

# PROPOSTA DE PROJETO

## SICOL

## SISTEMA DE INFORMAÇÃO PARA

## COLEÇÕES DE INTERESSE BIOTECNOLÓGICO

### 1. Objetivo

É o objetivo geral do **Programa Nacional de Biotecnologia e Recursos Genéticos** do Ministério da Ciência e Tecnologia elevar o nível de competitividade científica e tecnológica do país e patamar e equiparar a nível dos países desenvolvidos, acelerando os mecanismos de transferência dos conhecimentos e de tecnologias gerados pelas pesquisas ao setor produtivo de bens e serviços, com vistas à inovação e melhoria de produtos, processos e serviços biotecnológicos.

Uma das ações previstas neste Programa é a ampliação e o fortalecimento da infraestrutura de suporte à biotecnologia, incluindo as coleções de culturas e bancos de germoplasma, notadamente as de serviços de referência. Com um dos objetivos desta ação, encontra-se a implementação de um sistema de informação *online*.

Como objetivo específico, tem-se a estruturação de um sistema de informação que:

- sirva de elemento integrador às diversas e diferenciadas coleções (vegetal, animal, microrganismo e de células humanas)
- atenda à demanda por informação das coleções
- atenda à demanda por informação dos usuários de insumos biológicos na área de biotecnologia
- atenda à demanda por informação dos formuladores de políticas públicas

### 2. Introdução

Toda estratégia de informação depende da demanda do usuário e da disponibilidade do provedor de informação. O sistema de informação, objeto desse projeto, deverá servir de "ponte", auxiliando o provedor de informação a transformar o seu dado ou informação em



um produto acessível, útil e "utilizável" para o usuário final. Esquematicamente temos:

Considera-se como sendo o público -alvo de um sistema de informação, tanto o provedor, quanto o usuário da informação. A temática central dos sistemas de informação propostos são as coleções de interesse biotecnológico. O termo *coleções*, neste projeto, se restringe às instituições que fornecem infraestrutura básica para biotecnologia. São incluídas as



coleções"deserviço"e"dereferência" demicrorganismos,germoplasma(animalevegetale microrganismo),tecido celular(animalevegetal),DNAeplasmídeos.

São considerados "público -alvo" dos sistemas de informação o raproposto:

- as coleções (deserviço e referência);
- os usuários dos insumos biológicos dos serviços das coleções; e,
- os formuladores de política.

Definido o público alvo (tanto os provedores como os usuários de informação) é necessário definir uma estratégia de atuação visando a integração da "comunidade de coleções" e a promoção de sua interação com a comunidade de usuários de seus serviços e produtos e com os formuladores de políticas públicas.

Será necessário atuar nas seguintes linhas:

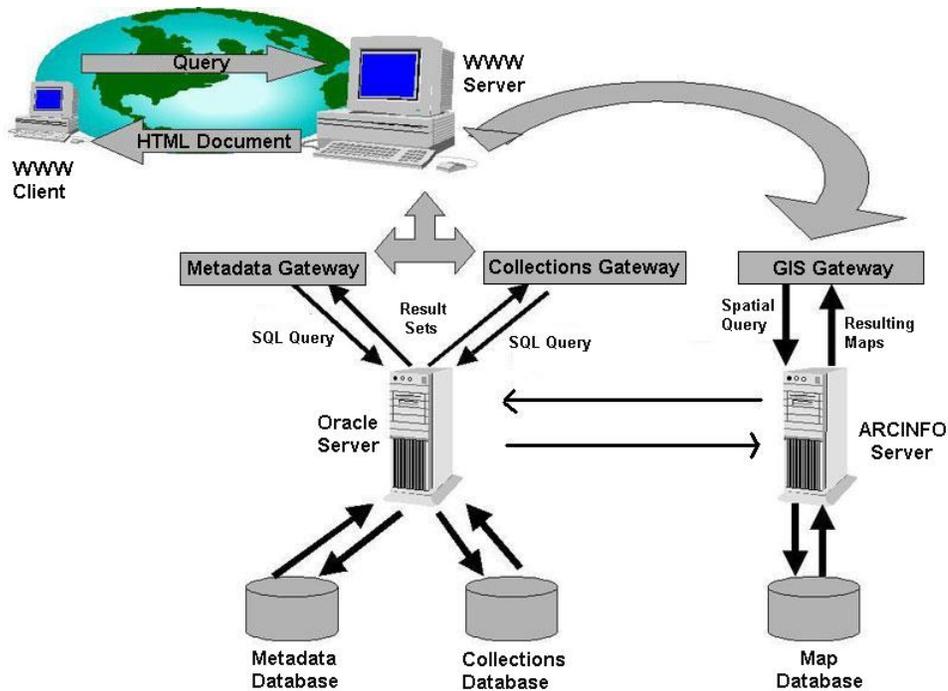
- na estruturação de um sistema de informação como elemento integrador
- no atendimento à demanda por informação dos usuários de insumos biológicos
- no atendimento à demanda por informação das próprias coleções
- no atendimento à demanda por informação dos formuladores de políticas públicas

Assim, será necessário realizar um diagnóstico para determinar qual a informação requerida por cada usuário -alvo e qual a informação existente nas coleções. Para a definição da arquitetura do sistema de informação, é necessário conhecer como o dado/informação está disponível, como está armazenado (em que sistema e formato) e as condições de conectividade da instituição à Internet.

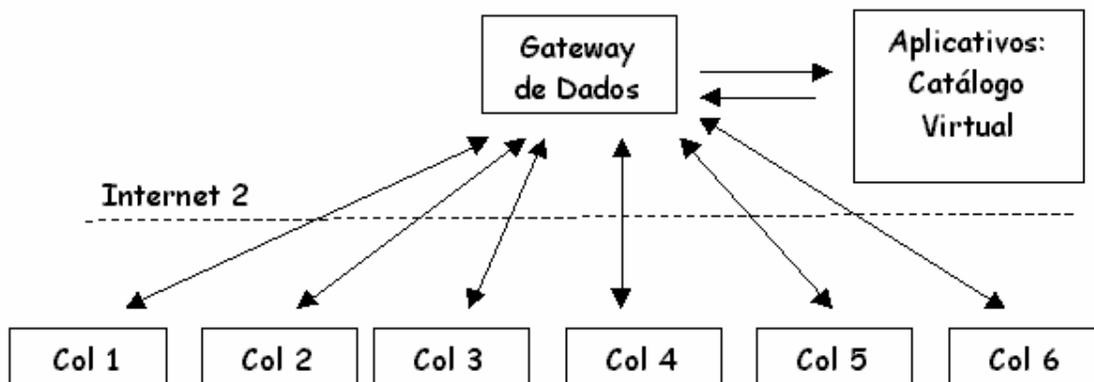
Com este diagnóstico pode-se definir a arquitetura da rede. De maneira simplificada podemos indicar que a rede de dados/informações pode ser centralizada, distribuída ou mista.

De forma simplificada, podemos dizer que em uma rede centralizada, todos os dados/informações são armazenados em um único lugar. É o sistema mais indicado quando os provedores de informação (nesse caso, as coleções) não têm capacidade computacional instalada e quando a linha de dados (conectividade à Internet) é muito lenta e/ou de baixa qualidade. Um exemplo de sistema centralizado é o sistema de informação ambiental que está sendo desenvolvido pelo programa Biota/Fapesp: O Instituto Virtual da Biodiversidade ([www.biotasp.org.br/sia](http://www.biotasp.org.br/sia)). Nesse caso, temos como provedores de informação a comunidade científica, pesquisadores que estão levantando dados sobre inventários de espécies do Estado de São Paulo. Não se pode exigir a cada pesquisador a estruturação e manutenção de um sistema de informação que atenda aos requisitos do programa, o que seria necessário para o caso de uma arquitetura distribuída. Cada pesquisador teria que ter em um servidor conectado a uma linha de dados de qualidade e garantir a manutenção deste serviço 24 horas por dia, 365 dias por ano, o que não faria sentido. Nesse caso, foi desenvolvido um sistema de alimentação remota de dados via Internet, dando ao pesquisador total domínio sobre os seus dados. Cada pesquisador tem seu *username* e senha para a entrada de seus dados, e somente ele pode modificá-los. Outro fator importante foi o fato de esta comunidade científica ter estabelecido uma ficha padrão de entrada de dados. Assim, temos uma ficha de entrada de dados com campos obrigatórios e vocabulário controlado. Trata-se do "metadado" da lista de espécies.

Para o usuário do sistema de informação tudo se passa de uma maneira transparente. Esquemáticamente temos:



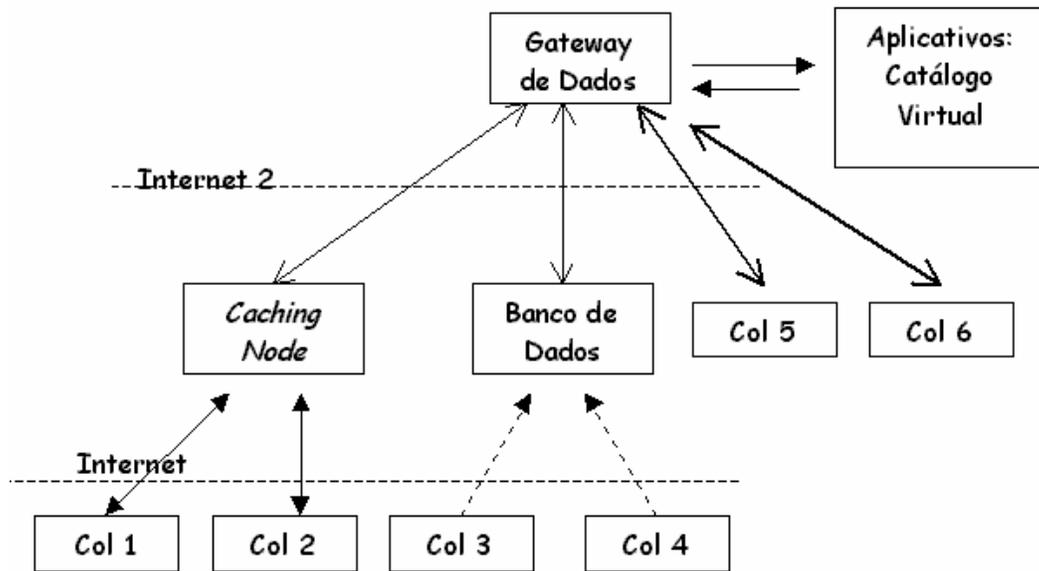
Nos casos de sistemas de informação e das coleções, existem algumas diferenças básicas. Trata-se de informações institucionais cuja padronização ainda está por ser determinada. Nesse contexto, faz sentido pensarem instituições como os responsáveis pela manutenção de seus próprios sistemas de informação. Assim, vale a pena estudar a possibilidade de implantação de um sistema de informação distribuído. Esquematicamente temos:



Este seria o esquema ideal. Uma situação em que todas as coleções estivessem conectadas à Internet 2, com capacidade de computacional instalada e todo o acervo informatizado. Pensando no "catálogo virtual" como aplicativo, seria criada uma página padrão no servidor central (gateway) que seria utilizada pelo usuário para fazer buscas. Ao receber uma pergunta do usuário, o gateway dispararia processos capazes de realizar a busca pela informação, localizar nas diversas coleções, agregando-as, num formato padrão, num relatório final único, e devolvendo-o ao usuário requisitante. Infelizmente, a

situaçãoprovavelmenteserádiferente.Imaginamosencontrarcoleçõesinformatizadascom excelenteconectividade,masencontraremostambémcoleçõesinformatizadasconectadas abaixavelocidadeeoutrasaindaquenãotêmseusistemasdeinformaçãoeestruturados *online*.Nestecaso,arquiteturadarededeveátenderatodassituações:

- acessoemtemporealàscoleçõesinformatizadas,comsistemas *online*eexcelente conectividade;
- estruturaçãode *cache*nodes (nóintermediários),espéciesde *mirrors*,queirão automaticamenteatualizaros dadosdascoleçõesemumcomputadorcom conectividadealtavelocidade;e,porfim,
- aalimentaçãodosdadosdecoleções *off-line*emumsistema *online*com



conectividadealtavelocidade.

Esquematicamentetemos:

Adefiniçãodaarquiteturasomenteseda ráapósodiagnósticodasituaçãodascoleções queintegrarãosistema.Resumidamentetemos:

- Dependendodo"grautecnológico"dascoleçõesosistemapoderáserdistribuído, centralizadooumisto.Qualquerquesejao"desenho"dosistemaéfundamentalque oprovedordosdados/informação( *datacustodian* )tenhaintegralresponsabilidadeedomíniosobreosseusdados/informação.
- Osistemadeinformaçãodevefacilitar/viabilizaroacessoaeintegraçãodos dados/informações.
- Osistemadeinformaçãodeverádesenvolveraplicativosdeinteresseparaosseus usuários

### 3. Etapas de Trabalho

#### A. Identificar as coleções a serem envolvidas

A primeira etapa de trabalho será identificar as coleções que deverão ser avaliadas para integrar o sistema de informação a ser criado. Temos já identificadas as seguintes coleções:

- EMBRAPA-CENARGEN: Brasília, DF
- FIOCRUZ: Rio de Janeiro, RJ e Salvador, BA
- UFRJ: Rio de Janeiro, RJ
- CCT - Fundação André Tosello: Campinas, SP
- Instituto Biológico: Campinas, SP
- IAC - Bancos de Germoplasma: Campinas, SP
- Instituto Evandro Chagas: Belém, PA
- Instituto Adolpho Lutz: São Paulo, SP

A definição das coleções participantes será realizada em comum acordo como MCT.

#### B. Diagnóstico

Para avaliar a arquitetura do sistema de informação a ser criado e o seu conteúdo básico, é necessário realizar um diagnóstico, *in loco*, das coleções. É necessário avaliar cada coleção com o provedor de informação, levantando dados a respeito da:

- organização da informação (conteúdo, campos obrigatórios, campos mínimos, vocabulário controlado, padrões adotados, metadados, etc.)
- informatização do acervo (*software* utilizado, formatos dos dados, etc.)
- conectividade (qualidade e velocidade de linha de dados)

Além da informação básica do acervo da coleção, é necessário levantar a demanda por informação de todos os usuários - alvo, ou seja, das próprias coleções, dos usuários de insumos biológicos e dos formuladores de política (*politymakers*). Uma vez identificada qual é a demanda por informação, será necessário identificar possíveis parceiros que possam ser integrados ao SICol.

#### C. Estudo da Arquitetura do Sistema

Com base no diagnóstico, uma vez definido o modelo do sistema de informação a ser desenvolvido, principalmente no tocante ao Catálogo Virtual (distribuído, centralizado, misto), será necessário realizar uma série de estudos, principalmente no tocante a protocolos de comunicação e de interoperabilidade (Z39.50, XML, etc.) e metadados.

Será fundamental também o desenvolvimento de *guidelines* para as coleções, orientando-as a respeito de sua forma de participação na rede. Provavelmente será necessário definir padrões para garantir a participação das 3 "categorias" de coleções:

- coleções informatizadas, com sistemas *online* e excelente conectividade: definição dos padrões para viabilizar o sistema de acesso aos dados em tempo real
- coleções informatizadas, com sistemas *online*, mas com conectividade de baixa qualidade e/ou velocidade: definição dos padrões visando a estruturação de *cache node* ou *mirrors*, que serão automaticamente atualizados com os dados das coleções
- coleções informatizadas, sem sistemas *online*, com conectividade de baixa qualidade: a definição dos padrões visando a alimentação dos dados de coleções *off-line*.

#### D. Desenvolvimento de um *website* para o SICol

Independentemente da conclusão do diagnóstico e do desenvolvimento da arquitetura do sistema de informação, é fundamental ainda no início do projeto, estruturar um *website* para o sistema de informação a ser criado. Se compreendermos o sistema de informação como elemento integrador, devemos procurar introduzir esta ferramenta o mais rapidamente possível, independentemente do seu caráter preliminar. Tem -se como atividades iniciais:

- criação do *website* com informações relevantes de interesse comum para o setor de coleções microbiológicas e de interesse biotecnológico
- criação de uma lista ou fórum de discussão
- criação de cadastros/diretórios (quem é quem - especialistas, instituições)
- criação de bancos de dados comuns ou o estabelecimento de *links* que dispõem de informações mais genéricas a respeito de: legislação, publicações, serviços, etc.

A estrutura do *website* dependerá do conteúdo a ser incluído, mas, em termos de tecnologia, deverá ser desenvolvido utilizando-se o método de criação de páginas dinamicamente.

#### 4. Prazo e Cronograma de Execução

O projeto terá duração de 06 (seis) meses, com acompanhamento contínuo da Coordenação Geral de Biotecnologia da Secretária de Política e Programas do MCT. O cronograma, com etapas do trabalho a serem discriminadas abaixo:

Etapas de Trabalho	Prazo de Execução (mês)					
	1	2	3	4	5	6
Identificação das Coleções Participantes	x					
Diagnóstico	x	X				
Estudo da Arquitetura			x	x	x	
Desenvolvimento do <i>Website</i>	x	X	x	x	x	x
Relatórios Técnicos			x			x

#### 5. Produtos Esperados

Após um período de 6 (seis) meses de trabalho, espera-se obter os seguintes produtos:

- um diagnóstico a respeito do estado da arte das coleções selecionadas para participar do levantamento, notadamente o seu sistema de informação e conectividade;
- um diagnóstico a respeito da demanda por informação do público - alvo do SICol;
- um estudo da arquitetura do sistema de informação a ser desenvolvido principalmente no catálogo virtual;
- *guidelines* para as coleções com orientações a respeito da melhor forma de participação na rede

- o website para o SICol
- a criação de uma lista ou fórum de discussão
- a criação de cadastro e/ou diretórios de especialistas, instituições

## **6. Responsável Técnico**

O responsável técnico e representante legal perante o Ministério da Ciência e Tecnologia é o Dr. Vanderlei Peres Canhos.

## **7. Desenvolvimento Futuro**

Uma vez definida a arquitetura, será necessário desenvolver o sistema, testá-lo, torná-lo disponível na rede e divulgá-lo. Através do diagnóstico será possível também definir quais são os produtos (aplicativos) a serem desenvolvidos. Acreditamos que alguns temas poderão ser:

- O Catálogo Virtual das Coleções; e
- O Catálogo de Serviços