

2159-PS

MODEL TO INTEGRATION AND ANALYSIS OF REMOTE EXPERIMENTS EXECUTED IN BEE WEBLABS

Marcelo Queiroz Leite (Universidade de São Paulo, SP, Brasil) - marcelo.queiroz@poli.usp.br
Leandro Halle Najm (Universidade de São Paulo, SP, Brasil) - leandro.najm@poli.usp.br
Pedro Luiz Pizzigatti Corrêa (Universidade de São Paulo, SP, Brasil) pedro.correa@poli.usp.br

The use of computer technology as tools to aid biologists, facilitated the implementation of experiments in remote environments, but, the little knowledge of researchers in this environment of technology changes has become an obstacle to its use. This paper discusses the integration of data collected by different technologies of sensors in hives, monitored by biologists and recorded remotely by a weblab tool called Bioabelha, available on the Internet for controlled access by researchers in order to obtain efficient means of data collection and support the scientific analysis using a specific analysis tool called Kepler. A case study is demonstrated in which it presents the result obtained with this integration.

Keywords: bioinformatics, kepler, metadata, sensors, weblab.

MODELO PARA INTEGRAÇÃO E ANÁLISE DE EXPERIMENTOS REMOTOS EXECUTADOS EM WEBLABS DE ABELHAS

O uso das tecnologias computacionais como ferramentas de auxílio aos biólogos, tem facilitado a execução de experimentos em ambientes remotos, por outro lado, o pouco conhecimento destes pesquisadores neste ambiente de mudanças tecnológicas tem se tornado um obstáculo ao seu uso. Este trabalho discute a integração de dados, coletados por diferentes tecnologias de sensores instalados em colméias, monitoradas por biólogos e cadastradas remotamente por uma ferramenta de weblab chamada Bioabelha, disponibilizada na Internet para acesso controlado por pesquisadores, a fim de se obter meios eficientes de coleta de dados e suporte a análises científicas utilizando uma ferramenta desenvolvida para esta atividade, neste caso a ferramenta Kepler. Um estudo de caso é demonstrado no qual se apresentam os resultados obtidos com esta integração.

Palavras-chave: bioinformática, kepler, metadata, sensores, weblab.

1. Introdução

O avanço das tecnologias computacionais tem se tornado um aliado aos cientistas. Ao permitir a execução de experimentos com um número cada vez maior de elementos em suas pesquisas, ao adicionar recursos para administrar o controle de variáveis relacionado aos experimentos com um número elevado de itens, precisão nas informações coletadas, meios para facilitar o compartilhamento dos resultados obtidos ao mesmo tempo em que permite o uso de informações resultantes de outros experimentos. Pode se perceber as contribuições que os avanços tecnológicos têm dado aos ambientes científicos executados com o auxílio da computação. Além disso, é preciso acrescentar o uso de ambientes simulados, a disponibilidade de redes de sensores, capazes de coletar informações a partir dos mais variados tipos de ambientes e menos invasivos ao habitat, a utilização de ferramentas computacionais para criação e também introdução de massas de dados em modelos matemáticos validando ou não estas análises, são outros exemplos dos campos da ciência que se beneficiaram com a evolução dos softwares e hardwares.

Dentro do contexto de pesquisas em biodiversidade, os laboratórios de acesso remoto, comumente conhecidos como *weblabs*, demonstram mais uma vez a capacidade que a computação tem de fornecer recursos que adicionam aos pesquisadores, meios de execução e monitoramento de experimentos mesmo em áreas ao qual o cientista não se encontra presente, permitindo uma redução de custos com laboratórios e equipamentos, e um ambiente de maior disponibilidade de acesso e horários, visto que um *weblab* pode funcionar 24 horas por dia e 7 dias por semana. Os *weblabs* geralmente são desenvolvidos tendo como objetivo fornecer recursos para atender a duas funcionalidades específicas (Zubia et al-a,2005:49):

- Coleta de dados – permitir ao(s) pesquisador (es) armazenar a maior quantidade possível de dados a respeito do experimento em execução;
- Controlar equipamentos à distância – por meio de redes de computadores privadas ou da própria Internet, utilizando serviços WEB estas ferramentas permitem aos cientistas efetuarem configurações específicas nos dispositivos controlados por meio do *weblab* em uso.

A análise dos dados obtidos por meio de um experimento executado com o auxílio de um *weblab*, nem sempre pode ser considerado uma tarefa simples. Geralmente como o foco dos *weblabs* não se concentra na análise, esta tarefa nem sempre pode ser considerada como uma atividade trivial,

Para que ocorra o processamento e a geração de resultados por meio dos dados coletados com o experimento, estes dados precisam ser disponibilizados a uma ferramenta capaz de efetuar a análise destas informações por meio de modelos matemáticos elaborados pelos cientistas responsáveis pelo experimento. Alguns projetos provêm uma arquitetura de integração para ferramentas específicas em análise, como é o caso do projeto SEEK (SEEK, 2004).

Este artigo descreve uma arquitetura em camadas cujo foco é o de permitir a integração de *weblabs* voltados a experimentos em biodiversidade, com ferramentas de análises científicas voltadas a esta área. Sua especificação foi inicialmente elaborada com o foco de atender ao

laboratório de abelhas da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo.

Dentre as características desta arquitetura destaca-se a capacidade de integrar um ou mais *weblabs* dentro de uma mesma análise, permitindo aos pesquisadores, análises conjuntas de experimentos executados em diferentes regiões.

O texto está organizado da seguinte forma. A seção 2 apresenta uma descrição de trabalhos relacionados com a proposta, tecnologias empregadas no trabalho e ferramentas de workflow e análises utilizadas no meio científico. A seção 3 descreve uma proposta de arquitetura a ser empregada para a integração de *weblabs* com ferramentas de análises. Na seção 4 é descrito o resultado do trabalho, em que um estudo de caso em andamento utiliza o modelo proposto. As conclusões e perspectivas são apresentadas na seção 5 e as referências citadas no trabalho estão presentes na seção 6.

2. Metodologia

Esta seção visa apresentar trabalhos diretamente ligados ao projeto proposto e também descreve tecnologias similares utilizando-as como referências para o desenvolvimento deste trabalho. O *weblab* utilizado no Laboratório de Abelhas, desenvolvida pelo LAA (Laboratório de Automação Agrícola da Escola Politécnica da USP) e utilizado pelo Instituto de Biociências e pela USP de Ribeirão Preto, sendo ambos utilizados para estudo de polinizadores, a escolha da ferramenta KEPLER (KEPLER, 2006) é efetuada por possuir um modelo de arquitetura de integração consistente, o que favorece a demonstração da solução proposta.

2.1. Arquitetura de um experimento

O uso de sistemas especialistas em análises e desenvolvimento de *workflows* científicos tem se mostrado eficientes ferramentas de auxílio na elaboração dos mais diversificados experimentos científicos. A Figura 1 apresenta o modelo de referência apresentado pelo projeto *myExperiment* (GOBLE et al., 2007:1) que descreve por meio de *workflows* as etapas inerentes ao ciclo de um experimento científico.

As camadas de um experimento do modelo proposto pelo *myExperiment* (MYEXPERIMENT, 2007) são descritas a seguir:

- a) Criação do experimento - neste item são estabelecidas as hipóteses a serem validadas pelo experimento, os resultados esperados e os pesquisadores envolvidos no experimento;
- b) Personalização - definição dos recursos utilizados na pesquisa e período de execução do experimento entre outros aspectos;
- c) Execução e monitoramento - etapa em que ocorrem as atividades envolvidas diretamente com o processo de execução do experimento;
- d) Análises e resultado do experimento - neste momento as informações coletadas devem ser utilizadas como dados de entrada para ferramentas que possam efetuar cálculos a partir de modelos matemáticos e outros tipos de análises de acordo com o experimento efetuado;

e) Compartilhamento de resultados - após o encerramento do experimento e a geração dos resultados obtidos com esta pesquisa, deve ser possível disponibilizar estas informações para avaliação e análises de outros pesquisadores;

f) Reutilização de dados para novos experimentos - após a finalização do trabalho de coleta de dados do experimento realizado tem se agora, a possibilidade de reuso destas informações como parâmetros de entrada ou comparação para outros experimentos.



Figura 1 - Ciclo de experimento do *myExperiment* (MYEXPERIMENT, 2007)

O trabalho de coleta de informações em um experimento, seja ele em um ambiente controlado, seja em um ambiente aberto, visando seu estudo por meio de ferramentas específicas em análises não deve ser definido como uma tarefa simples. Após o encerramento das atividades de coleta, os pesquisadores se deparam com uma quantidade de dados brutos, que neste momento devem ser analisados, filtrados e processados, disponibilizando desta maneira novas informações que serão os resultados encontrados neste experimento.

Dentro deste modelo proposto em camadas, tem-se como foco deste trabalho, a descrição de uma arquitetura para integração dos dados, obtidos após a execução de um experimento efetuado com o uso de *weblabs*, com ferramentas de análises capazes de processar as informações obtidas nestes experimentos. O modelo apresentado destina-se a pesquisas em biodiversidade, não significando que seu modelo não possa ser empregado em outras áreas da ciência que utilizem tecnologias similares em seus experimentos científicos.

Das tecnologias empregadas para elaboração da proposta, destaca-se a utilização de metadados específicos para integração de elementos de biodiversidade, onde se efetua um breve descritivo de algumas propostas existentes neste campo, descreve-se resumidamente uma arquitetura de referencia em integração de pesquisas científicas conhecida como SEEK, e por último são apresentados alguns softwares empregados nas áreas de modelagens de *workflows* e de análises científicas, com maior ênfase ao utilizado nesta pesquisa, a ferramenta KEPLER.

2.2. *Weblab*

Dentre os recursos disponibilizados aos pesquisadores, o *weblab* vem se tornando uma ferramenta de assistência com uma boa aceitação ao meio científico, pois por meio desta tecnologia, experimentos podem ser compartilhados e avaliados por vários cientistas, desde que estes estejam autorizados a integrarem a equipe responsável pelo experimento em execução. Um *weblab* ou laboratório de acesso remoto é um sistema integrado de hardware e software que permite o controle e o acompanhamento de um dispositivo eletrônico programável via WEB.

Weblabs são usualmente empregados nas universidades para que estudantes possam acessar os recursos dos laboratórios de qualquer lugar por meio de uma conexão com a Internet. (ZUBIA et al-b, 2005:2146).

O foco inicial de um *weblab* concentra-se na operação dos dispositivos eletrônicos e também na coleta das informações ao qual o experimento foi inicialmente preparado. Sendo uma ferramenta operada por meio de redes de computadores, há a necessidade de preparação de todo um ambiente para este controle. Dentre estes itens integrantes destacam-se:

- Servidor HTTP, em que se disponibiliza a aplicação que deve ser acessada pelo(s) pesquisador(es);
- Servidor de Banco de Dados, onde se encontra a base de dados responsável pelo armazenamento das informações adquiridas no experimento;
- Equipamentos específicos, aqui se enquadram os hardwares que são administrados pelo *weblab*, como exemplo destes dispositivos pode se destacar: redes de sensores, capazes de coletar diversos itens, como temperatura e umidade ou mesmo ambientes, webcams, responsáveis pela visualização e gravação do experimento e microfones que podem gravar o áudio no ambiente em que se efetua o experimento, neste trabalho utilizou-se redes de sensores sem fio MTS400 Crossbow e sensores em rede LonWorks.

Na Figura 2 visualiza-se a estrutura física do *weblab* para monitoramento no laboratório de abelhas.

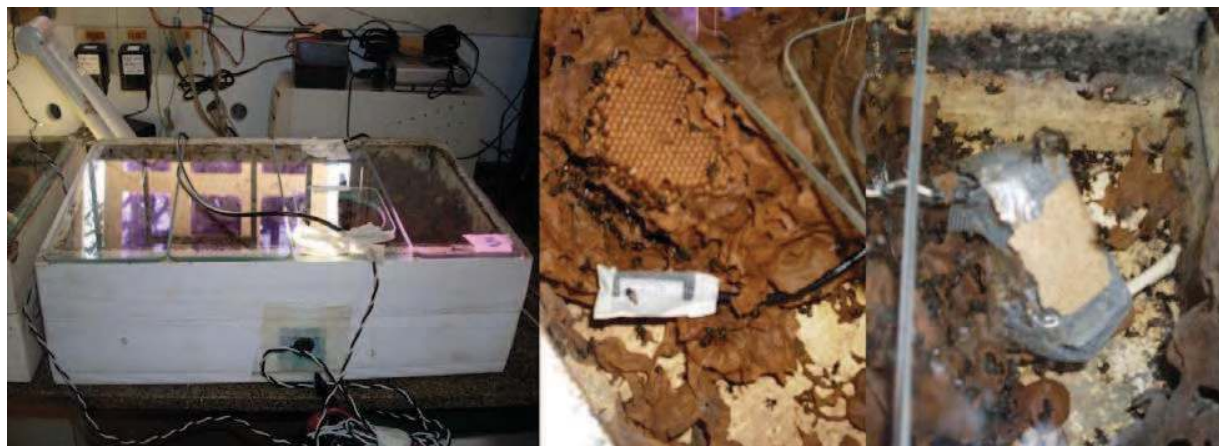


Figura 2 – Estrutura montada para monitoramento interno de colônias de abelhas

Os *weblabs* foram criados inicialmente com diversificadas tecnologias WEB. Projetos utilizando abordagens baseadas em objetos distribuídos foram propostas, como o R-LAB (HUA et al., 2003) que permite o acesso ao laboratório remoto por meio de navegadores, desde que estes sejam capazes de executar *applets* Java. Essas abordagens possuem em comum o limitado reuso de software e a baixa capacidade de customização, devido à baixa granularidade dos objetos distribuídos e a utilização de protocolos de interação comumente bloqueados por *firewalls*.

O projeto iLAB (ILAB, 2006) conduzido pelo MIT (*Massachusetts Institute of Technology*) foi o pioneiro em utilizar serviços WEB no domínio de *weblabs*. O iLAB desacopla as funcionalidades dos *weblabs* separando em pelo menos três classes de sessões, como apresentado na Figura 3.



Figura 3 - Modelo de arquitetura proposto pelo iLab (MIT,2006)

2.3. Bioabelha

O projeto Bioabelha é um *weblab* desenvolvido para estudo de polinizadores, neste *weblab* estão incluídos: uma base de dados de imagens de alta definição e vídeos para estudos taxonômicos e comportamentais em abelhas; um *weblab* de monitoramento em tempo real de uma colônia de abelhas, e também material para trabalhos de educação à distância, como pode ser visto na Figura 4.

Um banco de dados com padrões definidos de vídeos e imagens em alta definição de abelhas foi desenvolvido para ampliar a base de dados disponível no site Webbee (<http://www.webbee.org.br>). Este site contém informações sobre a biodiversidade de abelhas brasileiras em textos e imagens. Ele foi desenvolvido com o objetivo de ser utilizado em pesquisas e estudos sobre a biologia das abelhas, a sistemática e o comportamento, podendo ainda ser utilizado para cursos de educação a distância. Seu banco de dados original é restrito a imagens de baixa resolução, uma vez que é suposto ser acessado por uma variedade de usuários com diversos tipos de conexão à Internet.

O Webbee (WEBBEE, 2005), é um sistema de informações sobre polinizadores, mais especificamente sobre abelhas brasileiras nativas e foi criado justamente para organizar e centralizar as informações específicas sobre elas (SARAIVA et. al, 2003:78).



Figura 4 – Tela do *weblab* Bioabelha

2.4. Metadados

Metadados são dados que descrevem completamente os dados que representam, permitindo ao usuário decidir sobre a sua utilização da melhor forma possível (Almeida, 1999). São dados que permitem informar as pessoas sobre a existência de um conjunto de dados ligados as suas necessidades específicas, ou seja, dar informações sobre o dado. Segue se a relação com uma breve descrição dos metadados pesquisados, empregados em trabalhos de biodiversidade:

- EML (EML, 1997) – Utilizado para catalogar informações de dados ecológicos e ambientais, sua especificação é baseada em uma série de documentos XML podendo ser modular e extensível ao documento de dados ecológicos. Cada módulo EML é projetado para descrever uma parte lógica do metadados completo que deve ser incluído no conjunto de dados ecológicos. A arquitetura do EML foi desenvolvida para servir as necessidades de pesquisadores e comunidades sobre ecologia, beneficiando-se de trabalhos anteriores sobre metadados;
- DARWIN CORE (DWC, 2007) - foi projetado para facilitar a troca de informações sobre a ocorrência geográfica de organismos e da existência física de amostras em coleções bióticas. O padrão Darwin Core busca facilitar o intercambio de informações entre robustos bancos de dados distribuídos. Extensões para este padrão fornecem mecanismos para compartilhamento de informações adicionais, podendo estes dados serem informações específicas ou informações adicionais além do comum acordado ao âmbito do próprio Darwin Core;
- ABCD (ABCD, 2006) - é uma especificação de dados para unidades de coleção biológica, incluindo espécimes preservados e vivos. Destina-se inicialmente a apoiar o intercâmbio e a integração de coleções primárias de dados detalhados e dados de observação. A meta inicial da especificação de dados do projeto deveria ser abrangente e geral, para incluir uma ampla gama de conceitos que pudessem estar disponíveis em uma coleção de banco de dados, mas para uso apenas o mínimo de elementos necessários para tornar a especificação funcional. ABCD deliberadamente não abrange dados taxonômicos, tais como sinonímia, além do uso de nomes nas identificações. Da mesma forma, informações relacionadas a taxonomia, tais como áreas de distribuição, indicadores de valores, entre outros, também não estão incluídos. Estes elementos e conceitos são utilizados para fornecer uma maior compatibilidade com outras especificações no campo de coleções de dados biológicos tal como Darwin Core, EML entre outros
- PLINIAN CORE (PLINIANCORE, 2007) - Considerado um conjunto de conceitos que definem atributos básicos necessários para integrar e recuperar informações sobre espécies. Oferece também um suporte natural e estruturado ao relacionamento entre espécies. O esquema Plinian Core v.2.0 oferece um suporte natural e estruturado ao relacionamento entre espécies. É necessário verificar se ele atende aos requisitos dos principais serviços a serem oferecidos pela IABIN-PNT e suportados pela Webbee. Seu

desenvolvimento é coordenado pelo Instituto Nacional de Biodiversidade (INBio) da Costa Rica.

2.5. SEEK

O projeto SEEK (SEEK, 2004) (*Science Environment for Ecological Knowledge*, Ambiente Científico para Conhecimento Ecológico) é um sistema projetado não apenas para permitir a aquisição e o armazenamento de dados, mas também para prover a integração, transformação, análise e sintetização de dados de biodiversidade que até o momento de seu desenvolvimento eram considerados intratáveis. A infra-estrutura virtual SEEK abrange três camadas integradas; de acordo com a Figura 5.

- Sistema de Acesso a Informação – Camada capaz de permitir a integração de dados provenientes de fontes heterogêneas de informações que serão utilizadas para análises nas camadas superiores.
- Sistema de Mediação Semântica – Esta camada faz a mediação entre diferenças conceituais das diversas fontes de dados disponibilizadas pelo Ecogrid. Assim, dados provenientes de fontes diversas podem ser disponibilizados e agrupados, facilitando a construção de modelos para análises posteriores.
- Sistema de Análise e Modelagem – um ambiente visual automatizado (SEEK, 2004) onde ecologistas podem criar modificar e incorporar análises que compõem os fluxos de trabalho e novos modelos. Nesta camada a ferramenta utilizada como apoio para o pesquisador é o software Kepler (Altintas, 2004), uma interface visual para desenvolvimento de análises baseados em dados ecológicos, nas próximas seções serão comentados mais informações a respeito desta ferramenta.

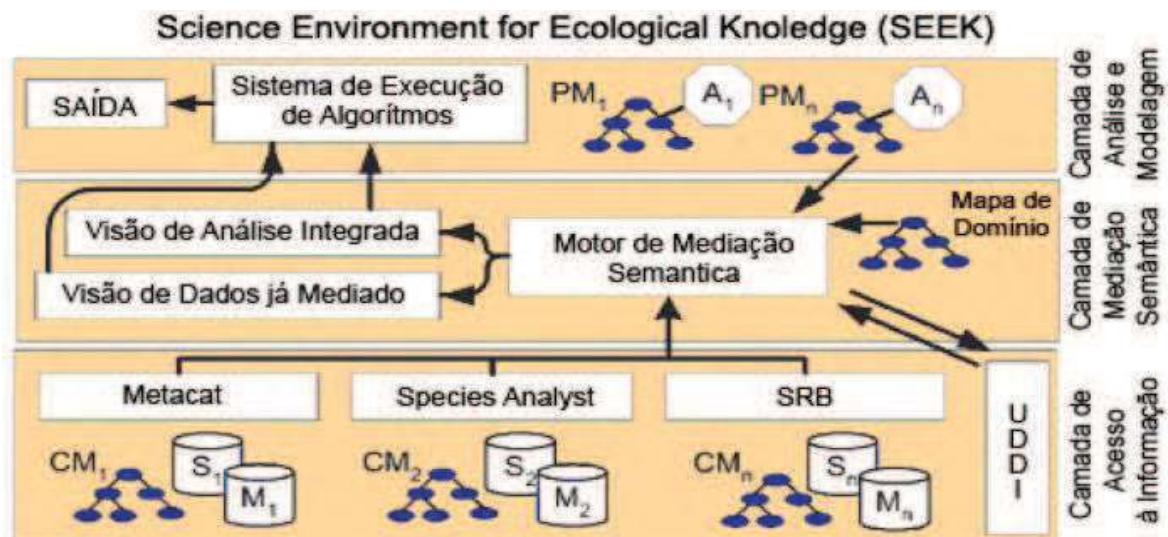


Figura 5 - Arquitetura em camadas do projeto SEEK (Ferreira, 2007)

2.6. Ferramentas de análises científicas

Pode se considerar que as ferramentas de análises científicas, proporcionam recursos que tem como objetivo facilitar o trabalho de elaboração e execução de análises especificadas pelo pesquisador. Atualmente existem conjuntos de ferramentas computacionais que permitem acelerar o processo de análise e descoberta. Nesta seção apresenta-se uma relação de algumas soluções de análises científicas utilizadas em diversos campos da ciência, o foco destas ferramentas encontra-se principalmente no processo e geração de resultados a partir de um conjunto de valores, dados coletados a partir do experimento realizado pelo pesquisador, estas informações são então, processadas e analisadas gerando os resultados finais do experimento em execução. Dentre estas ferramentas destaca-se:

2.6.1. Kepler

Segundo Ferreira (2007) apud (ALTINTAS, 2004) Kepler é uma aplicação baseada em Java que executa análises envolvendo coleções e dados ecológicos, permitindo a modelagem da sequência de processos utilizada em análises científicas.

Baseado no sistema Ptolemy (PTOLEMY, 1999), é uma plataforma robusta de apoio para vários modelos de suítes computacionais adequados para tipos distintos de análise (processamento de dados de sensor ou equações diferenciais integrais). Provê ao usuário uma interface gráfica e um ambiente de tempo de execução que pode executar *workflows* tanto de dentro do ambiente gráfico como a partir de linha de comando. Graças a sua flexibilidade ele pode ser utilizado em outras áreas do conhecimento além da Ecoinformática.

Em (KEPLER, 2006) são apresentados exemplos de outras áreas beneficiadas com o uso da ferramenta como, por exemplo, Biologia, Geologia, Oceanografia e Química. Na Figura 6 visualiza-se a ferramenta Kepler com um framework carregado em seu ambiente. *Workflows* desenvolvidos no Kepler podem ganhar desempenho com o poder computacional de tecnologias para processamento distribuído.

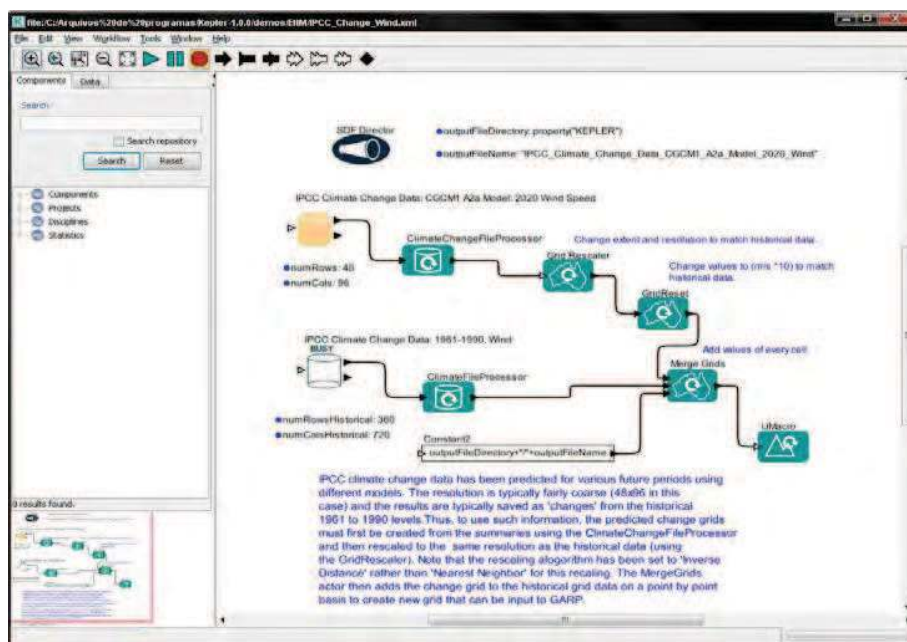


Figura 6 - Ambiente gráfico da ferramenta Kepler

2.6.2 VisTrails

É um sistema de gerenciamento de *workflow* científico (VISTRAILS, 2007) baseado em software-livre, foi desenvolvido pela Universidade de Utah. Fornece suporte para exploração e visualização de dados.

Considerando que os *workflows* são tradicionalmente utilizados para o processo de automatização de tarefas repetitivas, para aplicações que são por natureza de exploração, tais como simulações, análises de dados e visualização, pouco se repetem, neste caso, mudança é a norma.

A partir do momento que um cientista ou engenheiro gera e avalia os dados sobre as hipóteses em estudo, uma série de diferentes, porém relacionados, *workflows* são criados ao mesmo tempo em que um *workflow* é ajustado em um processo interativo. A ferramenta VisTrails tem como objetivo gerenciar estes *workflows* em rápida evolução. Na Figura 7 temos uma visão do ambiente da ferramenta VisTrails.

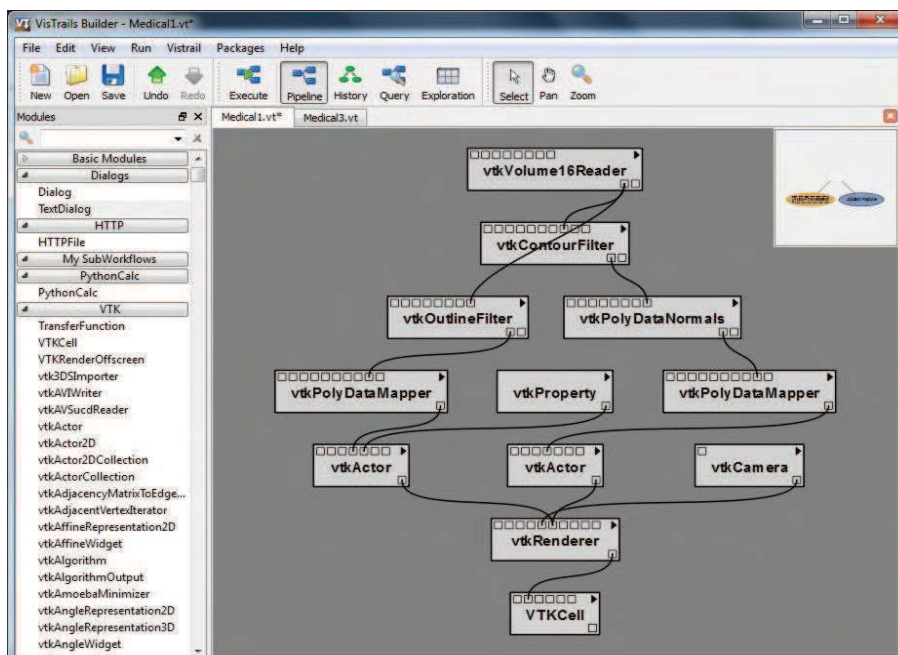


Figura 7 - Ambiente de modelagem do VisTrails

2.6.3. Taverna

O ambiente de gerenciamento de *workflow* Taverna, também é uma ferramenta baseada em software-livre, foi desenvolvida pelo projeto myGrid (GOBLE et al., 2007) como sendo a ferramenta responsável em realizar as análises por meio da execução de *workflows* científicos. O ambiente foi direcionado principalmente para a área de ciências biológicas, em que os cientistas não são especialistas em programação e serviços WEB (OINN et al., 2006).

A ferramenta permite a automatização dos métodos experimentais por meio da utilização de um número de diferentes serviços (tais como serviços da WEB) de um conjunto diversificado de domínios - a partir de biologia, química e medicina para a música, meteorologia e ciências sociais. Efetivamente permite que um cientista com pouco conhecimento em computação e poucos recursos técnicos, trabalhe para a construção de análises de alta complexidade por meio de dados públicos e privados e recursos computacionais, tudo a partir de um computador pessoal. A Figura 8 mostra a ambiente da ferramenta Taverna em ação durante o desenho de um fluxo de trabalho.

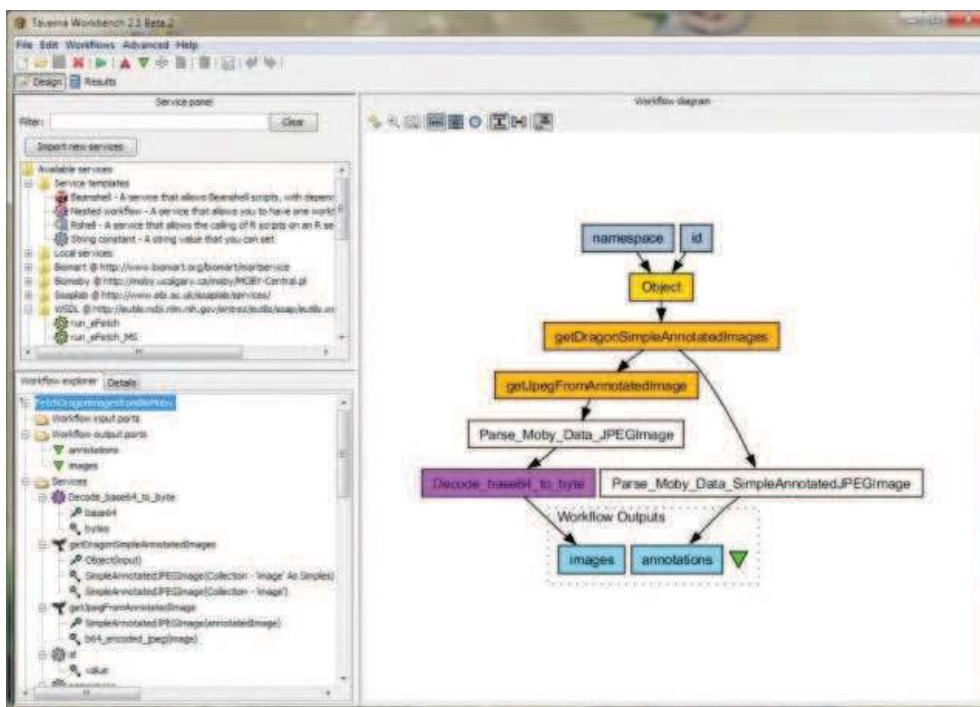


Figura 8 - Ambiente da ferramenta Taverna

3. Modelo de integração

Neste ponto é apresentado o modelo proposto para integração dos dados coletados por meio de laboratórios virtuais com ferramentas de análise. Esta proposta descreve uma arquitetura baseada em camadas ao qual são descritas todas as etapas seguidas da execução de um experimento, da coleta dos dados por meio de uma rede de sensores até o processamento e armazenamento dos resultados obtidos. Este modelo pode ser visualizado na Figura 9.

Na camada de aquisição de dados - Inicialmente têm se como foco o trabalho de estudo e coletas de dados sobre abelhas em suas colméias. Essa atividade engloba o monitoramento de temperatura e também de umidade relativa interna das colméias, essa atividade pode ser efetuada por meio de uma rede de sensores utilizando tecnologia LonWorks, muito utilizada em sistemas de controle predial, este modelo baseia se em nós interligados em rede, constituídos de sensores de reduzidas dimensões, aos quais se interligam por meio de cabos a um servidor, transferindo os dados coletados para uma rede TCP/IP. Seu modelo foi proposto em duas subcamadas denominadas:

- Módulo de redes de sensores - responsável pela conexão, controle, configuração e aquisição dos dados obtidos pelos sensores;
- Camada de acesso a informação – esta camada tem como finalidade a persistência dos dados brutos coletados pelos sensores de duas maneiras: na primeira, as informações são

armazenadas em um banco de dados da rede de sensores sem nenhum tratamento específico ou validação dos dados coletados; na segunda maneira tem como responsabilidade a persistência dos dados coletados na base de dados do *weblab*, neste momento as informações coletadas possuem tratamentos de validação e conversão de seu conteúdo de acordo com as especificações elaboradas no *weblab*. São utilizados também redes de sensores sem fio, que transmitem os dados por um módulo de comunicação sem fio até um gateway ligado a um computador, que é responsável por inserir os dados coletados por todos os nós sensores da rede na base de dados.

Na camada de *weblab* – têm-se o controle central do experimento e ainda, é nesta camada onde se gerencia a equipe responsável pela operação e manuseio dos equipamentos, pesquisadores envolvidos com o experimento, prazo de execução da coleta de dados, entre outras atividades. Suas subcamadas são descritas a seguir:

- Interface de Acesso WEB – apresenta aos usuários do *weblab* um portal de acesso aos recursos disponibilizados pelo módulo, desde que este esteja autorizado a operar a ferramenta;
- Regras de negócio – esta camada possui os algoritmos responsáveis pelo controle e funcionamento do *weblab* propriamente dito, e também a responsabilidade de, em conjunto com a camada de conexão Kepler, integrar os dados persistidos em seu banco de dados para posterior análise;
- Persistência a dados – garantir a persistência das informações disponibilizadas pelos pesquisadores e responsáveis pela operação do *weblab*.

Camada de conexão Kepler – a partir deste ponto o funcionamento do *weblab* e a coleta dos dados referentes ao experimento estão finalizados, existe a necessidade de proporcionar a equipe de pesquisadores, disponibilizarem essa massa de dados adquirida com a operação do *weblab* e da rede de sensores, sua atividade é dividida em duas subcamadas:

- Analisador de dados – ao ser executado recebe os dados por meio da camada de regras de negócio e efetua uma validação em seu conteúdo, com objetivo de garantir a integridade das informações que são disponibilizadas à ferramenta de análise;
- Gerador EML – após execução da validação das informações, pode-se gerar o arquivo a ser integrado na camada de análise. Esta etapa destina-se a criação de um arquivo XML que conterá as informações coletadas na execução do experimento, este arquivo deve estar de acordo com as especificações do padrão EML (*Ecological Metadata Language*) descrito na seção 2.4;

Camada SEEK – tem como objetivo permitir a aquisição, armazenamento e manipulação de dados para execução de análises, como descrito na seção 2.5.

Diante desta massa de dados coletadas na execução do experimento, é desenvolvido em conjunto com a equipe de pesquisadores, um *template* contendo modelos matemáticos

específicos para execução da análise destes dados coletados. Este *template* é inserido e executado dentro da ferramenta KEPLER que em conjunto com as informações coletadas pela execução do experimento e disponibilizadas a ferramenta, efetua as análises propostas pelos pesquisadores, seu resultado é adicionado a uma base de dados de análises científicas.

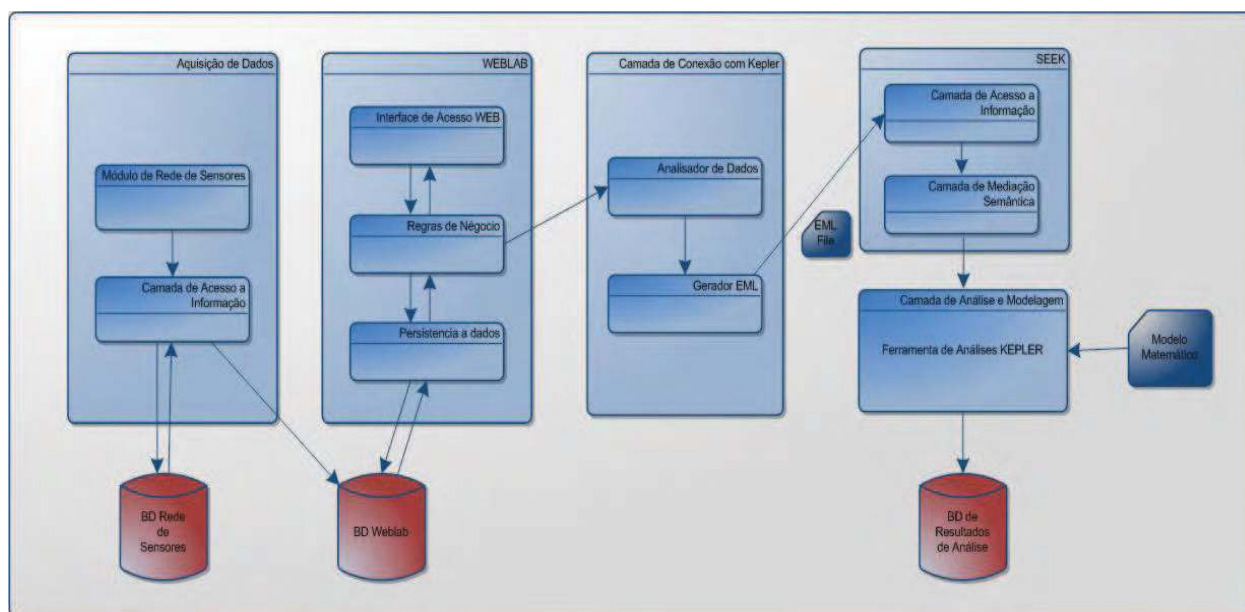


Figura 9 – Modelo empregado para integração de *weblabs* com ferramentas de análise

4. Resultados

No experimento realizado foi utilizada uma colônia de abelha sem ferrão da espécie *Scaptotrigona depilis* (Hymenoptera, Apidae, Meliponini), todas em caixas de madeira, padronizadas no formato de 50 x 30 x 14 cm (medidas externas), e 2 cm de espessura, cobertas por placas de vidro de 5 mm de espessura, (Figura 2). As caixas foram colocadas em uma sala no Laboratório de Abelhas, do Instituto de Biociências da USP, as quais se conectavam com o ambiente externo por meio de furos nas paredes, por onde mangueiras plásticas permitiam o fluxo das operárias forrageiras.

Por meio de furos nos vidros foram colocados dois sensores da rede LonWorks (região de cria e entrada/saída) e um nó sensor da RSSF (pote de alimento). Três sensores foram espalhados pela sala (dois sem fio e um da rede Lonworks) a fim de coleta dos dados do ambiente. Todos os sensores foram programados para coletar os dados a cada 10 minutos, e ficaram funcionando durante cinco dias ininterruptos.

Dados de temperatura e umidade foram os utilizados para a pesquisa, visto que os outros dados (pressão, luminosidade e aceleração) não eram relevantes neste momento.

4.1 Dados de temperatura

Analisando a localização das colméias, dois fatores influenciavam as condições internas: a temperatura interna da sala e a do ambiente externo, sendo esta última de importância significativa, devido a que o ambiente não apresenta nenhuma barreira fazendo fronteira, permitindo a passagem de ar, facilitando a troca de calor e umidade. Portanto, a comparação dos dados fica comprometida, conforme apresentado na Figura 2, a temperatura na entrada fica sempre menor que a da sala, esta não representando a temperatura ambiente.

Na Figura 10 também se pode observar que a temperatura da região de cria está sempre maior que a da sala, da entrada, e dos potes de alimento, sugerindo que *S. depilis* realiza a termorregulação dessa região (Engels, 1995).

Os dados da entrada da colméia representam um novo dado na literatura deste tipo de estudo, representando um bom indicativo da ventilação e estudo do controle da temperatura. Porém, devido à localização da colméia dentro da sala, fica impossibilitado de fazer a comparação destes dados com a temperatura da sala. Podemos inferir que a temperatura do ambiente externo estava mais fria que a da sala através destes dados.

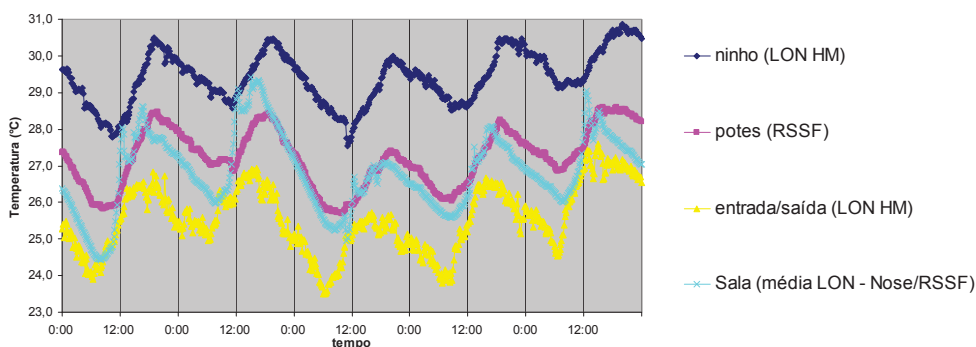


Figura 10. Temperatura a cada 10 minutos nos diferentes locais da colméia

Como nos demais trabalhos com termorregulação nas abelhas sem ferrão, as maiores temperaturas encontradas foram na região das células de cria, assim como a maior média ($29,4\text{ }^{\circ}\text{C} \pm 0,73$) (Tabela 1), ligeiramente inferior à temperatura encontrada por Engels *et al.* (1995) trabalhando com uma espécie do mesmo gênero, *Scaptotrigona postica*, que encontrou uma média de $32 \pm 3\text{ }^{\circ}\text{C}$ nesta região.

Tabela 1. Temperatura e umidade relativa (RH) média nos locais da colméia.

	Ninho	Potes	Entrada/Saída	Sala
Temperaturas médias ($^{\circ}\text{C}$)	29.4 (SD 0.73)	27.2 (SD 0.8)	25.6 (SD 0.91)	26.8 (SD 1.06)
RH médias (%)	80 (SD 1.5)	72 (SD 2.2)	64 (SD 5.3)	54 (SD 5.6)

4.2 Dados de umidade relativa

O acompanhamento constante da umidade relativa (RH) em várias partes de uma colônia representa um avanço pioneiro e muito significativo nas pesquisas com meliponíneos (Nicolson, 2009), uma vez que não são encontrados dados deste tipo na literatura. Estima-se que as abelhas não coletam água diretamente, mas que a umidade presente no ninho provenha da desidratação do néctar e do metabolismo das abelhas (Nogueira-Neto, 1997).

Como se pode observar na Figura 11, a RH na região da cria é mantido maior e mais constante que nas outras partes estudadas. Isto pode ser o resultado de uma manutenção da RH ativa das operárias e das estruturas do ninho, como o invólucro, ou então apenas um subproduto da presença deste invólucro para outros fins.

As operárias do meliponíneos realizam a ventilação do ninho por meio do batimento das asas na entrada, para controlar as condições climáticas internas (Nogueira-Neto, 1997; Nicolson, 2009). Ainda não existem estudos relacionando isto com a umidade, mas observando a Figura 11, supõe-se que a alta variação da RH na entrada da colônia poderia indicar a ventilação, podendo-se também fazer inferências sobre este comportamento em trabalhos futuros.

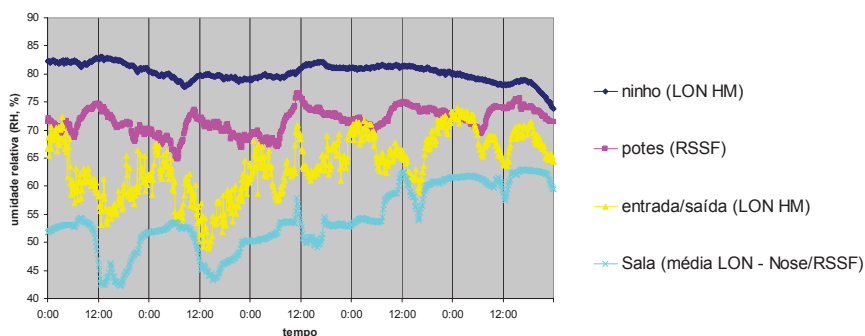


Figura 11. Umidade relativa (RH) nos diferentes locais da colmeia

5. Conclusões

Este artigo apresentou os requisitos e a arquitetura de integração de dados coletados com experimentos executados em ambientes remotos com a ferramenta de análise Kepler, o modelo proposto, porém, não se limita a uma ferramenta específica, sua arquitetura pode ser adaptada para disponibilizar camadas específicas de integração com outras ferramentas de análise, podendo assim beneficiar o pesquisador com o uso de ambientes computacionais distribuídos (*grids*), ou mesmo efetuar análises comparativas, avaliando as discrepâncias – se ocorrerem – entre as ferramentas.

Assim, o ambiente apresentado pretende contribuir com as lacunas apresentadas, oferecendo um ambiente de apoio à utilização de cientistas e pesquisadores em elementos de biodiversidade.

A validação da proposta está sendo acompanhada pela ajuda de especialistas na área, os pesquisadores do laboratório de Biologia da USP e do laboratório de abelhas da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo.

6. Referências

ABCD - **ABCD Schema 2.06 ratified TDWG Standard**. Disponível em <<http://bgbm3.bgbm.fu-berlin.de/TDWG/CODATA/Schema/default.htm>>. Acesso em 12 dez. 2009.

ALMEIDA, L.F.B. **A Metodologia de Disseminação de Informação Geográfica e os Metadados**. Tese de Doutorado. Centro de Ciências Matemáticas e da Natureza – UFRJ, 1999.

DWC – **DarwinCore Group**. Disponível em <<http://www.tdwg.org/activities/darwincore/>>. Acesso em 10 dez. 2009.

EML – **Ecological Metadata Language**. Disponível em <<http://knb.ecoinformatics.org/software/eml>>. Acesso em 12 dez. 2009.

ENGELS W. et al., **Thermoregulation in the nest of the Neotropical stingless bee *Scaptotrigona postica* and a hypothesis on the evolution of temperature homeostasis in highly eusocial bees**. Stud. Neotrop. Fauna Environ. 1995. p. 193–205.

FERREIRA, M.S.J. **Uma arquitetura de sistemas distribuídos para weblabs de serviços ambientais**. Dissertação de Mestrado. Escola Politécnica da Universidade de São Paulo 2007.

GIANNINI, T.C. et al. **Webeelabs: weblabs on Brazilian bees**. Proceedings of the 5th International Conference on Ecological Informatics ISEI5, 2006.

GOBLE, C. et al. **myExperiment: social networking for workflow-using e-scientists**. In: WORKS, Monterey, California, USA. Jun. 2007 p.1-2.

HUA, J. et al. **Web Enabled Laboratory (R-LAB) Framework**. In: ASEE/IEEE Frontiers in Education Conference, Boulder, USA. 2003.

ILAB – **iLabs: Internet access to real labs – anywhere, anytime**. Disponível em <<http://lcampus.mit.edu/iLabs/>>. Acesso em 4 jan. 2010.

KEPLER – **Kepler: An Extensible System for Scientific Workflows**. Disponível em <<http://kepler.ecoinformatics.org>>. Acesso em 17 dez. 2009.

MYEXPERIMENT – **The myExperiment Virtual Research Environment**. Disponível em <<http://www.myexperiment.org>>. Acesso em 18 dez. 2009.

NICOLSON, S. W., **Water homeostasis in bees, with the emphasis on sociality**. The Journal of Experimental Biology 212, 2009. p. 429-434.

NOGUEIRA-NETO, **Vida e criação das abelhas indígenas sem ferrão**. 1997. 442p.

OINN, T. et al., Taverna: **Lessons in creating a workflow environment for the Experience life sciences**. Concurrency Computat. Pract. Exper., vol. 18. 2006 p. 1067-1100.

PLINIANCORE – **Plinian Core Project**. Disponível em <<http://www.pliniancore.org/en/inicio.htm>>. Acesso em 10 dez. 2009.

PTOLEMY – **The Ptolemy Project**. Disponível em <<http://ptolemy.eecs.berkeley.edu>>. Acesso em 17 dez. 2009.

SARAIVA et al. **WebBee – a Web-based Information Network on Bees** – Revista de Engenharia de Computação e Sistemas Digitais, n.1, v.1, nov. 2003 p.77-86.

SEEK – **Science Environment for Ecological Knowledge**. Disponível em <<http://seek.ecoinformatics.org>>. Acesso em 22 dez. 2009.

VISTRAILS – **Open Source Scientific Workflow**. Disponível em <<http://vistrails.org/>>. Acesso em 4 jan. 2010.

WEBBEE – **Uma rede de informações sobre diversidade brasileira em abelhas**. Disponível em <<http://www.webbee.org.br>>. Acesso em 18 dez. 2009.

ZUBIA, J.G. et al. **Suitability and Implementation of a WebLab in Engineering**. In: IEEE Conference on Emerging Technologies and Factory Automation. 2005. p. 49-56.

_____, **Towards a Canonical Software Architecture for Multi-Device WebLabs**. In: 31st Conference of IEEE. nov. 2005. p. 2146-2151.