

rede de informação colaborativa sobre a biodiversidade brasileira¹

splink.cria.org.br

breve histórico

O desenvolvimento da rede *speciesLink* teve início em 2001 no contexto do programa Biota-Fapesp. Foi oficialmente lançada *on-line* em outubro de 2002 com dados de uma única coleção biológica. Até 2005 havia somente a participação de coleções biológicas do estado de São Paulo, mas cresceu ao longo dos anos, com apoio do CNPq, Finep e outras agências do país e do exterior, contando hoje com a participação de pelo menos uma coleção biológica em cada estado da união².

Graças à elaboração do documento *Diretrizes e estratégias para a modernização de coleções biológicas brasileiras e a consolidação de sistemas integrados de informação sobre biodiversidade*³ e a várias ações estaduais e programas federais, muitos decorrentes desse documento, como o PPBio, ProTaxa e, mais recentemente o INCT, a rede evoluiu⁴ e se consolidou.



Ao longo dos seus 17 anos de existência, essa *e*-infraestrutura tem sido mantida com recursos de projetos, o que tem se mostrado cada vez mais difícil e complexo, ironicamente, graças ao seu êxito. O número cada vez maior de parceiros e consequente manutenção de mais dados, exige uma rotina expressiva, que inclui manutenção e atualização de *hardware* e *software*, *backup*, suporte e *helpdesk* a provedores e usuários dos dados. Essa rotina não se enquadra nas políticas de apoio a projetos, voltados

especificamente à pesquisa e desenvolvimento de novos sistemas e não há políticas de apoio de longo prazo para a manutenção de *e*-infraestruturas de interesse público.

termo aditivo MCTIC – RNP

No final de 2018 foi dado um importante passo para assegurar a permanência dessa e-infraestrutura com a celebração do 20º. termo aditivo ao Contrato de Gestão celebrado entre o MCTIC e a RNP. Esse termo aditivo assegura o repasse de quinhentos mil reais à RNP para apoiar o CRIA na migração do sistema *speciesLink* para a infraestrutura em nuvem da RNP, incluindo suporte e manutenção durante um período de 12 meses. Todos os sistemas públicos de informação, dados, ferramentas e aplicativos foram transferidos para o Centro de Dados Compartilhados da RNP em Recife e estão em operação desde abril de 2019. Portanto os sistemas agora estão em um ambiente seguro com *hardware* compatível com as suas necessidades.

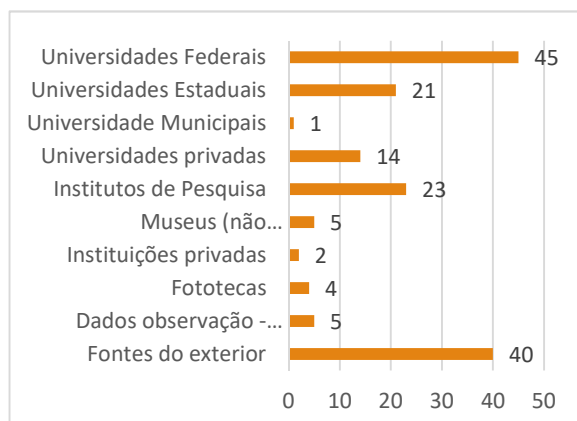
parceiros

É fundamental importância a rede de pesquisadores que foi criada graças à organização e compartilhamento de dados abertos *online* e ao financiamento recebido para aumentar o seu conteúdo, desenvolver novas ferramentas e aplicativos e promover a articulação da comunidade de interesse.

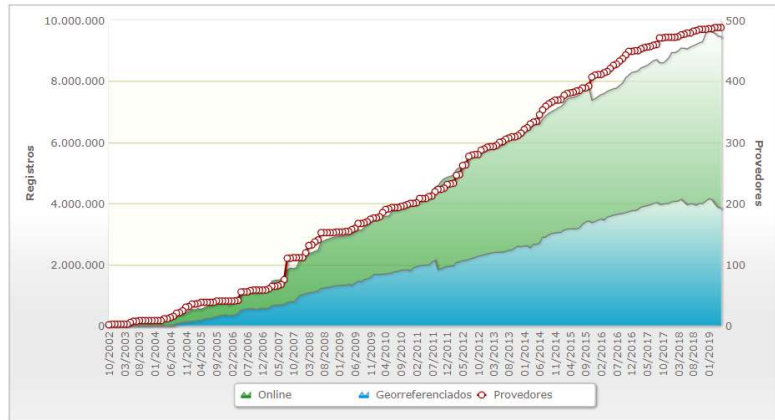
Importantes parceiros da rede são

- o CRIA, responsável pelo desenvolvimento e manutenção da e-infraestrutura de dados e ferramentas de acesso público e aberto e pelo apoio aos provedores de dados
- a Rede Nacional de Ensino e Pesquisa, RNP, responsável por
 - abrigar os sistemas da rede em seu Centro de Compartilhamento de dados, CDC em Recife
 - inclusão do CRIA na RedeComep de Campinas, o que possibilita a gestão remota dos sistemas
 - viabilizar a participação das universidades e instituições públicas do país, tanto como provedores como usuários dos dados, através da rede Ipê e demais redes metropolitanas do país
- os provedores de dados brasileiros. A rede hoje integra 489 conjuntos de dados⁵ provenientes de 81 universidades; 23 Institutos de pesquisas, 5 museus (não universitários), 2 instituições privadas, 4 fototecas (pesquisadores) e 5 bases de dados de observação de projetos.
- os provedores de dados do exterior. 16% dos dados da rede *speciesLink* são repatriados de 39 coleções de 6 países e de um banco de dados sobre a biodiversidade da Amazônia Colombiana.

É importante observar a participação expressiva das universidades na rede. Além do compartilhamento de dados, elas cumprem um papel relevante na pesquisa e na formação de recursos humanos, tanto no uso das e-infraestruturas como também no compartilhamento aberto de dados e conhecimento. Hoje, 95% dos herbários que compartilham seus dados através o INCT – Herbário Virtual da Flora e dos Fungos estão associados a programas de pós-graduação.



A participação de várias instituições também se dá através de redes temáticas e geográficas. No início foram estruturadas redes estaduais como Biota/Fapesp, TaxonLine (PR) e as redes Capixaba e Mineira. Também foram estruturadas redes temáticas como o INCT – Herbário Virtual da Flora e dos Fungos⁶, a rede SiCol (coleções de interesse biotecnológico)⁷, a rede de polinizadores, agora recebendo apoio para acervos de abelhas, e acervos históricos como a *Flora Brasiliensis*⁸ (parceria com a Unicamp e o Jardim Botânico de Missouri) e os Herbários Virtuais de Saint-Hilaire⁹ (parceria com o Museu Nacional de História Natural de Paris e o Instituto de Botânica de São Paulo) e Glaziou¹⁰ (parceria com o Museu de Paris e o Museu Nacional).



São também parceiras algumas e-infraestruturas internacionais, como *Global Biodiversity Information Facility – GBIF*; *Integrated Digitized Biocollections – iDigBio*, EUA; *Comisión Nacional para el Conocimiento y uso de la Biodiversidad – Conabio*, México; *species2000* (Catálogo da Vida); e, *Group on Earth Observations – Biodiversity Observation Network - GeoBon*, entre muitos outros.

conteúdo disponível on-line

Hoje são disponibilizados mais de **9,5 milhões** de registros associados a espécimes depositados em coleções biológicas, ou a observações feitas em campo, provenientes de **489 conjuntos** de dados. São documentados dados de mais de **128 mil espécies**. Pouco mais de **7,6 milhões** desses registros têm **coordenadas geográficas** e pouco mais de **218 mil** são referentes a **tipos nomenclaturais**, de grande importância para a ciência. A rede disponibiliza ainda mais de **2,2 milhões de imagens** desse material, muitas em alta resolução. Esse conjunto de dados também contém mais de **100 mil registros de espécies ameaçadas** de extinção.



A cada ano, a rede *speciesLink* cresce de forma expressiva, tanto em relação ao número de provedores de dados, em seu conteúdo, registros textuais e imagens, e no seu uso. Em 2018 foram integrados os dados dos acervos de **18 novas coleções** além da atualização dos dados dos provedores existentes. Foram integrados à rede mais de **750 mil novos registros** e mais de **290 mil novas imagens**, o que significa uma média de mais de **2 mil novos registros** e **cerca de 800 novas imagens** inseridas na rede **por dia**. O índice de atualização em maio de 2019 é de 60%, indicando que 60% das coleções atualizaram seus dados ao menos uma vez nos últimos 12 meses, sendo que várias atualizam seus dados semanalmente. Poderíamos elevar esses índices se houvesse um maior apoio às coleções e recursos adequados ao CRIA para ampliar o atendimento, lembrando que mais de **16 milhões de registros** das coleções participantes ainda **não estão online**.

serviços às coleções

speciesLink

dataCleaning
english

Essa ferramenta tem a finalidade de identificar possíveis erros e facilitar a padronização dos dados. Geralmente são erros pequenos e simples, onde uma letra é adicionada, trocada ou omitida. São erros difíceis de serem encontrados em um grande conjunto de dados pela pequena diferença da ortografia. Uma premissa do trabalho é não interferir na informação. Nenhum dado é modificado, o sistema apenas indica os registros "suspeitos", recomendando ao autor que confira e escolha entre corrigir ou não a informação. A ferramenta ainda está em desenvolvimento e todas as sugestões serão bem vindas.

Selecione uma coleção:

coleção: ALCB	
total de registros indexados	130880
- sem coordenadas geográficas	36191
- georreferenciados	94689
- georreferenciados bloqueados	0
- no mar	4001
- número do catálogo em branco	6072
	[susp]
registros repetidos	
número do catálogo	23354
registros duplicados	4026
nome + número do coletor	12523
última atualização - logs de erro	
da coleção: 13-05-2019	do dataCleaning: 14-05-2019



perfil da coleção
estatísticas do dataCleaning
análise da dispersão das coordenadas geográficas

dados taxonômicos	
inventário	nome científico - coletor - determinador - tipos
reino	não encontrado
família	não encontrado
genero	não encontrado
espécie	2826 suspeito(s)
subespécie	6 suspeito(s)
autor	8203 suspeito(s)
duplicata	7364 suspeito(s)
outras inconsistências	44 suspeito(s)

dados geográficos	
inventário	país - estado - município
nome do país/estado	681 suspeito(s)
outlier	137 suspeito(s)
long/lat fora do limite mundial	217 suspeito(s)
long/lat iguais	25 suspeito(s)
long ou lat com zero	360 suspeito(s)
long/lat mar (Brasil)	1489 suspeito(s)
nome do município (Brasil)	17770 suspeito(s)
análise da coordenada (Brasil)	93 suspeito(s)
outras inconsistências	1679 suspeito(s)

data da coleta	
data da coleta menor que 1931	2491 suspeito(s)
ano da coleta maior que atualização	não encontrado
ano da coleta maior que identificação	229 suspeito(s)

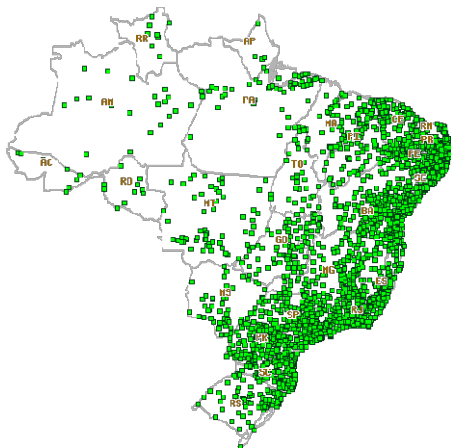
sugestões para campos em branco	
long/lat (Brasil)	26177 sugestões
nome país/estado	114 sugestões
nome município (Brasil)	1049 sugestões

busca
dataCleaning

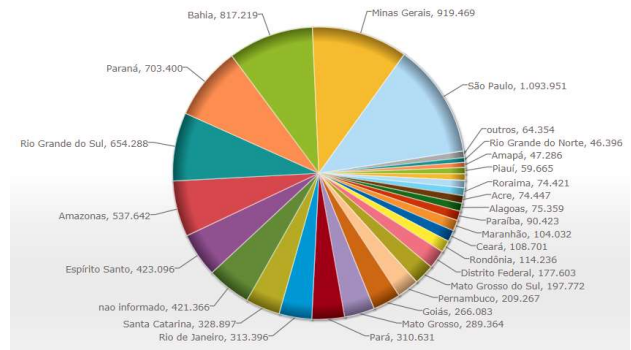
email
Centro de Referência em Informação Ambiental, CRIA

O CRIA oferece apoio às coleções no **compartilhamento** dos registros e imagens, na avaliação da **qualidade** dos dados, e na **valoração** do seu trabalho. Um dos princípios básicos da rede é o da mínima interferência na rotina das coleções, o que garante a elas a possibilidade de escolha do sistema de gerenciamento de seus dados que melhor atende às suas necessidades. A rede *speciesLink* tem mecanismos que permitem à coleção **total controle sobre o compartilhamento de seus dados**, que é feito da forma mais simples e menos impositiva possível.

Uma outra importante diretriz da rede é a da manutenção da **responsabilidade sobre os dados pelo provedor**. Nenhuma modificação é feita nos dados enviados à rede. Ao invés disso, a rede disponibiliza uma série de relatórios sobre a qualidade dos dados, denominado *dataCleaning*, o que facilita enormemente a correção de eventuais erros pela coleção em seu sistema local.



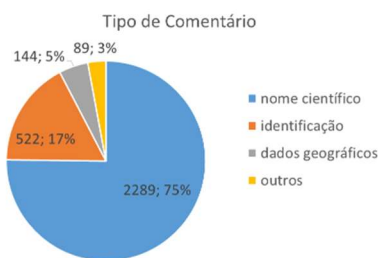
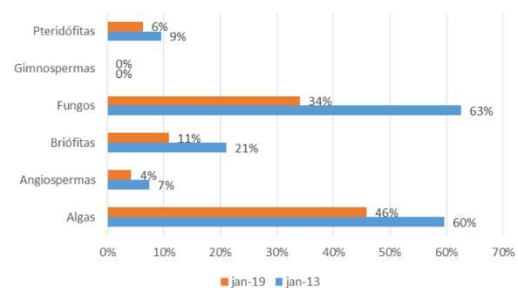
A rede *speciesLink* disponibiliza dinamicamente vários **infográficos** que são **indicadores** da rede e suas sub-redes, assim como de cada coleção, demonstrando a evolução de registros compartilhados, composição do acervo (por estado, família, data de coleta, etc.) e mapas. Essa informação é de grande importância para as coleções que a utilizam na produção de relatórios sobre seus acervos para suas instituições e para as agências de fomento.



Por fim, no escopo do projeto INCT – Herbário Virtual da Flora e dos Fungos foram desenvolvidas ferramentas que, entre outros usos, são fundamentais para orientar novas coletas e priorizar a digitação dos dados: **Lacunas**¹¹ e **BioGeo**¹². O primeiro indica o *status* dos dados para as espécies listadas na Flora do Brasil 2020, apontando as lacunas geográficas e taxonômicas do Herbário Virtual. BioGeo é um *workflow* para a produção e publicação de modelos de distribuição de espécies baseados em seus nichos ecológicos.

O sistema Lacunas é também uma importante ferramenta para avaliar a evolução qualitativa dos dados do Herbário Virtual da Flora e dos Fungos. Comparando a análise das lacunas taxonômicas do relatório de janeiro de 2013 com janeiro de 2019, o número de espécies de fungos sem registros no herbário virtual diminuiu de 60 a 34% das espécies citadas na Flora 2020.

Porcentagem de espécies da Flora 2020 sem registros no Herbário Virtual



Com relação ao BioGeo, hoje, 4.778 espécies da Lista da Flora do Brasil de 2014 (11%) têm pelo menos um modelo de distribuição geográfica.

No final de 2012, o CRIA desenvolveu um mecanismo para que usuários da rede *speciesLink* possam enviar seus comentários em relação a registros específicos. Os comentários são enviados ao curador e anexado ao registro como uma anotação. No período foram anexadas mais de 2,8 mil anotações sobre o nome da espécie, sinalizando a identificação via rede de espécimes com imagens online (cibertaxonomia).

serviços oferecidos a outras redes

A rede *speciesLink* alimenta ainda, com atualizações mensais, outras redes de informação semelhantes como a rede *Global Biodiversity Information Facility*, GBIF, as redes americanas *Integrated Digitized Biocollections*, *iDigBio*, e *Vertnet*, além do *Sistema de Informação sobre a Biodiversidade Brasileira*, SiBBr. Para esses sistemas, em 2015, graças ao apoio do

CNPq, foi desenvolvido um serviço de compartilhamento de dados - IPT (*Integrated Publishing Toolkit*)¹³ no âmbito do projeto INCT – Herbário Virtual. O IPT do CRIA hoje disponibiliza quase 5 milhões de registros.

A rede colabora também com a *Flora do Brasil 2020* disponibilizando dados e imagens que são associadas aos registros de nomes, assim como com o CNCFlora – *Centro Nacional de Conservação da Flora*, que utilizou os dados da rede para avaliar o *status* de conservação das espécies da flora brasileira.

uso dos dados

A rede *speciesLink* mantém um indicador dinâmico *on-line* sobre o uso dos dados da rede¹⁴. Esse indicador contabiliza apenas os dados que realmente são utilizados pelos usuários na produção de gráficos, mapas, listas, fichas ou *download* através da interface web.

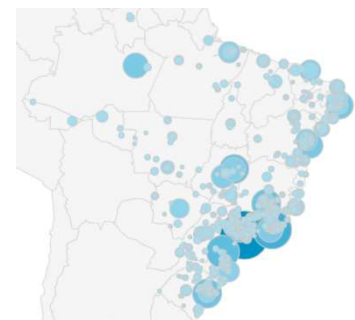


O ano de 2019 apresenta uma média diária de cerca de **1,9 milhão de registros e 9 mil imagens utilizadas por dia**. Ao longo dos últimos cinco anos, o uso representa cerca de **818 vezes** o acervo médio disponível *on-line*.

Outro indicador importante em relação ao uso dos dados é o número de publicações que citam a rede *speciesLink* e/ou os aplicativos e ferramentas disponíveis online. Através da busca no **Google Scholar** foram recuperados **463 artigos, 6 dissertações** de mestrado e um trabalho de conclusão de curso que citam a rede *speciesLink* ou ferramentas desenvolvidas pelo CRIA. O GBIF atribui um identificador único (DOI) a todos os downloads enviados por email, solicitando que o usuário cite o DOI caso publique algum artigo usando os dados. Esse DOI é associado a cada provedor cujos dados foram incluídos no download. Dessa maneira é possível associar os artigos publicados aos dados utilizados. Em 2018, **76 artigos** citam o uso dos dados de dados compartilhados pelo IPT do CRIA ao GBIF. São, portanto, mais de 500 artigos que, em 2018, citaram o uso da e-infraestrutura *speciesLink* e seus aplicativos.

usuários

Em 2018 a interface de busca foi acessada por 130 mil visitantes únicos. Cerca de 94% dos acessos à rede *speciesLink* são originados no Brasil. Esse fato é importante uma vez que um dos principais focos da rede é fomentar a pesquisa nacional sobre biodiversidade e alimentar os processos de formulação de políticas públicas e tomada de decisão.



considerações finais

Os resultados alcançados pela rede *speciesLink* sem dúvida representam uma história de sucesso no estabelecimento de uma rede de provedores e usuários de dados em biodiversidade, trabalhando em sintonia e de forma inovadora com as equipes do CRIA e da RNP no desenvolvimento dessa *e*-infraestrutura de acesso livre e aberto. Essa rede social, apoiada pelo desenvolvimento contínuo de tecnologias de informação e comunicação, é responsável pela inovação constante do sistema de informação.

Esse sucesso, entretanto, com um volume cada vez maior de dados e uma demanda cada vez mais crescente por novas funcionalidades e serviços aumenta a pressão financeira, abalando a sustentabilidade da rede *speciesLink* no curto prazo.

O Centro de Dados Compartilhados (CDC) da RNP, provendo infraestrutura física como serviço, com apoio financeiro adequado, certamente representa uma solução de parte do problema. A outra parte, sob responsabilidade do CRIA, que também precisa de apoio financeiro adequado e de longo prazo, refere-se à manutenção dos sistemas de informação e suporte aos provedores de dados e usuários do sistema.

Novos desenvolvimentos poderão ser financiados através de editais e projetos de curto prazo. Entretanto, o *core* da rede precisa estar garantido. Portanto, o apoio de longo prazo aos serviços do CDC da RNP e do CRIA para a manutenção e atualização da *e*-infraestrutura é a solução que buscamos para garantir o serviço contínuo da rede *speciesLink* para a educação, pesquisa e formulação de políticas públicas.

¹ rede *speciesLink*: splink.cria.org.br

² distribuição geográfica das coleções biológicas da rede *speciesLink*: www.splink.org.br/showNetwork

³ Diretrizes e estratégias para a modernização de coleções biológicas brasileiras e a consolidação de sistemas integrados de informação sobre biodiversidade: www.cria.org.br/cgee/col/

⁴ Indicadores da rede *speciesLink*: splink.cria.org.br/indicators/

⁵ Network manager: splink.cria.org.br/manager

⁶ INCT Herbário Virtual da Flora e dos Fungos: inct.splink.org.br

⁷ SIColNet: sicol.splink.org.br

⁸ Flora brasiliensis: fb.cria.org.br

⁹ Herbário Virtual A. de Saint-Hilaire: hvsh.cria.org.br

¹⁰ Herbário Virtual Auguste Glazou: glaziou.cria.org.br

¹¹ Lacunas de conhecimento da flora e dos fungos do Brasil: lacunas.inct.florabrasil.net

¹² Biogeografia da Flora e Fungos do Brasil: biogeo.inct.florabrasil.net

¹³ Integrated Publishing Toolkit: ipt1.cria.org.br/ipt

¹⁴ Estatísticas do uso dos dados da rede *speciesLink*: www.splink.org.br/showUsage