

Biologia: BioKepler

- [Krótki opis usługi](#)
- [Aktywowanie usługi](#)
- [Pierwsze kroki](#)
 - [Instalacja oprogramowania x2go](#)
 - [Konfiguracja klienta x2go:](#)
 - [Dalsza konfiguracja:](#)
 - [Łączenie ze środowiskiem pracy użytkownika BioKepler \(na zasobach chmurowych\)](#)
 - [Uruchomienie oprogramowania BioKepler \(poprzez ikonę kepler.sh\).](#)
 - [Załadowanie i uruchomienie przykładowego scenariusza](#)
- [Zaawansowane użycie](#)
- [Gdzie szukać dalszych informacji?](#)

Krótki opis usługi

Usługa jest przeznaczona dla biologów i bioinformatyków.

Usługa **BioKepler** składa się z kilku elementów: dostarczane jest całe wirtualne środowisko pracy dostępne poprzez zdalny desktop: maszyna wirtualna uruchomiona na zasobach chmurowych, a na niej zainstalowany zestaw popularnych narzędzi i aplikacji bioinformatycznych oraz oprogramowanie BioKepler. BioKepler jest modulem platformy zarządzania naukowymi scenariuszami obliczeniowymi - **Keplera**, który służy do uruchamiania zestawu narzędzi bioinformatycznych w rozproszonych środowiskach obliczeniowych takich jak PLGrid. Użytkownik może tworzyć całe scenariusze obliczeniowe (ang. *workflows* lub *pipelines*), składające się z wielu kroków, a poszczególne komponenty mogą być uruchamiane na różnych zasobach obliczeniowych.

Przykładowe komponenty BioKeplera (tzw. aktorzy): 2bwt-builder, DynamicTrim.pl, RNA_parse.pl SOAPdenovo31mer, blast_rRNA.pl, blastall, bowtie, bowtie-build, bowtie2-build, breakdancer-max, bwa_align bwa_index, cd-hit cd-hit-454, cd-hit-est, clustalw, clustalx, cuffdiff, cytoscape, faa_stat.pl, fastq2fasta.py, fraggene_scan.pl, fraggene_scan_parse.pl, hmm_rRNA.pl, hmmsearch.pl, hmmsearch_parse.pl, kegg.pl, kegg_parse.pl, metagene.pl, mother, muscle orf_2_tbl.pl, orf_finder, qc_filter.pl, qc_filter_fastq.pl, qiime rasmol, rpsblast.pl, rpsblast_parse.pl, samtools, soap, ssake, tRNAscan-SE.pl, tophat, uchime, velvet, velvet

[Środowisko Kepler](#), w którym scenariusze są tworzone graficznie poprzez łączenie aktorów (elementarnych cegiełek wykonujących określone zadanie) przy użyciu kabli (które transportują dane), ma wbudowaną bibliotekę ponad 400 aktorów (np. 'wyrażenie matematyczne', 'czytnik danych tabelarycznych', 'kod Java', 'skrypt Python', 'narysuj wykres'). Gotowe scenariusze mogą być uruchamiane w trybie graficznym bądź z linii komend.

Aktywowanie usługi

Aby skorzystać z usługi **BioKepler**, należy mieć aktywne [konto w Infrastrukturze PLGrid](#) oraz aktywną [afiliację](#).

Do usługi **BioKepler** może uzyskać dostęp każdy użytkownik PLGrid, który jest użytkownikiem usługi **Molecular Biology Data Analysis Toolkit**. W celu uzyskania dostępu do tej usługi należy po zalogowaniu się w [Portalu PLGrid](#) w lewym menu wybrać zakładkę **Usługi**. Po przejściu do tej zakładki należy kliknąć zielony przycisk **Zarządzaj usługami**, znajdujący się w prawym górnym rogu. Spowoduje to przejście do [Katalogu Aplikacji i Usług](#) (KAiU), gdzie należy wyszukać usługę **Molecular Biology Data Analysis Toolkit**, a następnie o nią aplikować.

Przyznanie usługi następuje automatycznie. Gdy usługa zostanie przyznana, pojawi się na liście usług w portalu PLGrid ze statusem „active” Status usługi będzie również widoczny w KAiU jako *”Status użytkownika: Usługa aktywna — usługa dostępna dla użytkownika”*.

Pierwsze kroki

Poniżej przedstawione są następujące kroki:

- [instalacja oprogramowania x2go](#)
- [konfiguracja x2go](#)
- [łączenie ze środowiskiem pracy użytkownika BioKepler \(na zasobach chmurowych\)](#)
- [uruchomienie oprogramowania BioKepler](#)
- [załadowanie i uruchomienie przykładowego scenariusza \(ang. Workflows\)](#)

Instalacja oprogramowania x2go

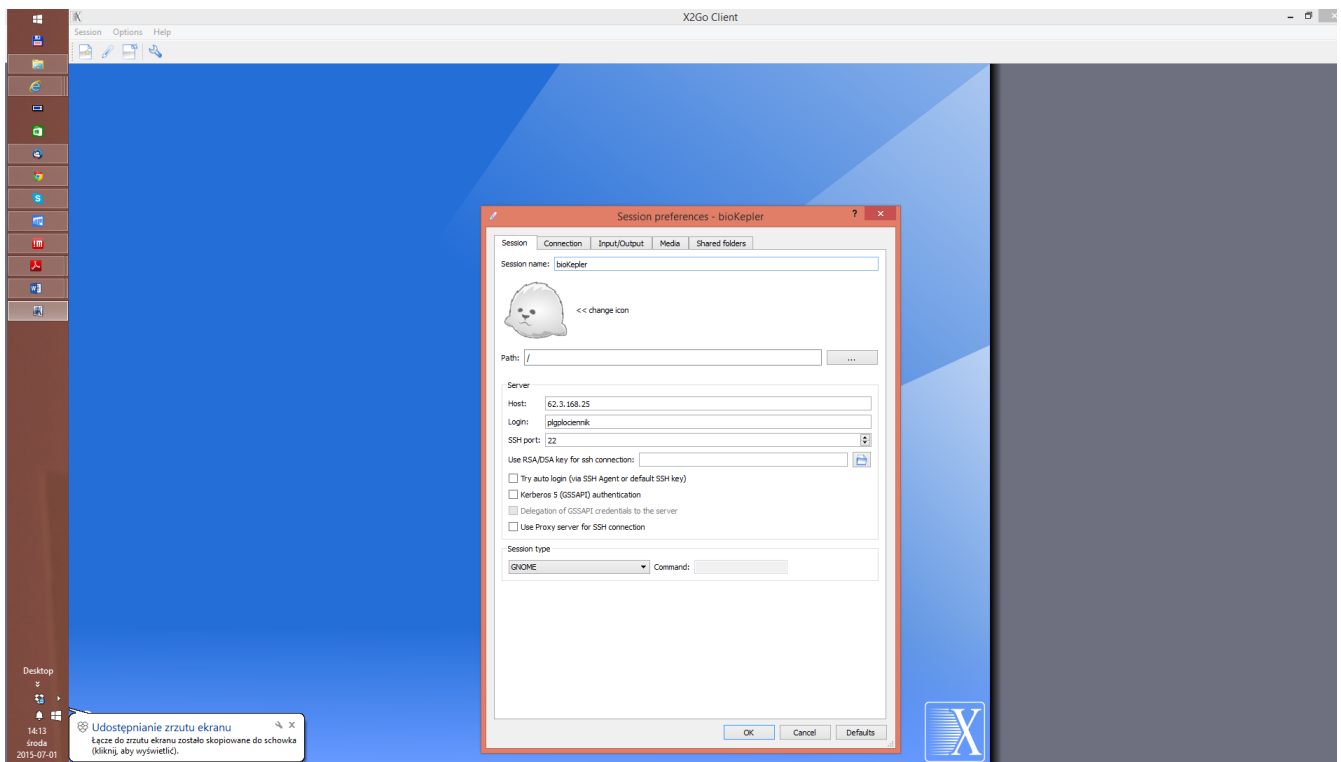
1. Pobranie oprogramowania x2go (klient) ze strony: <http://wiki.x2go.org/doku.php/doc:installation:x2goclient> - na odpowiednią platformę.
2. Instalacja klienta x2go

Konfiguracja klienta x2go:

1. Session name: BioKepler
2. Host: 62.3.168.25
3. login: plgusername

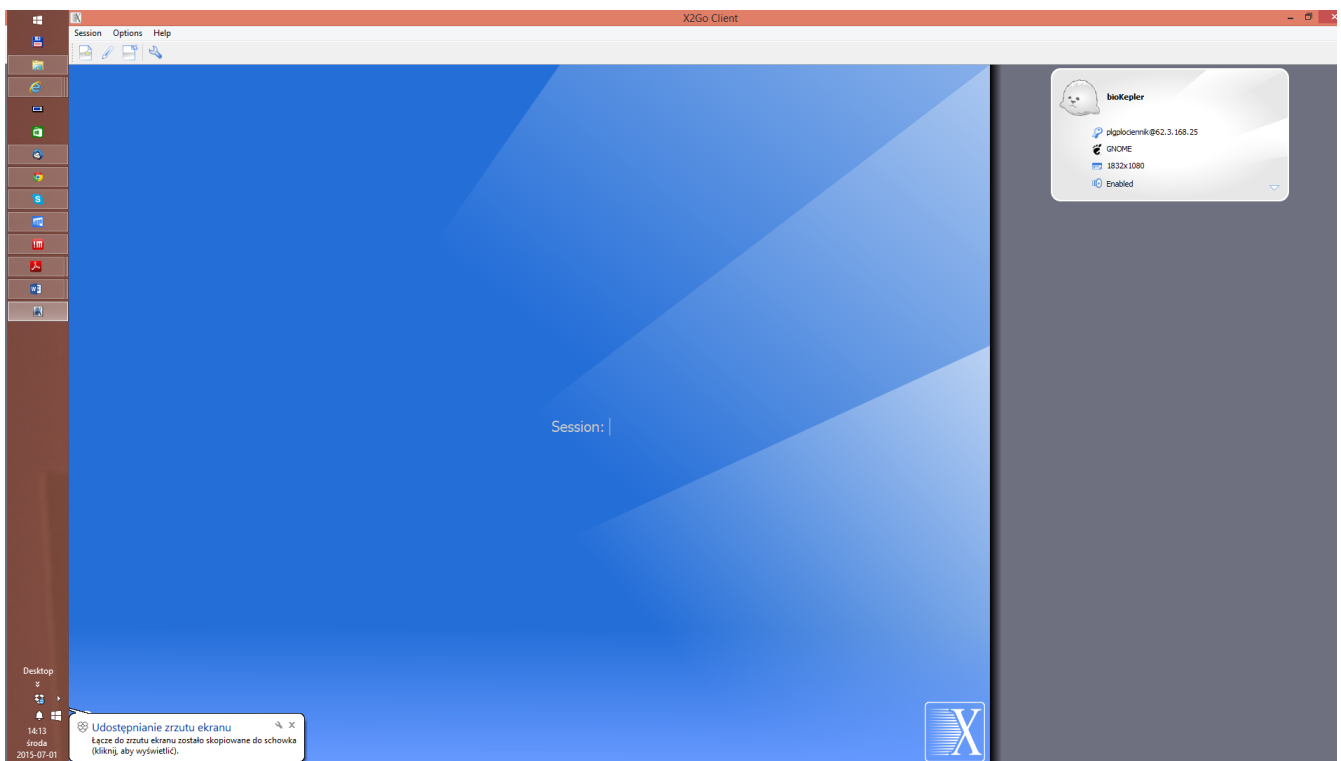
4. sessiontype: GNOME
5. "OK" doda nową sesję do listy dostępnych sesji na pasku po prawej stronie

Uwaga: można stworzyć sobie poprzez X2Go kilka sesji, na raz może być aktywna tylko jedna z nich.



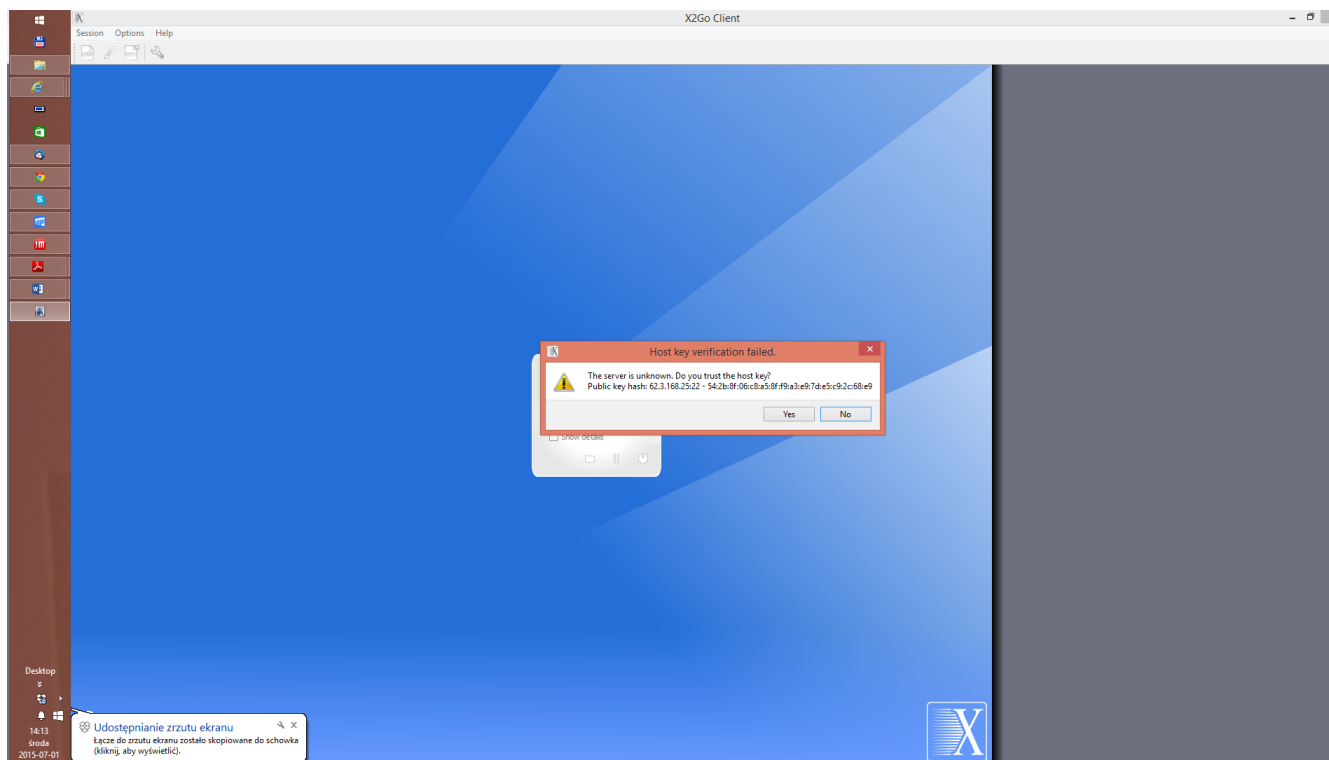
Dalsza konfiguracja:

1. Zmiana rozmiaru zdalnego desktopu: wg. preferencji

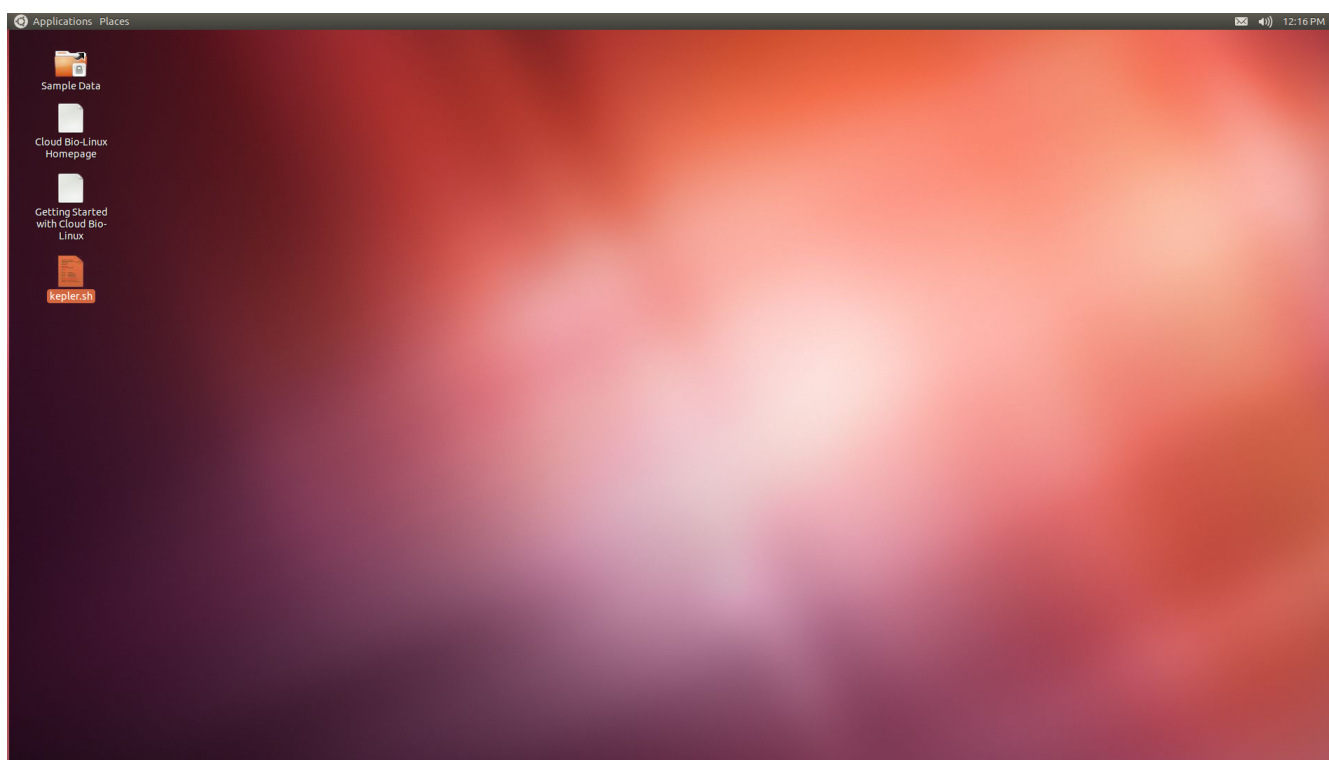


Łączenie ze środowiskiem pracy użytkownika BioKepler (na zasobach chmurowych)

1. Wybór skonfigurowanej sesji
2. podanie hasła
3. potwierdzenie zaufania dla kluczy



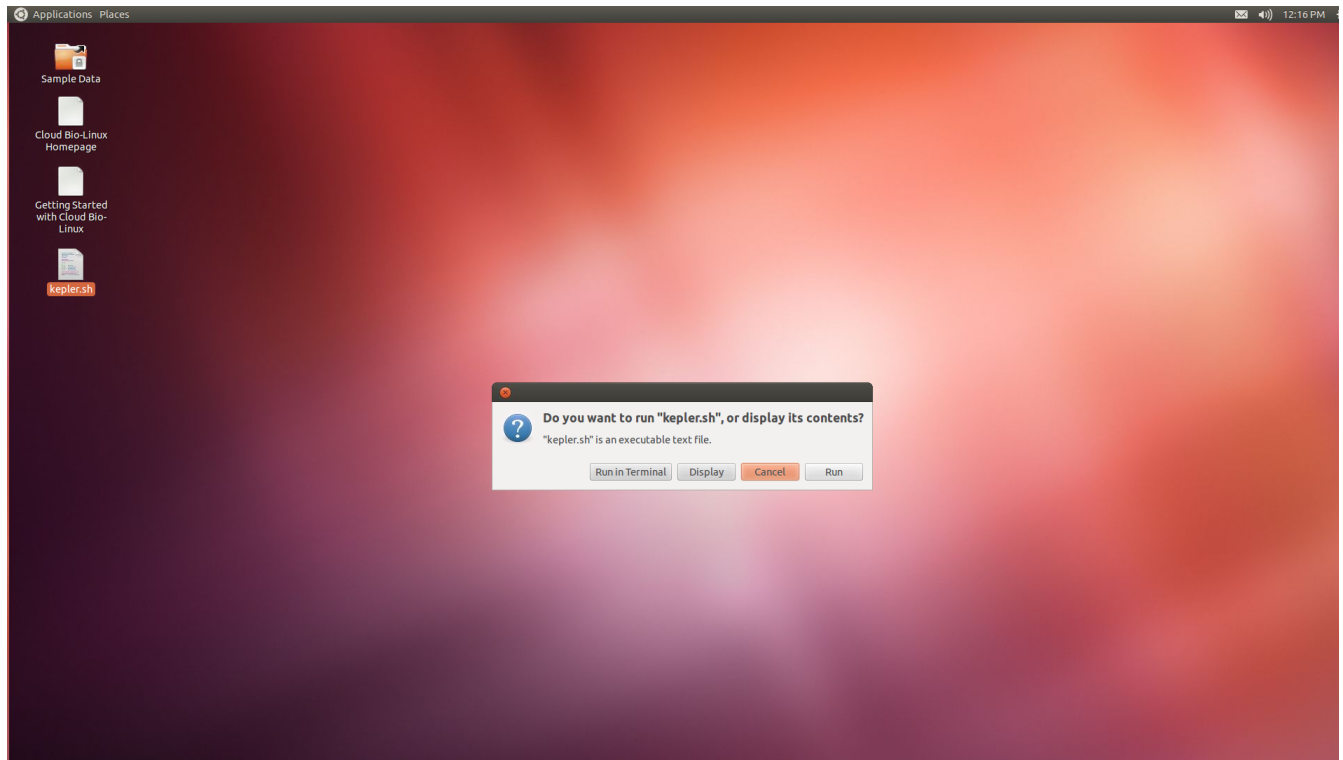
Po połączeniu ze zdalnym środowiskiem pracy, powinien pojawić się zdalny pulpit, na której między innymi znajduje się ikona do narzędzia BioKepler: kepler.sh



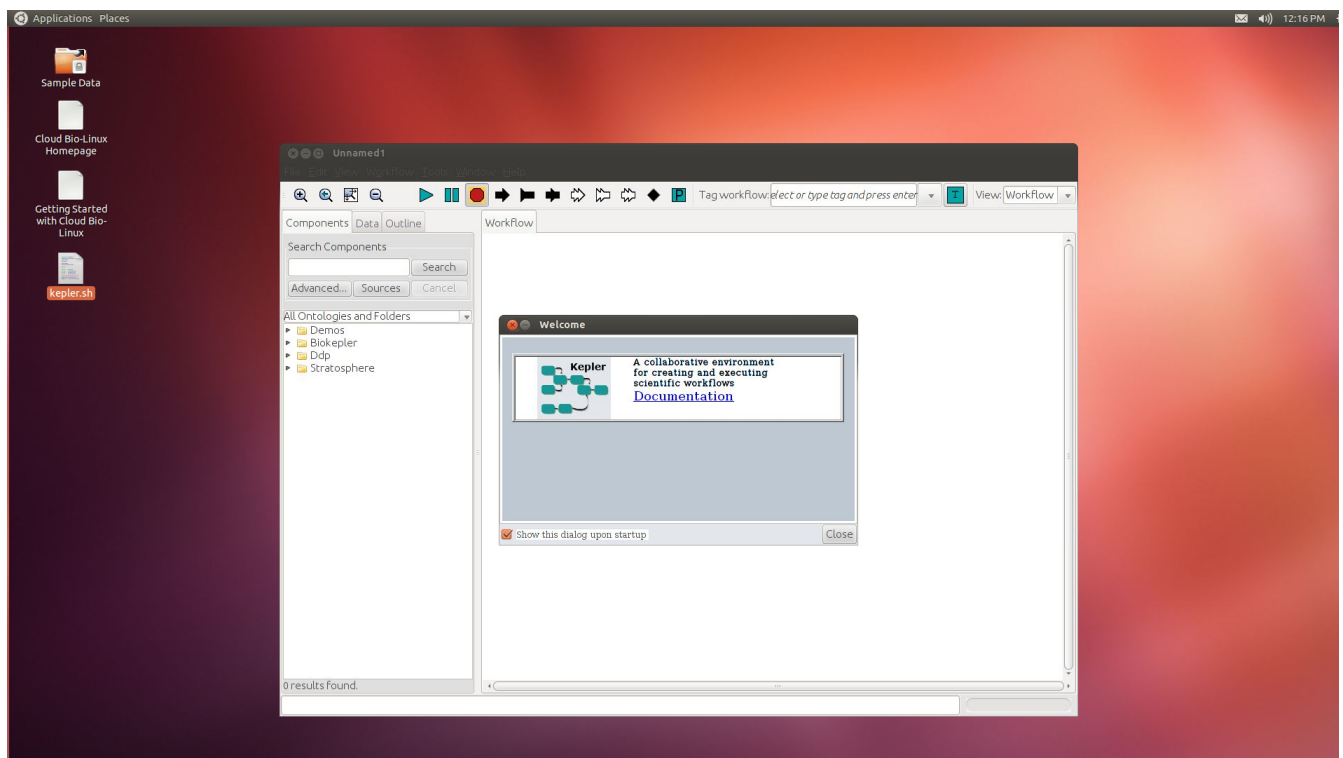
Uruchomienie oprogramowania BioKepler (poprzez ikonę kepler.sh).

"Run in terminal" da możliwość obserwowania komunikatów w konsoli.

Pierwsze uruchomienie BioKeplera będzie trwało dłużej ze względu na automatyczne tworzenie katalogów i kopiowaniu plików do katalogu domowego..



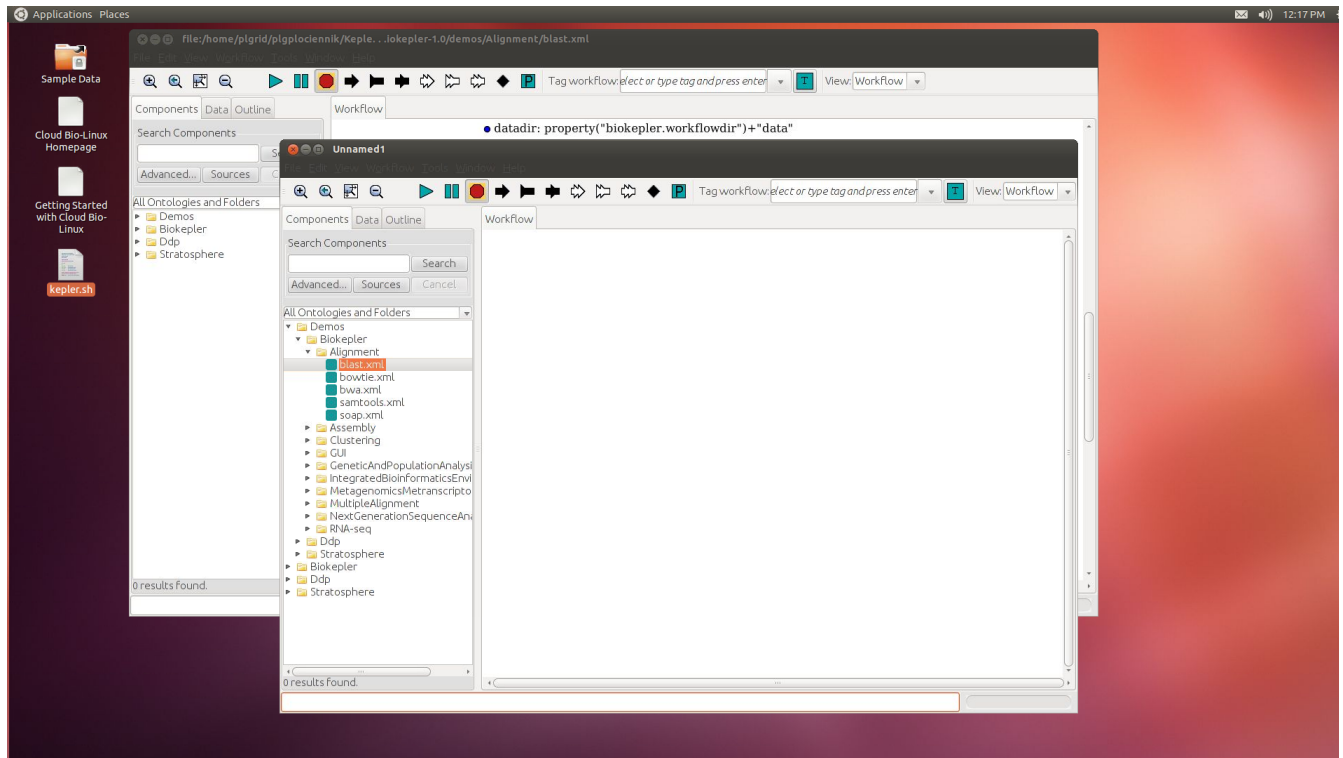
Po uruchomieniu pojawi się ekran powitalny aplikacji, który należy zamknąć(można zaznaczyć na dole okna, żeby nie pojawiał się przy następnych uruchomieniach).



Łaďadowanie i uruchomienie przykładowego scenariusza

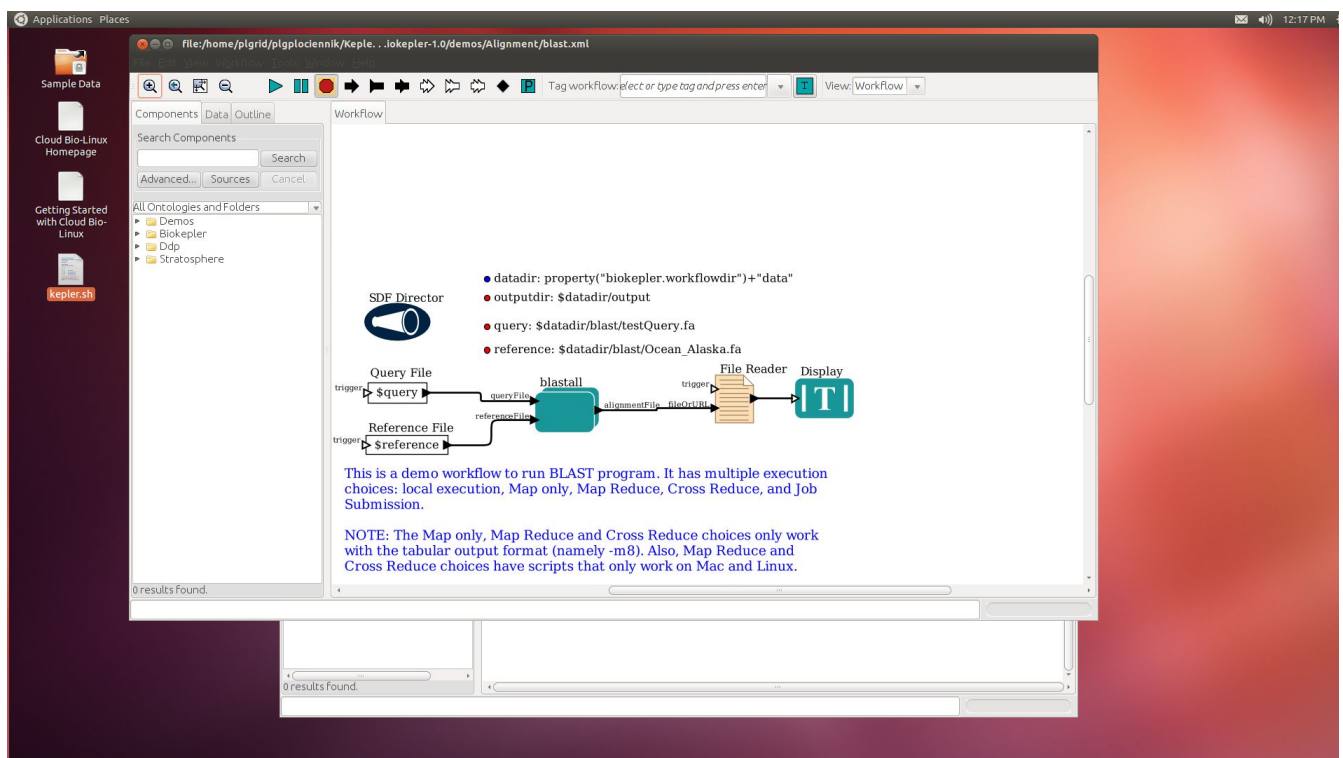
Z zakładki po lewej stronie przestrzeni roboczej (dashboard) można wybrać gotowe scenariusze lub utworzyć samemu nowy scenariusz. Dla przykładu wpisując "blast" lub wybierając "Demos BioKepler Alignment Blast.xml"

i klikając podwójnie na ikonę można otworzyć przykładowy workflow



Poniewaű workflow jest prekonfigurowany i sű dostępne przykładowe dane wejściowe, można od razu przejűć do uruchomienia scenariusza.

Moűna to osiűgnűć poprzez naciűnięcie ikony "press"



Wynik uruchomienia tego scenariusza powinien się pojawić w osobnym oknie (blast.Display)

The screenshot shows the Biokepler application interface. The main window displays a workflow named 'blast.xml' with the following parameters:

- dataDir: property("biokepler.workflowdir")+"data"
- outputDir: \$dataDir/output
- query: \$dataDir/blast/testQuery.fa
- reference: \$dataDir/blast/Ocean_Alaska.fa

A 'blast.Display' window is open, showing a table of search results. The table has columns for 'Query ID', 'Subject ID', 'Score', 'E-value', 'Bits', 'Query Length', 'Subject Length', 'Query Start', 'Query End', 'Subject Start', and 'Subject End'. The results are as follows:

Query ID	Subject ID	Score	E-value	Bits	Query Length	Subject Length	Query Start	Query End	Subject Start	Subject End
69.7233679	17539182	85.71	98	14	0	2	99	12	109	
69.7233679	17534724	94.87	39	2	0	49	87	4	42	
69.7237027	17550370	89.83	59	6	0	66	124	17	75	
69.7237027	17546601	89.29	56	6	0	69	124	19	74	
69.7237027	17545076	84.62	65	10	0	61	125	88	24	
69.7240509	17544901	93.94	33	2	0	7	39	37	5	
69.7240509	17538493	93.94	33	2	0	7	39	37	5	
69.7240509	17525674	93.94	33	2	0	7	39	37	5	
69.7243158	17544866	83.91	87	14	0	1	87	89	3	
69.7243158	17536071	94.29	35	2	0	74	108	62	28	
69.7246439	17525209	86.73	98	13	0	23	120	112	15	
69.7246439	17546412	87.69	65	8	0	29	93	39	103	
69.7246439	17525202	86.36	66	8	1	29	93	17	82	
69.7246749	17542322	90.00	40	4	0	6	45	40	1	
69.7248173	17549965	100.00	24	0	0	81	104	50	73	
69.7248173	17540078	100.00	24	0	0	81	104	65	88	
69.7248173	17531612	100.00	24	0	0	81	104	75	98	
69.7248173	17529086	100.00	24	0	0	81	104	59	82	
69.7248173	17523468	100.00	24	0	0	81	104	60	83	
69.7248173	17522675	100.00	24	0	0	81	104	66	89	
69.7253771	17522617	84.85	66	10	0	52	117	1	66	
69.7261661	17532552	85.15	101	14	1	3	103	103	4	

The status bar indicates '0 results found' and 'execution finished: 3674 ms. Memory: 464960K Free: 341684K (73%)'.

Zaawansowane użycie

Analogicznie do przykładu przedstawionego w sekcji pierwsze kroki, można uruchomić kilkanaście innych gotowych scenariuszy.

Poniżej przykładowe wyniki uruchomień przykładowych scenariuszy:

Soap

The screenshot shows the Biokepler application interface. The main window displays a workflow named 'soap.xml' with the following parameters:

- dataDir: property("biokepler.workflowdir")+"data"
- outputDir: \$dataDir/output
- querySequence: \$dataDir/soap/test.fna
- referenceSequence: \$dataDir/soap/NC_009513.fna

A 'soap alignment display' window is open, showing a table of alignment results. The table has columns for 'Query ID', 'Subject ID', 'Score', 'E-value', 'Bits', 'Query Length', 'Subject Length', 'Query Start', 'Query End', 'Subject Start', and 'Subject End'. The results are as follows:

Query ID	Subject ID	Score	E-value	Bits	Query Length	Subject Length	Query Start	Query End	Subject Start	Subject End
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3

The status bar indicates '0 results found' and 'execution finished: 3674 ms. Memory: 464960K Free: 341684K (73%)'.

Assembly/velvet

The screenshot shows the BioKepler application interface. The main window displays a workflow diagram with components like 'SDF Director' and 'velvet.Contig display'. A terminal window shows the execution of the velvet tool, displaying sequence data for various nodes. A text box provides information about the velvet tool version and download link.

Velvet tool (version 1.2.03) for de novo assembly can be downloaded from:
<http://www.ebi.ac.uk/~zerbino/velvet>.

SAMTOOLS

The screenshot shows the BioKepler application interface. The main window displays a workflow diagram with components like 'samtools.SAM File Display' and 'SAM File Display'. A terminal window shows the execution of the samtools tool, displaying sequence data for various nodes. A text box provides information about the samtools tool version and download link.

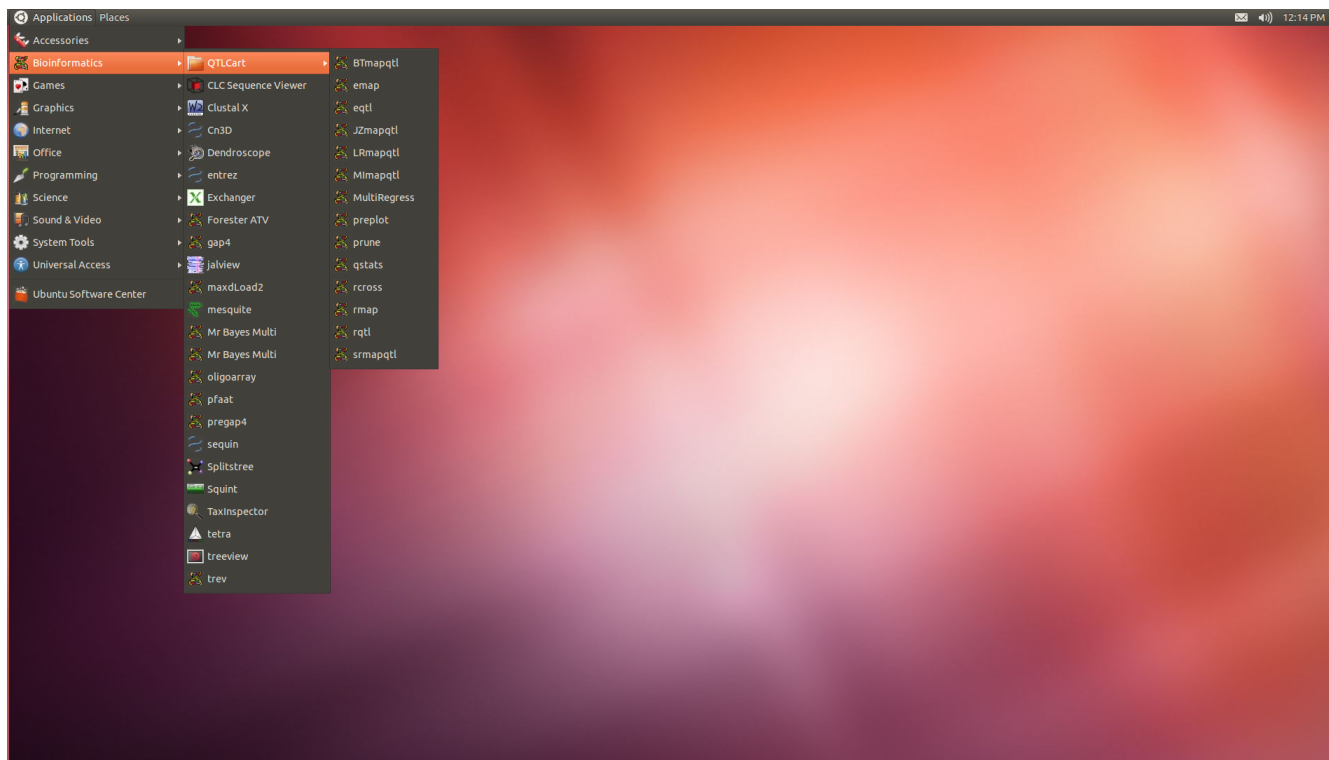
bioKepler team (biokepler.org)
 April-22-2013

NOTE: This workflow requires samtools to be installed and accessible from \$PATH.

The samtools package (version 0.1.18) for processing files in SAM format can be downloaded from:
<http://samtools.sourceforge.net>.

Zestaw kilkuset najczęściej używanych programów domenowych jest dostępnych na ścieżce uruchomieniowej, może być również wykorzystany poprzez BioKeplera.

Część narzędzi bioinformatycznych jest również dostępnych z poziomu menu (na przykład: Dendroscope, cn3d, sequin, splitsree, tetra, trev, squint, etc)



W sprawie konstruowania innych scenariuszy/pomocy przy konfigurowaniu obecnych prosimy kontaktować się przez [helpdesk PLGrid](#).

Gdzie szukać dalszych informacji?

Szczegółowe informacje o użytkowaniu infrastruktury PLGrid znajdują się w [Podręczniku Użytkownika](#).

Szczegółowa dokumentacja programu BioKepler znajduje się na stronie: <http://www.biokepler.org/userguide>

Szczegółowa dokumentacja platformy Kepler znajduje się w pliku: [dokumentacja.pdf](#)

Informacje o usługach dziedzinowych *Biologia* dostępne są na stronie: <http://biologia.plgrid.pl/>

Uzyskanie informacji/helpdesk PLGrid: [dokumentacji o pomocy](#)