and Technology, Scientific Publishers OWN, Poznań, 2007, vol. 13, str. 101-112.
- [48] M. Lawenda, N. Meyer, M. Stroiński et al. *Effective Job Management in the Virtual Laboratory*, Springer LNCS 3470, P.M.A. Sloot et al. (red.), *Advances in Grid Computing – EGC 2005*, European Grid Conference, Amsterdam, February 14-16, 2005, str. 414-423.
- [49] M. Okoń, M. Lawenda, M. Stroiński, N. Meyer *Virtual Laboratory as a Remote and Interactive Access to the Scientific Instrumentation Embedded in Grid Environment*, eScience 2006, 2nd IEEE International Conference, Amsterdam, Dec. 4-6, 2006.
- [50] T. Prokosch, N. Meyer, M. Lawenda et al. *What is remote Instrumentation?*, INGRID 2008 – Instrumenting the Grid, Lacco Ameno, Ischia, Włochy, kwiecień 9-11, 2008, Springer (w druku); <http://www.ingrid.cnit.it/Presentazioni/01-11.pdf>
- [51] M. Okoń, D. Stokłosa, D. Kaliszczan et al. *Grid Integration of Future Arrays of Broadband Radio-Telescopes moving towards e-VLBI*, INGRID 2007 – Instrumenting the Grid, (red.) F. Davoli et al. *Grid Enabled Remote Instrumentation*, Springer, 2008, str. 571-588.
- [52] <http://www.dorii.eu>
- [53] Wprowadzenie do projektu Expres, T. Charles Yun.
www.expres-eu.org/DOCS/internetworking-tcyun-009.ppt
- [54] D. Stokłosa *Workflow Management in Remote Instrumentation Infrastructure – Based on e-VLBI Experiences*, materiały z konferencji INGRID 2008 – Instrumenting the Grid, kwiecień 9-11, 2008, Ischia, Włochy, materiały konferencyjne Springer (w druku).
<http://www.ingrid.cnit.it/Presentazioni/06-01.pdf>
- [55] G.P. Collins *Machina do odkryć*, Świat Nauki, kwiecień 2008, nr 4(200) str. 37-43.
- [56] J.P. Nassalski *Polska częśćka w LHC*, Świat Nauki, kwiecień 2008, nr 4(200), str. 52-57.

- [57] H. Newman et al. *The UltraLight Project: The Network as an Integrated and Manager Resource for Data-Intensive Science*, Computing Science and Engineering, 2005 (vol. 7, No 6), str. 38-47.
- [58] www.telepath.poznan.pl
- [59] J. Błaszczyński, M. Kosiedowski, C. Mazurek et al. *Telemedical Portal „Telemedycyna Wielkopolska”*, MIT 2006 Medical Informatics & Technology, (red.) E. Pietka et al. Wisła 2006, str. 230-235.
- [60] M. Kosiedowski, C. Mazurek, M. Stroiński, J. Węglarz *Grid-supported Medical Digital Library*, Studies in Health Technology and Informatics 126, 127-136 (2007).
- [61] J. Błaszczyński, M. Kosiedowski, C. Mazurek, Sz. Wilk *Ontologies for Knowledge Modeling and Creating User Interface in the Framework of Telemedical Portal*, [w:] H. Stormer et al. (red.): European Conference on eHealth 2006. Fribourg, Switzerland, October 12-13, 2006. Lecture Notes in Informatics P-91. Gesellschaft für Informatik, Bonn, 2006, 275-286.
- [62] J. Nabrzyski, J. Pukacki, M. Tsiknakis et al. *A Semantic Grid Infrastructure Enabling Integrated Access and Analysis of Multilevel Biomedical Data in Support of Postgenomic Clinical Trials on Cancer*, [w:] IEEE Transactions on Information Technology in Biomedicine, vol. 12, 2007.
- [63] *Future for European Grids: Grids and Service-Oriented Knowledge Utilities, Vision and research directions 2010 and beyond*, Next Generation Grids Expert Group report 3, University Society and Media, January 2006.
- [64] G. Zervas, E. Escalona, R. Nejabati et al. *Phosphorus Grid-enabled GMPLS Control Plane (G2MPLS): Architectures, Services and Interfaces (UESsex, Nextworks, PSNC, AIT)*. Accepted for publication in the IEEE Communications Magazine under the Special call on *Multi-Domain Optical Networks: Issues and Challenges*.
- [65] <http://www.ist-phosphorus.eu/standardization.php>
- [66] C. Citko S. Starzak, M. Stroiński et al. *Koncepcja programu PIONIER2 – Zaawansowane Platformy Usługowe, Program na lata 2007-2013*, Poznań-Kraków-Łódź-Białystok, czerwiec-listopad 2006 r.

On the importance of the e-infrastructure for science

In this article the European scientific e-infrastructure is described with special emphasis on the contribution of the Polish scientific community. Examples of this infrastructure for radio-astronomy, high-energy physics and medicine are presented. Directions of further development in the field are pointed out.

Key words: e-infrastructure, computer networks, grids, information and telecommunication technology