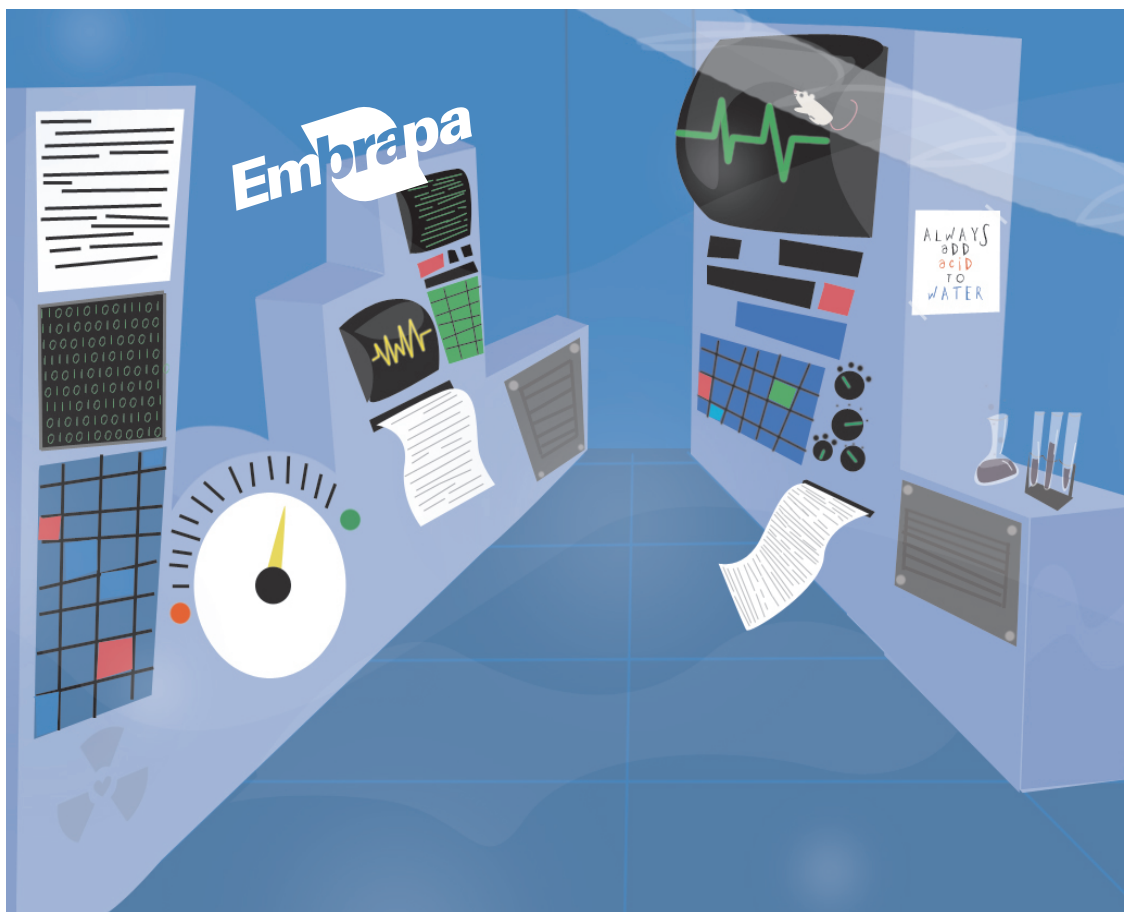


e-Science na Embrapa



ISSN 1677-9274
Dezembro, 2011

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 117

e-Science na Embrapa

Glauber José Vaz

Embrapa Informática Agropecuária
Campinas, SP
2011

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041 - 13083-886 - Campinas, SP
Fone: (19) 3211-5700 - Fax: (19) 3211-5754
www.cnptia.embrapa.br
sac@cnptia.embrapa.br

Comitê de Publicações

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa, Stanley Robson de Medeiros Oliveira, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo, Paula Regina Kuser Falcão*

Supervisão editorial: *Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Edição eletrônica: *Suzilei Almeida Carneiro, Neide Makiko Furukawa*

Arte capa: *Suzilei Almeida Carneiro*

Fotos da capa: *Imagens livres disponíveis em <<http://www.stock.schng>>*

Secretária: *Carla Cristiane Osawa*

1ª edição on-line 2011

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP) Embrapa Informática Agropecuária

Vaz, Glauber José.

e-Science na Embrapa / José Glauber Vaz. - Campinas : Embrapa Informática Agropecuária, 2011.

58p. : il. - (Documentos / Embrapa Informática Agropecuária; ISSN 1677-9274, 117).

1. e-Science. 2. Ciência da computação. 3. Infraestrutura científica. I. Embrapa Informática Agropecuária. II. Título. III. Série.

CDD 004 (21. ed.)

© Embrapa 2011

Autor

Glauber José Vaz

Mestre em Ciência da Computação

Analista da Embrapa Informática Agropecuária

Av. André Tosello, 209, Barão Geraldo

Caixa Postal 6041 - 13083-970 - Campinas, SP

Telefone: (19) 3211-5839

e-mail: glauber@cnptia.embrapa.br

Apresentação

A Ciência da Computação já vem sendo considerada o terceiro pilar da Ciência, ao lado da teoria e da experimentação. No entanto, muitos pesquisadores já consideram essa área do conhecimento a responsável pela emergência de um quarto paradigma científico, *e-Science*.

Atualmente, grandes desafios, como o tratamento de uma imensa quantidade de dados e a crescente necessidade de colaboração entre cientistas das mais variadas áreas, exigem mais do que uma automatização da Ciência. São necessários novos métodos e uma próxima geração de infraestrutura para se fazer Ciência e se resolver problemas, além de meios eficazes para que cientistas realmente trabalhem colaborativamente. A Computação está presente em todos esses aspectos, e, em essência, é disso que se trata a *e-Science*.

A Embrapa, por ser uma empresa de pesquisa, deve assumir um papel relevante nas iniciativas relacionadas a *e-Science*. A associação da Computação ao domínio agropecuário pode prover as soluções de que a Embrapa necessita, a fim de se promover um maior avanço nas pesquisas realizadas na empresa. Nesse contexto, a Embrapa Informática Agropecuária assume um papel de extrema importância.

Este documento explica em detalhes o que vem sendo chamado de *e-Science* e outros termos relacionados a esse conceito. Também apresenta propostas do que a Embrapa pode fazer em relação à aplicação das ideias preconizadas pela *e-Science*, conforme suas prioridades estratégicas.

Kleber Xavier Sampaio de Souza

Chefe-geral

Embrapa Informática Agropecuária

Sumário

1 Introdução	9
1.1 Programas de <i>e-Science</i>	10
1.2 Ciência no Brasil	11
1.3 Objetivos	13
1.4 Organização	14
2 e-Science, Computação e Ciência	15
2.1 O papel da Computação na Ciência	15
2.2 A emergência de um quarto paradigma científico	18
2.3 O que é <i>e-Science</i> ?	19
3 Infraestrutura para a pesquisa científica	21
3.1 Recursos de dados e informações	24
3.2 Software	27
3.3 Capital humano	28
3.4 Estrutura organizacional	28
3.5 Ambiente virtual de pesquisa	29
3.6 Considerações sobre alguns termos relacionados a <i>e-Science</i>	31
3.7 Futuro da e-infraestrutura e da <i>e-Science</i>	32
4 Propostas	34
4.1 Embrapa, <i>e-Science</i> e a 4ª CNCTI	35
4.1.1 Infraestrutura científica	36
4.1.2 Colaboração e multidisciplinaridade	37
4.1.3 Computação e TICs	38
4.1.4 Grande volume de dados	38

4.2	Prioridades estratégicas.....	39
4.3	Embrapa Informática Agropecuária	40
4.4	Laboratório Multiusuário de Bioinformática.....	45
4.5	Formação de recursos humanos para <i>e-Science</i>	48
5	Conclusão	50
6	Referências	53

e-Science na Embrapa

Glauber José Vaz

1 Introdução

A Ciência da Computação já é reconhecida como terceiro pilar da Ciência. No entanto, a área tem sido responsável, segundo pesquisadores do mundo todo, pela emergência de um quarto paradigma científico, baseado em grande volume de dados, que vem sendo chamado de *e-Science*. Esse termo surgiu na virada do milênio e representa não apenas o futuro da pesquisa científica, mas uma nova revolução na Ciência, que passou a contar com ferramentas, conceitos e teoremas da Ciência da Computação em sua própria construção (EMMOTT, 2006). Portanto, não se trata apenas de automatizar a ciência, mas de aplicar novos métodos que podem revolucionar a maneira de se fazer ciência e que possibilitam o tratamento de grandes desafios que não poderiam ser resolvidos de outra maneira (RESEARCH COUNCILS UK, 2009).

Como os dados e a computação direcionam cada vez mais a ciência, surgem desafios tecnológicos importantes que envolvem a realização de captura, análise, modelagem e visualização da informação científica, da melhor maneira possível, com a finalidade de auxiliar a tomada de decisão de cientistas, formuladores de políticas e o público em geral (HEY et al., 2009).

A Empresa Brasileira de Pesquisas Agropecuária (Embrapa) lida cotidianamente com um imenso volume de dados, gerados e coletados constantemente. Sequências genômicas, imagens de satélites e simulações climá-

ticas são apenas alguns exemplos de domínios que costumam envolver grande volume de dados e que são tratados na empresa. A fim de se obter um maior avanço nas pesquisas, é necessário criar mecanismos capazes de extrair informação e conhecimento a partir desses dados, e a pesquisa em Computação associada ao domínio agropecuário pode prover as soluções de que a Embrapa necessita. Esse é justamente um dos papéis que a unidade Embrapa Informática Agropecuária tem desempenhado nos últimos anos.

Para uma empresa que pretende ser referência em Tecnologia da Informação (TI) no país (EMBRAPA, 2010) e que é citada frequentemente em publicações científicas, inclusive naquelas que são reconhecidas internacionalmente, a Embrapa deve assumir um papel relevante nas iniciativas relacionadas a *e-Science*, até porque são fundamentais para que a empresa ocupe a liderança mundial na geração de conhecimento, tecnologia e inovação para a produção sustentável de alimentos, fibras e agroenergia.

Em nível nacional, é importante a intensificação de iniciativas relacionadas a *e-Science*, o que significa o reconhecimento de que a inovação baseada no conhecimento é cada vez mais crítica para o bem-estar social e econômico (RESEARCH COUNCILS UK, 2009).

1.1 Programas de *e-Science*

Em 2001, o Reino Unido lançou um programa pioneiro, que recebeu investimentos da ordem de £250 milhões, para estimular o desenvolvimento de *e-Science* em todos os campos de pesquisa (ATKINSON et al., 2009). Quando foi lançado por Taylor (2001), o programa de *e-Science* do Reino Unido visava a prover infraestrutura e facilidades necessárias para a pesquisa colaborativa, acelerar a emergência da próxima geração de padrões de plataforma aberta para serviços globais de informação, resolver os principais desafios em processamento, comunicação e armazenamento de grandes volumes de dados e prover soluções genéricas para as necessidades de disciplinas e aplicações individuais. Em revisão feita em 2009 (RESEARCH COUNCILS UK, 2009), o programa foi considerado um sucesso: teve um impacto acadêmico muito alto, foi precursor de vários outros programas similares em todo o mundo, colocou o Reino Unido em

primeiro plano no desenvolvimento de *e-Science* e contribuiu para que a *e-Science* fosse aceita como um paradigma de pesquisa válido.

Programas de *e-Science* promovem a colaboração entre os pesquisadores da Computação com os de outras áreas e ainda fornecem uma infraestrutura científica baseada em tecnologias da informação e da comunicação (TICs). Os resultados alcançados por cientistas do Natural Environment Research Council (NERC), Reino Unido, ilustram as vantagens dessa abordagem de se compartilhar métodos, ferramentas, dados e computação de alto desempenho. Em 3 anos, eles publicaram 13 artigos na Nature (ATKINSON et al., 2009).

No Brasil, mais de uma década depois do início do programa do Reino Unido, há poucos cientistas que têm conhecimento ou interesse sobre o tema e tampouco há um programa nacional integrado. Se por um lado, o avanço da *e-Science* em outros países constitui uma boa direção para o desenvolvimento da *e-Science* brasileira, por outro, percebe-se o atraso do país nesse cenário.

E a falta de iniciativas nessa área apresenta vários riscos (ATKINSON et al., 2009):

- Perda de competitividade, pois inibe-se a agilidade na inovação à medida que não se consegue compartilhar recursos digitais, nem aumentar a colaboração.
- Baixo retorno nos investimentos, pois além de se perder oportunidades de compartilhamento, há duplicação e fragmentação excessiva de recursos e esforços.
- Falta de disseminação das abordagens usadas pelos pesquisadores.
- Perda de influência internacional.

O Brasil, portanto, precisa tomar medidas mais ousadas para desenvolver *e-Science*.

1.2 Ciência no Brasil

Artigos recentes da *The Economist* (SCIENCE..., 2011) e da *Science* (REGALADO, 2010) destacaram a evolução da Ciência no Brasil, que deseja ser, até 2020, uma das 10 maiores potências científicas mundiais.

Líderes em medicina tropical, bioenergia e biologia de plantas, os brasileiros, que eram responsáveis por 1,7% dos artigos científicos no mundo em 2002, alcançaram a marca de 2,7% em 2008.

Na *The Economist* (SCIENCE..., 2011), afirma-se que, fazendo sua própria ciência, os países tropicais em desenvolvimento garantem que não se resolvam apenas os problemas de países ricos em clima temperado. A Embrapa, citada na *Science* (REGALADO, 2010) como uma das principais unidades agrícolas do mundo, é fundamental para resolver problemas extremamente importantes do Brasil e outras nações, uma vez que os problemas agropecuários encontrados no país não apresentam as mesmas características dos problemas dos países desenvolvidos.

Criar soluções visando ao desenvolvimento sustentável no Brasil é o principal compromisso da Embrapa e esse viés está cada vez mais forte nas chamadas Ciências da Terra e do Ambiente, que agora atravessam uma fase em que o principal objetivo é resolver grandes problemas, e não apenas publicar artigos científicos, considerados meios para a criação de soluções. Essa ciência de aplicações ambientais diferencia-se da pesquisa básica por ser: guiada pelas necessidades; baseada em grande volume de dados; influenciada por circunstâncias externas, políticas e econômicas; útil, mesmo quando incompleta; entre outras características. As pesquisas em mudanças climáticas, que também são desenvolvidas na Embrapa, representam bem essa nova fase das Ciências da Terra e do Ambiente (DOZIER; GAIL, 2009).

No campo político, Jeffrey Sachs (2010), diretor do *Earth Institute*, afirma em artigo da *Nature* (NORVIG et al., 2010) que o mundo precisa de um sistema eficaz de governança global para conduzir o desenvolvimento sustentável. E uma das medidas sugeridas é que a política deva considerar os conhecimentos técnicos. Ele destaca que em negociações internacionais, despende-se muito tempo com discussões sobre questões legais dos acordos e pouco tempo para as questões técnicas. Há uma tendência em anunciar metas sem estratégias técnicas e então não cumprir as metas. As decisões políticas devem ser melhor sustentadas nos conhecimentos técnicos e a Embrapa tem um papel fundamental para o Brasil nesse contexto em questões relacionadas à agropecuária.

No entanto, o artigo da *Science* (REGALADO, 2010) destaca a fragilidade da educação pública no país e a produção incipiente de patentes e de artigos com alto impacto. Como exemplo, afirma que desde a publicação na

Nature do trabalho envolvendo o sequenciamento da *Xylella fastidiosa*, não houve um artigo de grande repercussão na área de genômica. Portanto, é necessário haver um esforço para se obter bons resultados e possibilitar que eles sejam duradouros e sustentem novos trabalhos de impacto. Uma infraestrutura que viabilize a *e-Science* ajuda muito, e, em um primeiro instante, esforços podem ser concentrados em áreas estratégicas como as ambientais e a bioinformática.

Um programa nacional de *e-Science* traria uma série de benefícios, mas as instituições brasileiras já podem se adaptar a esse paradigma, adequando sua infraestrutura para desfrutar das vantagens proporcionadas pela *e-Science* e preparar-se para um futuro que, necessariamente, passa pela integração das instituições envolvidas com pesquisa. A Embrapa já está se preparando para essas novas circunstâncias, com a criação, por exemplo, de laboratórios multiusuários que visem ao compartilhamento de infraestrutura. Mas ainda há muito a ser feito.

1.3 Objetivos

Este documento visa a fornecer informações importantes sobre *e-Science* e o que vem sendo feito a respeito do tema. É rico em fontes bibliográficas, uma vez que o tema é extremamente abrangente e torna-se impossível aprofundar nos inúmeros tópicos aqui abordados. Assim, importantes referências à literatura são feitas para oferecer um bom ponto de partida para pesquisadores que desejam atuar em áreas específicas envolvidas pela *e-Science*.

Com base nessas informações, procura-se indicar caminhos para que a Embrapa se insira no contexto global de *e-Science*, organize melhor seus dados e informações, elabore estratégias fundamentadas nas experiências e nos estudos já realizados e conceba, da melhor maneira possível, os laboratórios multiusuários, uma das prioridades da empresa.

Construir uma plataforma de *e-Science* não é trivial e consome muitos recursos. Este documento fornece diretrizes para sua implementação sem se aprofundar em assuntos específicos. A ideia é que se estimule e facilite o desenvolvimento de *e-Science* na Embrapa por meio de projetos de pesquisa.

Neste documento, ainda fica clara a importância da Embrapa para a ciência e o desenvolvimento sustentável brasileiros em um mundo que depende cada vez mais da inovação baseada no conhecimento e, mais especificamente, a importância da Embrapa Informática Agropecuária devido às suas pesquisas em Computação associada a domínios agropecuários.

1.4 Organização

Este documento é organizado de maneira que um leitor com formação em qualquer área do conhecimento possa compreender seu conteúdo e suas principais ideias, mesmo que não esteja acostumado à linguagem associada à Computação.

Assim, depois de considerar a situação da Ciência no Brasil, a importância dos programas de *e-Science* e os objetivos deste documento, a seção 2 começa por exibir as maneiras como a Computação pode estar associada à Ciência. Para isso, expõe o uso de vários termos usados nessa área, como, por exemplo, Ciência da Computação, Informática e Ciência Computacional. Também explica o termo *e-Science*, que vem sendo utilizado para representar um novo paradigma científico.

A infraestrutura necessária para o avanço científico é tratada na seção 3. O foco é dado ao papel da Computação e a alguns dos principais aspectos envolvidos na chamada e-infraestrutura: recursos de dados e informações, software, capital humano e estrutura organizacional. Ainda são tratados nessa seção o ambiente virtual de pesquisa, um dos componentes mais importantes de uma plataforma de *e-Science*, questões relacionadas à terminologia associada à e-infraestrutura e à *e-Science*, e considerações sobre o futuro desses 'e-termos'.

A seção 4 apresenta propostas para a Embrapa de iniciativas relacionadas a *e-Science* fortemente baseadas na 4ª Conferência Nacional de Ciência, Tecnologia e Inovação para o Desenvolvimento Sustentável e nas prioridades estratégicas da empresa. A principal delas envolve o Laboratório Multiusuário de Bioinformática e uma plataforma de *e-Science* para problemas da bioinformática. Também é proposto um curso sobre *e-Science* na Universidade para a preparação de profissionais competentes nesse novo paradigma.

Finalmente, a seção 5 conclui o documento com reflexões e considerações sobre a *e-Science* e sobre o que a Embrapa e o Brasil podem fazer a respeito.

2 e-Science, Computação e Ciência

Para entender o que significa *e-Science*, é necessário, em primeiro lugar, compreender o papel da Computação na Ciência, e como ela vem sendo responsável pelo avanço da pesquisa científica, promovendo a emergência do que é considerado por especialistas um quarto paradigma científico. Em seguida, são enumeradas várias definições encontradas na literatura para *e-Science*, a fim de se caracterizar melhor o que esse termo representa e de se extrair suas ideias fundamentais.

2.1 O papel da Computação na Ciência

Para se compreender melhor o papel da Computação na Ciência, é necessário caracterizar e diferenciar o que representam termos como informática, Ciência Computacional, Ciência da Computação e pensamento computacional. Depois disso, fica mais fácil entender o que tem sido chamado de *e-Science*, foco deste trabalho.

O termo Ciência Computacional (*Computational Science*), segundo a Sociedade Brasileira de Computação (2006), vem sendo usado para designar modelos, algoritmos e ferramentas computacionais para a solução de sistemas complexos de diferentes naturezas. Nessa mesma linha, o Grupo Ciência 2020, composto por mais de 30 reconhecidos cientistas de 12 diferentes nacionalidades e de várias áreas do conhecimento, inclusive Ciência da Computação, afirma que o termo é normalmente utilizado para denotar a aplicação da computação como suporte à Ciência (EMMOTT, 2006). Em maiores detalhes, um relatório elaborado pelo Comitê Assessor de TI para o presidente dos EUA (BENIOFF et al., 2005) define e caracteriza Ciência Computacional. Afirma que pode ser considerada simples-

mente a aplicação de recursos computacionais à solução de problemas do mundo real, mas para ressaltar seu caráter multidisciplinar, a Ciência Computacional foi definida como um campo multidisciplinar que vem crescendo rapidamente e que usa computação avançada para entender e resolver problemas complexos. Esta última definição, portanto, se aproxima do que é Ciência da Computação.

Informática (*informatics*) é outro termo que está intimamente relacionado à Ciência Computacional. Seu significado varia de acordo com o lugar onde é usado, mas no Brasil, assim como nos Estados Unidos, está relacionado à computação aplicada ou à computação no contexto de outro domínio. Informática, portanto, pode ser definida como uma disciplina que resolve problemas por meio da aplicação da computação no contexto de domínio do problema (GROTH; MACKIE-MASON, 2010).

Ciência da Computação (*computer science*), de acordo com Wing (2006), refere-se ao estudo da computação: o que e como pode ser computado, e vai muito além da capacidade de se programar um computador. Os conceitos da Ciência da Computação fornecem níveis de abstração que permitem aos cientistas de diferentes campos entender e aprender soluções uns com os outros (EMMOTT, 2006). Desta forma, a Ciência da Computação é, talvez, melhor caracterizada pelo pensamento computacional (*computational thinking*), termo introduzido e muito bem explicado por Wing (2006). Resumidamente, pensamento computacional consiste em se adotar uma abordagem de resolver problemas, projetar sistemas e entender o comportamento humano, usando conceitos fundamentais da Ciência da Computação (WING, 2005). É importante salientar que as bases teóricas da Computação são muito sólidas e visam responder a uma questão essencial: Quais são as capacidades fundamentais e as limitações dos computadores? (SIPSER, 2006). A Teoria da Computação procura determinar o que pode e o que não pode ser computado, com que velocidade, com quanto de memória, e ainda com que tipo de modelo computacional (SIPSER, 2006).

Em resumo, enquanto Ciência da Computação tem a computação como atividade-fim, a Ciência Computacional está relacionada ao uso de ferramentas computacionais para resolver problemas nas diferentes áreas do conhecimento. No entanto, há uma estreita relação entre esses termos, pois, ao mesmo tempo que a Ciência Computacional permite estudos antes impossíveis em vários domínios, os avanços destes últimos geram

desafios de pesquisa em Computação (SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2006). É importante destacar, porém, que é comum o uso desses termos com sentidos ligeiramente diferentes. O próprio relatório produzido pelo governo norte-americano (BENIOFF et al., 2005) aproxima o significado de Ciência Computacional ao de Ciência da Computação afirmando que a primeira consiste no uso de capacidades computacionais avançadas para entender e resolver problemas complexos, o que está intimamente relacionado ao pensamento computacional, característico da segunda.

A Tabela 2.1 exhibe algumas diferenças que podem ser apontadas entre Ciência Computacional e Ciência da Computação, conforme são normalmente utilizadas.

Tabela 2.1. Comparação entre Ciência Computacional e Ciência da Computação.

Termo	Ciência Computacional	Ciência da Computação
Ênfase	Aplicações da Computação	Conceitos da Computação
Termo correlato	Informática	Pensamento computacional
Abordagem ¹	Multidisciplinar	Interdisciplinar
Papel da Computação	Atividade-meio	Atividade-fim

O grupo Ciência 2020 (EMMOTT, 2006) considera que a visão de Ciência Computacional é um estado intermediário, insustentável e indesejável. A Computação tem muito mais a contribuir à Ciência do que simplesmente a sua aplicação aos problemas, e é esta interseção da Ciência da Computação com as demais ciências que representa os fundamentos de uma nova revolução na Ciência.

¹ Enquanto na multidisciplinaridade, os objetivos próprios de cada disciplina (área de conhecimento) são preservados e as articulações entre elas são tênues, na interdisciplinaridade, o que se busca é o estabelecimento de uma intercomunicação efetiva entre as disciplinas, por meio do enriquecimento das relações entre elas (MACHADO, 2002).

2.2 A emergência de um quarto paradigma científico

O relatório elaborado pelo governo norte-americano (BENIOFF et al., 2005), e já citado anteriormente neste documento, aponta a Ciência Computacional como o terceiro pilar da pesquisa científica, juntamente com a teoria e a experimentação, pois permite que pesquisadores construam e testem modelos de fenômenos complexos que não podem ser replicados em laboratórios, e que tratem grandes volumes de dados de maneira rápida e econômica. A Sociedade Brasileira de Computação (2006) também reconhece a importância de se tratar grandes volumes de dados enumerando como um dos cinco grandes desafios atuais para a Computação no Brasil a “gestão da informação em grandes volumes de dados multimídia distribuídos”.

Porém, Jim Gray, prêmio Turing² de 1998, defendia que a Computação representa mais do que um terceiro pilar: representa um quarto paradigma. A Figura 2.1 ilustra a evolução dos paradigmas científicos na visão de Gray (2009), que já identificava nos experimentos científicos, e não nas aplicações comerciais, os maiores desafios para tratamento de dados no futuro (BELL et al., 2009).

Para entender a natureza, os cientistas reconheceram, na teoria e na experimentação, paradigmas científicos básicos. Porém, nas últimas décadas, simulações computacionais tornaram-se um terceiro paradigma essencial, uma ferramenta para que cientistas explorem domínios que são inacessíveis à teoria e à experimentação, como por exemplo, a predição das mudanças de clima e a evolução do Universo. Com o grande volume de dados gerado pelas simulações e pelos experimentos científicos, um quarto paradigma está emergindo, que consiste de técnicas e tecnologias necessárias para desenvolver uma ciência baseada em grande volume de dados (*data-intensive science*). Segundo artigo da *Wired* (ANDERSON, 2008), hoje é possível fazer ciência analisando dados sem ter hipóteses prévias sobre eles, com algoritmos que podem encontrar padrões, mesmo que não haja teorias ou modelos coerentes sobre os dados analisados. Não é mais necessário seguir uma abordagem de hipótese – modelo – teste.

² O prêmio Turing é considerado o “Prêmio Nobel da Computação” e é concedido anualmente pela Association for Computing Machinery (ACM) por contribuições no campo da Computação.

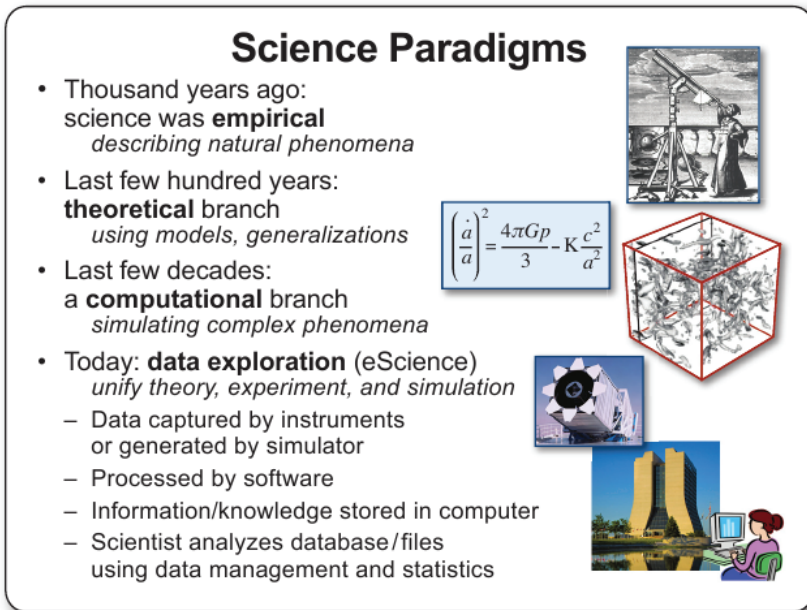


Figura 2.1. Paradigmas científicos na visão de Gray.

Fonte: Gray (2009).

À síntese da Tecnologia da Informação com a Ciência, que permite lidar com desafios em escalas antes inimagináveis, Jim Gray deu o nome de *e-Science* (BELL et al., 2009).

2.3 O que é e-Science?

Não há um consenso em relação à definição de *e-Science*, mas várias são apresentadas pela comunidade científica. Esse termo foi cunhado por John Taylor, quando era diretor geral dos Conselhos de Pesquisa do Reino Unido, e usado para envolver a coleção de ferramentas e tecnologias necessárias para suportar a pesquisa científica do século 21, caracterizada pela natureza colaborativa e multidisciplinar, pelo grande volume de dados e pelo importante papel da Tecnologia da Informação (TI).

Em 2001, quando foi lançado o programa de *e-Science* do Reino Unido, uma das primeiras iniciativas relacionadas ao tema, *e-Science* foi definida

como a pesquisa feita por colaborações globalmente distribuídas e possibilitadas pela internet. Ainda foi caracterizada pelo uso de grandes coleções de dados, recursos de computação em teraescala³ e pela visualização de alto desempenho (RESEARCH COUNCILS UK, 2009).

Em outra definição de *e-Science* atribuída a Taylor (HEY; TREFETHEN, 2003; THE LONDON E-SCIENCE CENTRE, 2011), o termo foi definido como a colaboração global em áreas chave da Ciência e a próxima geração de infraestrutura que vai possibilitá-la.

De maneira mais sucinta, conforme mencionado anteriormente, Jim Gray definiu *e-Science* como a síntese de Tecnologia da Informação e Ciência, que permite resolver desafios em escalas antes inimagináveis. Para ele, *e-Science* é onde a “TI se encontra com os cientistas” (GRAY, 2009).

De maneira alternativa, o professor Malcolm Atkinson usa o termo *e-Science* para denotar o desenvolvimento sistemático de métodos de pesquisa que exploram o pensamento computacional avançado. A ênfase, neste caso, está na abordagem fundamentada na Ciência da Computação para resolver problemas.

Getov (2008) também foca nos métodos, afirmando que *e-Science* envolve os novos métodos revolucionários para a descoberta e a colaboração científica experimental, incluindo simulações e ambientes virtuais, que são viabilizados pela moderna infraestrutura de computação em grade.

O significado do “e” de *e-Science* normalmente não é explicado, mas segundo documento do Research Councils UK (RCUK⁴) (2011), sem dúvida, começou com o significado de *electronic* (eletrônico) e agora representa melhor *enhanced* (melhorada) ou *enabled* (habilitada).

Como pôde ser observado, há várias definições para *e-Science*, mas é possível destacar suas principais características: o uso de uma nova geração de infraestrutura científica, seu caráter colaborativo e multi/interdisciplinar e o grande volume de dados, além do importante papel desem-

³ Teraescala refere-se à ordem de grandeza 10^{12} . Poder de processamento em teraescala equivale à execução de trilhões de operações por segundo.

⁴ O Research Councils UK (RCUK 2011) é uma parceria estratégica entre os Conselhos de Pesquisa do Reino Unido criada em 2002 para permitir que eles trabalhem colaborativamente de maneira mais eficaz e gerem maior impacto de suas atividades de pesquisa, treinamento e inovação, contribuindo assim, na realização dos objetivos governamentais para a Ciência e a inovação.

penhado pela Computação na pesquisa científica. Portanto, *e-Science* denota a ciência desenvolvida com o mínimo possível de limitações de recursos computacionais e colaborativos, recursos estes que envolvem desde software, hardware e curadoria de dados até uma estrutura organizacional que estimule a cooperação entre os pesquisadores e potencialize os benefícios alcançados com a infraestrutura disponível.

A próxima seção é dedicada a essa infraestrutura necessária para o desenvolvimento da pesquisa científica.

3 Infraestrutura para a pesquisa científica

Taylor (2001) frequentemente relacionava *e-Science* à infraestrutura que a viabilizaria. A essa infraestrutura, ele deu o nome de Grid, uma nova geração de serviços de informação composta de *middleware*, software e hardware para acessar, processar, comunicar e armazenar grande volume de dados. Um termo correlato é e-infraestrutura (*e-infrastructure*), que denota equipamentos digitais, software, serviços, ferramentas, portais, implantações, equipes operacionais, serviços de suporte e treinamento que oferecem serviços de computação, dados e comunicação para os pesquisadores (ATKINSON et al., 2009). Esse termo também é definido como a infraestrutura computacional distribuída que oferece acesso compartilhado a grandes coleções de dados, ferramentas avançadas de TIC para análise de dados, recursos computacionais em grande escala e visualização de alto desempenho (POTHEN, 2006).

O termo e-infraestrutura é amplamente utilizado na Europa, mas também recebe outros nomes como ciberinfraestrutura (*cyberinfrastructure*), nos Estados Unidos, ou infraestrutura da informação (BOWKER et al., 2007). A National Science Foundation (NSF) dos Estados Unidos usou o termo ciberinfraestrutura para descrever a aplicação integrada, ubíqua e cada vez mais pervasiva de abordagens da computação científica e da tecnologia da informação (WOOLEY, 2003). No entanto, é importante destacar que além da ubiquidade da computação, é necessário haver qualidade. A própria Sociedade Brasileira de Computação (2006) apontou como grande desafio da Computação no Brasil o “Desenvolvimento Tecnológico de Qualidade: sistemas disponíveis, corretos, seguros, escaláveis, persistentes e ubíquos”.

Um relatório produzido no Reino Unido (POTHEN, 2006) aponta os principais benefícios e recursos que uma e-infraestrutura deve oferecer aos pesquisadores. Uma vez que constituem uma boa visão do que é uma e-infraestrutura, esses recursos são enumerados a seguir.

- Acesso a sistemas, serviços, redes e recursos onde for necessário.
- Meios para facilmente descobrir recursos e conseguir utilizá-los apropriadamente.
- Certeza da integridade, autenticidade e qualidade dos serviços e recursos usados.
- Garantia de que seus resultados sejam acessíveis, no tempo corrente e no futuro.
- Infraestrutura física independente do local para combinar computação e informação de várias fontes de dados.
- Tecnologia avançada para a pesquisa colaborativa.
- Formação e qualificação necessárias para explorar os serviços e recursos disponíveis.

A ideia é permitir que os pesquisadores possam:

- Explorar o poder de aplicações e tecnologias de informação avançadas para melhorar continuamente o processo de pesquisa.
- Colaborar e comunicar, com segurança, com outros pesquisadores, transitando em várias disciplinas, instituições e setores.
- Maximizar o potencial de tecnologias avançadas para dar suporte à inovação e à experimentação.
- Compartilhar resultados com outros pesquisadores e reutilizá-los no futuro.
- Relacionar-se com a indústria no suporte a objetivos econômicos mais amplos.

Assim, possibilita-se:

- O aumento da transferência de conhecimento e o desenvolvimento de aplicações comerciais a partir de resultados de pesquisa.
- O acompanhamento dos resultados de pesquisas por parte dos financiadores.

- A proteção da privacidade individual e do trabalho, satisfazendo restrições éticas, legais e regulatórias.
- A proteção da propriedade intelectual.
- A preservação da produção de informação digital como uma parte vital da herança intelectual e cultural da nação.

O relatório (POTHEN, 2006) também enfatiza a relevância de temas transversais na elaboração de um plano de e-infraestrutura:

- A transferência de tecnologia, preocupação constante da Embrapa, deve ser possibilitada pela e-infraestrutura para estimular o melhor aproveitamento do conhecimento produzido e da tecnologia desenvolvida para objetivos sociais e econômicos.
- Interoperabilidade e padronização, para facilitar a comunicação e a cooperação.
- Pesquisa e Desenvolvimento em TICs para desenvolver e manter uma e-infraestrutura adequada. Neste ponto, a Embrapa Informática Agropecuária tem um papel fundamental.
- As mudanças de cultura são ainda mais importantes do que as técnicas, pois o sucesso de uma e-infraestrutura depende da participação ativa dos variados grupos, instituições e disciplinas para o desenvolvimento de uma cultura de compartilhamento da produção científica e de um senso de coletividade da infraestrutura.
- A coordenação efetiva é muito importante para assegurar a clareza dos papéis e responsabilidades de todos os envolvidos e evitar desperdícios de recursos.
- A qualidade deve ser garantida.
- Os pesquisadores devem estar aptos para explorar os recursos oferecidos e auxiliar no desenvolvimento de novos serviços e ferramentas.

Caracterizada a e-infraestrutura, pode-se afirmar que o principal objetivo do movimento global de *e-Science* é determinar as maneiras de se usar as TICs, via uma e-infraestrutura, como fundamento para avançar a pesquisa e criar maiores benefícios com seus resultados (RESEARCH COUNCILS UK, 2009).

Segundo relatório da Research Councils UK (2009), *e-Science* serve como uma plataforma para o avanço da pesquisa. E o conceito de plataforma

considerado inclui não somente os sistemas que atendem às definições tradicionais da computação (software), mas também outras formas importantes de infraestrutura: organizacionais (grupos formais e informais que prestram serviços de *e-Science*), capital humano (conhecimento e experiência), e recursos de dados e informações (sistemas que suportam o crescente volume de dados gerados pela pesquisa).

A seguir, é analisada cada uma dessas dimensões: recursos de dados e informações, sistemas, capital humano e estruturas organizacionais. Depois disso, trata-se do ambiente virtual de pesquisa, componente muito importante de uma plataforma de *e-Science*. Também são feitas considerações sobre alguns dos principais termos relacionados a *e-Science* tratados nesta seção, a fim de tornar mais claras as diferenças entre eles. Concluem, esta seção, algumas reflexões sobre o futuro da e-infraestrutura e da *e-Science*.

3.1 Recursos de dados e informações

O ideal é que os cientistas consigam olhar para os dados facilmente, onde quer que eles estejam, e ter poder de processamento suficiente para usar os algoritmos que são adequados para processar estes dados. As plataformas atuais, portanto, devem permitir que os cientistas superem as barreiras computacionais que existem entre eles e os dados (EMMOTT, 2006).

Uma primeira etapa para a implantação de uma e-infraestrutura diz respeito ao cuidado com os dados, cujo ciclo de vida é exibido na Figura 3.1. Em outra abordagem para se lidar com os dados, as atividades necessárias são representadas por um conjunto de verbos: criar, armazenar, descrever, identificar, registrar, descobrir, acessar e explorar (BALL, 2010; BURTON; TRELOAR, 2009). Esses esquemas ajudam a compreender melhor o que deve ser feito para se ter uma e-infraestrutura adequada no que diz respeito à manipulação de dados.

Assim como a boa administração dos dados, seu compartilhamento também apresenta uma série de vantagens, que são apontadas por um guia (VAN DEN EYNDEN, 2009) disponibilizado pela *UK Data Archive* (2011), um centro especializado em obter, curar e prover acesso a dados. A maioria dos conselhos de pesquisa já impõe ou incentiva o compartilhamento

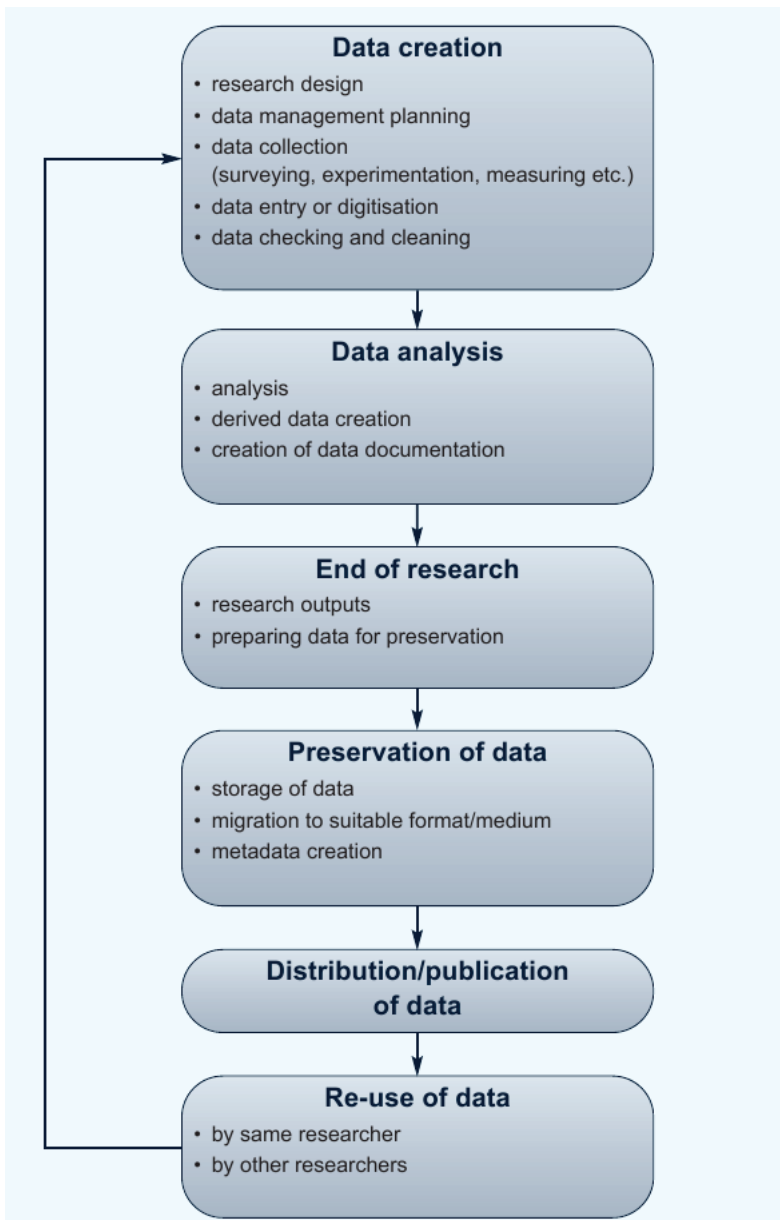


Figura 3.1. Ciclo de vida dos dados.

Fonte: Van den Eynden (2009).

de dados, cobrando planos de administração dos dados nas solicitações de financiamento. Além disso, a fim de manter sua credibilidade, publicações importantes requerem a manutenção dos conjuntos de dados para verificação e suporte à reprodutibilidade dos experimentos (HODSON, 2010; REPRODUCIBLE..., 2010). Há várias recomendações a cientistas, agências de fomento e editores de revista para que as pesquisas possam ser reproduzidas (REPRODUCIBLE..., 2010) e ter seus dados compartilhados, até mesmo antes de sua publicação (BIRNEY, 2009).

Há muitos fatores que podem representar obstáculos nessa questão: privacidade, propriedade intelectual, cultura acadêmica e tempo para preparação dos dados para compartilhamento, entre outros. No entanto, o processo de administração de dados deve ser racionalizado em um mundo de ciência baseada em dados. Os editores e universidades já estão pensando em maneiras de se adaptar para refletir e lidar com essa mudança de cultura (HODSON, 2010). A integração de dados e código às publicações, por exemplo, é uma tendência e a Embrapa também deve se preocupar com isso, uma vez que é produtora de conhecimento científico e é mantida com financiamento público. A garantia de reprodutibilidade das pesquisas, o compartilhamento dos dados e sua adequada administração devem ser suportados por uma e-infraestrutura de qualidade.

Segundo relatório sobre ciberinfraestrutura para as Ciências Biológicas (WOOLEY, 2003), uma de suas principais funções é prover: ferramentas para captura, armazenamento e administração de dados; ferramentas para organizar, encontrar e analisar os dados para obter informações; a conexão de análises experimentais e teóricas, e sua interação com simulações e modelos baseados nessa informação; e ainda a integração de diferentes aspectos dessa informação para fornecer um repositório de conhecimento⁵.

Para se extrair informação e conhecimento relevantes a partir de dados brutos, muitas ferramentas novas são necessárias.

⁵ As diferenças entre dado, informação e conhecimento são apontadas por Setzer (2001). Resumidamente, enquanto dado é uma sequência de símbolos quantificáveis e é puramente sintático, informação envolve a compreensão, tem significado, contém semântica. Conhecimento já não depende apenas de uma interpretação pessoal, como a informação, pois requer uma vivência do objeto do conhecimento. Portanto, este último termo é caracterizado como uma abstração interior, pessoal, de algo que foi experimentado por alguém.

3.2 Software

Jim Gray (2009) já apontava que muitos dados coletados não são sistematicamente tratados ou publicados e é dever dos cientistas da computação criar ferramentas genéricas para os pesquisadores coletarem, analisarem e processarem os dados. Ao mesmo tempo, elas devem ser intuitivas, fáceis de usar e robustas para que sejam amplamente usadas pelos cientistas das diversas áreas (LEHNING et al., 2009). O desafio, portanto, consiste em criar ferramentas computacionais genéricas que possam ser facilmente adaptadas para a utilização em vários domínios, garantindo que os pesquisadores, independentemente de sua área de atuação, tenham a sensação de que a ferramenta foi construída e configurada especificamente para seu uso.

O desenvolvimento de software, no entanto, apesar de não ser pesquisa básica, requer uma massa crítica de profissionais qualificados. É necessário entender que o desenvolvimento e o suporte de software é, pelo menos, tão importante para a ciência moderna quanto telescópios e aceleradores e que há uma grande diferença entre um protótipo e o software que constitui infraestrutura de fato. Este *gap* de engenharia de software continua a ser um problema na pesquisa em Computação, pois deve haver responsáveis por fazer a engenharia e a manutenção das ferramentas necessárias. Como esta é uma tarefa cara, pode ser vista como despesa de recursos que poderiam ser direcionados para a pesquisa. Essa falta de reconhecimento aos esforços de engenharia despendidos em tecnologia da informação para desenvolver a infraestrutura é considerada preocupante para o desenvolvimento da *e-Science* (RESEARCH COUNCILS UK, 2009).

Segundo Goodman e Wong (2009), os desafios relacionados à análise e à visualização são muito parecidos nas diferentes áreas científicas. Atualmente, os cientistas lidam com várias possibilidades para realizar uma mesma tarefa de análise, mas nenhum conjunto específico de ferramentas possibilita realizar todas as tarefas da maneira ideal. No futuro, os próprios cientistas, com a colaboração ou não de especialistas em dados, poderão projetar o software necessário, integrando aplicações modulares sob demanda. É necessário, portanto, prover formas de conectar e reusar componentes já existentes a fim de se projetar novas ferramentas para o avanço da descoberta em grandes volumes de dados. Um exemplo de sof-

ware que obteve sucesso nesse contexto é o Taverna (OINN et al., 2004), uma ferramenta para a composição de *workflows*⁶ em bioinformática.

O crescimento ininterrupto do volume de dados gera uma demanda crescente por melhores ferramentas que suportem todo o ciclo de pesquisa, desde a captura e curadoria dos dados até sua análise e visualização (GRAY, 2009). A e-infraestrutura cria melhores condições para o desenvolvimento, o suporte e a utilização dessas ferramentas, mas não é completa se não envolve capital humano de qualidade e adequada estrutura organizacional.

3.3 Capital humano

Os conceitos de Ciência da Computação são importantes não só para pesquisadores da área como para qualquer campo do conhecimento. O chamado pensamento computacional, conforme apresentado na seção 2, deve ser desenvolvido desde a formação básica. Além disso, como já foi destacado anteriormente, o sucesso de uma e-infraestrutura depende muito de uma cultura de participação ativa em diferentes ambientes e de compartilhamento de infraestrutura e de resultados. Também são fundamentais profissionais competentes tecnicamente e na coordenação de equipes. Qualquer investimento é insuficiente para a construção de uma infraestrutura científica se não envolve capital humano de qualidade.

3.4 Estrutura organizacional

Wilbanks (2009) rejeita a posição de que o grande volume de dados representa um quarto paradigma científico. Para ele, o que precisa ser destruído é a ideia de que cientistas existem como indivíduos isolados. Eles devem se ver como nós conectados que transmitem dados, testam teorias e acessam simulações de outros cientistas. Além disso, devem disponibilizar

⁶ Gouble e de Roure (2009) apontam as vantagens de workflows na e-Science.

seu conhecimento em rede, pois nada é tão rapidamente escalável quanto uma rede aberta. Para ilustrar sua eficiência, basta considerar a internet, uma rede pública que cresce sem uma autoridade central, permite a inovação sem pedidos de permissão e amplia e descobre mercados. Também é exemplo do poder de uma rede aberta o software livre, construído em modelo distribuído com a junção de pequenas contribuições e sob padronizações técnicas e legais. Portanto, os cientistas devem trabalhar em conjunto, e a melhor maneira de se fazer isso é por meio de redes.

Abbott (2009) também caracteriza uma organização que pode abrir caminho para abordagens integradoras em direção a um futuro sustentável. Tal organização deve integrar diferentes campos da Ciência, equilibrar ciência e tecnologia, focar no pensamento sistêmico, suportar abordagens flexíveis e interdisciplinares para resolver problemas de longo prazo, integrar a criação e o uso de conhecimento, e ainda ponderar realizações individuais e coletivas.

Bowker et al. (2007) utilizam uma interessante ideia de que as organizações também podem ser vistas como processadores de informação para afirmar que os mecanismos tecnológicos só podem substituir os organizacionais e humanos à medida que estes estejam preparados para essa substituição.

Mais do que simplesmente adotar tecnologia no local de trabalho, é necessário adotá-la apropriadamente. Nesse ponto, a infraestrutura institucional é tão importante quanto a infraestrutura tecnológica e, entre as questões mais importantes, podem ser enumeradas: criar sistemas e estruturas organizacionais que motivem os usuários a usar a tecnologia apropriadamente, adequar a tecnologia ao processo de trabalho de seus usuários e ajudar as pessoas a ter consciência de seu trabalho e coordená-lo com o trabalho dos outros. Essas questões são estudadas em uma área denominada Trabalho Cooperativo Suportado por Computador - Computer Supported Cooperative Work (CSCW) (PRATT et al., 2004.).

3.5 Ambiente virtual de pesquisa

Inúmeros recursos podem fazer parte de um ambiente virtual de pesquisa (AVP), *virtual research environment* (VRE), como os enumerados a seguir:

- site de informações gerais;
- *blogs*;
- *wikis*;
- busca em bases de dados heterogêneas;
- softwares e tutoriais correspondentes;
- serviços de armazenamento e acesso a dados com segurança;
- serviços de computação de alto desempenho;
- canais de comunicação com integrantes da equipe;
- ambientes virtuais geoespaciais;
- fóruns de discussão;
- recursos para comunicação com áudio e vídeo;
- mecanismos de votação e de enquetes para *feedbacks* e auxílio na tomada de decisões;
- bibliotecas de publicações, de componentes de software e de modelos digitais;
- ferramentas de *workflow*;
- recursos para dispositivos móveis.

Um AVP pode ser visto como um *framework* em que ferramentas, serviços e recursos podem ser conectados. Tem potencial para a multidisciplinaridade, tanto em seu uso quanto em seu desenvolvimento e pode envolver também questões administrativas e de transferência de tecnologias (FRASER, 2005). Assim, permite uma redução drástica das barreiras de tempo e distância, tanto geográfica quanto disciplinar ou organizacional (RESEARCH COUNCILS UK, 2009). No entanto, um AVP não é efetivo se não estiver integrado às políticas e à infraestrutura de pesquisa existentes. Nesse caso, em vez de um ambiente de pesquisa, torna-se um portal Web sub-utilizado. Então, para se avaliar os benefícios de um AVP, deve-se levar em consideração a complexidade do desenvolvimento técnico e da mudança de cultura, a disponibilidade de uma e-infraestrutura básica e o suporte institucional, com suas estratégias, políticas e financiamento (FRASER, 2005).

Uma infraestrutura comum para um AVP, mesmo que seja apenas em nível de políticas e padrões acordados, facilita o reuso e o compartilhamento de

ferramentas, dados e resultados. Serviços e ferramentas genéricas, como, por exemplo, de comunicação, colaboração ou aquelas que possibilitam buscas inteligentes, podem fazer parte de um *framework* comum a vários projetos de pesquisa. Além disso, uma ferramenta criada para um fim específico tem potencial para ser usada em outros contextos.

No Reino Unido, o JISC (2011) é responsável por definir um *framework* comum e padrões associados, e incentivar seu uso para a construção de AVPs específicos. Com isso, procura-se maximizar os benefícios alcançados com os investimentos feitos nessa área. A ênfase desse centro, portanto, é na arquitetura e nos padrões, mais do que nas aplicações específicas (ALLAN et al., 2004; JISC, 2011).

É fundamental que a Embrapa também busque um *framework* comum adequado para dar suporte a seus inúmeros projetos de pesquisa. Depois disso, é necessário investir em AVPs dedicados a problemas específicos, como os relacionados à bioinformática, às mudanças climáticas e à biodiversidade.

3.6 Considerações sobre alguns termos relacionados a e-Science

Para concluir esta seção, faz-se necessário estabelecer diferenças claras entre os termos utilizados neste documento. Apesar de, muitas vezes, representarem conceitos diferentes em fontes bibliográficas distintas, os termos são bem definidos no presente trabalho para que não haja ambiguidade e o texto possa ser bem compreendido.

O termo *e-Science* foi amplamente discutido na seção 2 e representa a ciência desenvolvida com o mínimo de restrições de recursos computacionais, aí incluídos aspectos como estruturas organizacionais e capital humano. A e-infraestrutura, ou ciberinfraestrutura, é justamente o que fornece estes recursos que dão suporte não só à *e-Science*, como também ao comércio eletrônico (*e-commerce*), ao governo eletrônico (*e-government*) e à educação à distância (*e-learning*), por exemplo. Já o *Grid* é associado especificamente à infraestrutura de *e-Science*. Portanto, e-infraestrutura normalmente tem uma conotação de maior abrangência do que o *Grid*, apesar de ser normalmente associada a *e-Science*.

O termo “infraestrutura científica” é usado para designar o conjunto de recursos que permitem o desenvolvimento da Ciência. Logo, a e-infraestrutura é um de seus componentes, uma vez que laboratórios de Biologia ou Química, por exemplo, não fazem parte de uma e-infraestrutura, mas compõem a infraestrutura científica. No entanto, a e-infraestrutura serve não só à Ciência, mas também a outras áreas, como o comércio eletrônico, conforme descrito anteriormente.

A palavra “plataforma” é bastante genérica e chegou a ser usada, neste texto, para envolver, além de recursos técnicos, aspectos organizacionais e de recursos humanos, conforme foi utilizada pelo Research Councils UK (2009). Porém, alguns autores, como Pettit et al. (2010), utilizam o termo “plataforma de *e-Science*” para representar o ambiente virtual de pesquisa, o que é apenas um elemento, ainda que importantíssimo, da infraestrutura científica. Então, é mais adequado utilizar, para esse fim, o termo “ambiente virtual de pesquisa” do que “plataforma de *e-Science*”. Este último é utilizado de uma maneira mais genérica para tudo o que possa representar uma base para o desenvolvimento de *e-Science*.

Em relação aos idiomas, para os termos em que a tradução pode ser feita sem perda de significado ou com palavras parecidas, casos de e-infraestrutura e ciberinfraestrutura, foram utilizados termos na língua portuguesa. No entanto, foram mantidos termos que aparecem com frequência em seu idioma original mesmo em documentos de língua portuguesa, como é o caso de *e-Science*. Além disso, para se garantir uma maior precisão e facilitar novas pesquisas na literatura internacional, normalmente, neste documento, os termos traduzidos foram acompanhados de seus correspondentes na língua original, pelo menos em sua primeira ocorrência.

3.7 Futuro da e-infraestrutura e da *e-Science*

Para finalizar, são apresentadas algumas considerações sobre o futuro da e-infraestrutura e da *e-Science*. Tende a desaparecer o prefixo ‘e’ desses termos, que normalmente significa ‘eletrônico’, e ‘melhorada’ no caso de *e-Science* (ver seção 2). Groth e MacKie-Mason (2010) afirmam que ‘*e-commerce*’ (comércio eletrônico), de um novo campo, passou a ser absorvido por ‘*commerce*’ (comércio) em que a internet e a computação

correspondem simplesmente a tecnologias em uma complexa infraestrutura.

Muitos termos ainda tornam explícito o caráter eletrônico e a relação com TI por que a infraestrutura digital ainda está em fase de desenvolvimento, mas quando ela de fato tornar-se ubíqua e pervasiva a ponto de não se perceber mais a existência desse suporte, a ênfase no aspecto tecnológico será dissipada.

Desta maneira, a partir do momento em que houver uma e-infraestrutura mundial consolidada e a colaboração no contexto científico passar a ser regra geral, a chamada *e-Science* será equivalente à Ciência, sem a ênfase nas abordagens computacionais, uma vez que a e-infraestrutura passará a ser condição para se fazer Ciência de qualidade.

No entanto, para se chegar a esse ponto, é necessário muito esforço, investimento e tempo. A adequada construção de uma e-infraestrutura e seu uso apresentam inúmeros desafios. Alguns caminhos são apontados por Atkinson et al. (2002), que enumeram as seguintes necessidades: atenção às questões semânticas, sistemas ubíquos e confiáveis, computação autônoma⁷ para reduzir custos e complexidade, e rápida construção de conjuntos de serviços personalizados. No que se refere ao uso, destacam as pesquisas para novas formas de comunidade e para a maior compreensão dos impactos sócio-econômicos gerados por essa infraestrutura.

O objetivo primordial das tecnologias atuais é armazenar, gerenciar, indexar e computar dados de maneira eficiente (PARASTATIDIS, 2009), o que está mais relacionado à Ciência Computacional, conforme já discutido na seção 2. No entanto, há um longo caminho ainda para a criação de sistemas computacionais que possam automaticamente descobrir, adquirir, organizar, analisar, correlacionar, interpretar, inferir e raciocinar sobre a informação disponível na internet, oculta nos discos rígidos dos pesquisadores ou ainda nos cérebros das pessoas. Tampouco há infraestrutura capaz de gerenciar e processar conhecimento em uma escala global (PARASTATIDIS, 2009). Esses recursos estão associados à *e-Science*, o que torna claro que ainda não há limites estabelecidos para seu avanço. Portanto, ainda há muito a se fazer até mesmo na chamada Ciência

⁷ A computação autônoma lida com sistemas que se auto-gerenciam com o mínimo de interferência humana (ATKINSON et al., 2002).

Computacional, mas a *e-Science* é que deve guiar as decisões relacionadas à Computação na Ciência.

4 Propostas

Nos últimos anos, foram lançados, em vários países, grandes programas nacionais relacionados a *e-Science* com o objetivo de se construir a infraestrutura necessária para uma pesquisa multidisciplinar e colaborativa. O Brasil também deve dar atenção a esse tema, e a Embrapa, que visa a ser referência na viabilização de soluções de tecnologia da informação no âmbito das instituições públicas de Ciência, Tecnologia e Inovação (CT&I) (EMBRAPA, 2010), deve assumir um papel protagônico no país, com destaque para a unidade Embrapa Informática Agropecuária, principal responsável pelas pesquisas em Computação na organização.

Com um suporte computacional adequado, os pesquisadores da Embrapa podem focar suas atividades no cumprimento da missão da empresa: viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira. Portanto, uma plataforma de *e-Science* que atenda às suas necessidades é fundamental para que a Embrapa torne-se líder mundial na geração de conhecimento, tecnologia e inovação para a produção sustentável de alimentos, fibras e agroenergia.

Para apresentar propostas relacionadas a *e-Science*, é importante levar em consideração o que foi discutido na 4ª Conferência Nacional de Ciência, Tecnologia e Inovação para o Desenvolvimento Sustentável (CNCTI) e as estratégias estipuladas nos planos diretores da Embrapa. Com base nas prioridades estratégicas apontadas, propõe-se o desenvolvimento de uma plataforma de *e-Science* que, inicialmente, tem o objetivo de atender aos pesquisadores de bioinformática da Embrapa, mas que tem o potencial de servir à empresa toda e a parceiros espalhados pelo globo em aplicações diversas.

Ainda são apresentados, nesta seção, maiores detalhes sobre a estrutura de pesquisa da Embrapa Informática Agropecuária, unidade que ficaria responsável por tal plataforma, e sobre o Laboratório Multiusuário

de Bioinformática, que seria o pioneiro em sua utilização. Finalmente, é sugerido o oferecimento de cursos sobre *e-Science* na Universidade, uma vez que a formação de recursos humanos é parte essencial de qualquer iniciativa envolvendo *e-Science*.

4.1 Embrapa, *e-Science* e a 4ª CNCTI

Propostas relacionadas à *e-Science* estão totalmente alinhadas com as diretrizes dos planos diretores da Embrapa e com as recomendações feitas durante a 4ª Conferência Nacional de Ciência, Tecnologia e Inovação para o Desenvolvimento Sustentável (CNCTI). Realizada em maio de 2010 com mais de 4 mil participantes, a conferência foi convocada por decreto presidencial para discutir uma política de Estado para ciência, tecnologia e inovação com vistas ao desenvolvimento sustentável, e foi precedida de encontros estaduais, fóruns de discussão e conferências regionais. O chamado Livro Azul (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010b) sintetiza as principais contribuições da 4ª CNCTI, que teve como fios condutores o desenvolvimento sustentável e a inovação. Acompanha este volume a consolidação das recomendações da conferência (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010a), que apresenta um conjunto de sugestões feitas a partir dos pontos que se tornaram recorrentes nas discussões, dos principais desafios a serem enfrentados na formulação da política nacional de ciência, tecnologia e inovação para os próximos anos.

O Livro Azul (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010b) afirma que nenhum país que aspire a ser moderno e desenvolvido pode abrir mão de investir seriamente na área de Tecnologias de Informação e Comunicação (TICs). Recomenda, além do fortalecimento da pesquisa e desenvolvimento, investimentos na infraestrutura e na formação de recursos humanos na área. Essas recomendações são fortemente relacionadas a *e-Science*, que também foi explicitamente citada como foco de investimentos.

Na seção 2, foram apresentadas as principais características de *e-Science*. A primeira delas foi o uso de uma nova geração de infraestrutura científica, considerada em maiores detalhes na seção 3 deste documento. Esse tema

foi levado em consideração na 4ª CNCTI e nos documentos estratégicos da Embrapa, assim como as outras das principais características de *e-Science*: colaboração e multidisciplinaridade, grande volume de dados e o papel fundamental desempenhado pela Computação.

4.1.1 Infraestrutura científica

O V Plano Diretor da Embrapa (EMBRAPA, 2008a) aponta como ameaças importantes a insuficiência de investimentos em infraestrutura para o desenvolvimento de PD&I nas áreas de fronteira do conhecimento e o risco de obsolescência da infraestrutura laboratorial de PD&I agropecuária do Brasil. Com isso, estabelece a diretriz de assegurar a atualização contínua dos processos de PD&I e da infraestrutura, associando-a à estratégia de assegurar a manutenção, a atualização e a máxima utilização da infraestrutura laboratorial, de tecnologia da informação e dos campos experimentais (EMBRAPA, 2008a).

Além disso, dos quatro grupos de recomendações (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010a) para a ampliação e fortalecimento da base de pesquisa científica e tecnológica nacional, um foi dedicado à ampliação e à modernização da infraestrutura de pesquisa. Entre as recomendações, foram enumerados: acréscimo de investimentos em infraestrutura; criação de novas instalações de uso multi-institucional, especialmente em áreas estratégicas como a biotecnologia; fortalecimento do papel das unidades de pesquisa ligadas aos ministérios que forneçam a infraestrutura adequada, promovam ações estruturantes e sirvam de âncoras para grandes projetos científicos e tecnológicos de interesse da sociedade brasileira.

A construção de uma plataforma de *e-Science* contribui significativamente para a atualização da infraestrutura científica. Além disso, facilita a interação entre pesquisadores e permite o uso multi-institucional de recursos, a fim de melhor aproveitá-los. Nesse cenário, instituições de pesquisa como a Embrapa têm papel fundamental na atuação em grandes projetos científicos.

4.1.2 Colaboração e multidisciplinaridade

A segunda característica de *e-Science* apontada na seção 2 foi seu caráter colaborativo e multidisciplinar. Uma das recomendações do Livro Azul (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010b) é a de promover sinergia entre instituições, projetos e recursos humanos para a ciência e a tecnologia, um dos principais objetivos de uma plataforma de *e-Science*. A sinergia é essencial não só para impedir o desperdício de recursos mas, sobretudo, porque é básica para a inovação, que depende de interações e trocas.

Tal plataforma também pode ser vista como implementação de diversas estratégias enumeradas no Plano Diretor da Embrapa (EMBRAPA, 2008a) associadas à diretriz de ampliação da atuação em redes para aumentar a sinergia, a capacidade e a velocidade da inovação, a fim de se enfrentar os desafios institucionais e organizacionais da Embrapa. As estratégias sugeridas envolvem a criação de mecanismos para intensificar a atuação da empresa em ambientes multidisciplinares e multi-institucionais; a promoção de arranjos institucionais e metodologias inovadoras; a estimulação da cultura de cooperação e organização de alianças estratégicas, e o desenvolvimento de competências organizacionais e infraestrutura para a inovação tecnológica em rede; ampliação da cooperação tecnológica, científica e organizacional com empresas e instituições mais inovadoras. Também constitui estratégia da empresa para o fortalecimento da comunicação institucional promover ações e produtos que integrem a comunicação interna, criando um ambiente cooperativo e de livre circulação de ideias (EMBRAPA, 2008a).

A Cooperação, citada frequentemente no Plano Diretor da Embrapa, é um dos valores cultivados na empresa (EMBRAPA, 2008a). Assim, são valorizadas as atitudes cooperativas, a construção de alianças institucionais e a atuação em redes para compartilhar competências e ampliar a capacidade de inovação. Apesar de “colaboração” e “cooperação” serem termos usados frequentemente como sinônimos, Bair (1989) aponta algumas diferenças entre eles. Ainda assim, ambos envolvem o trabalho em grupo para o cumprimento de objetivos comuns. E esse trabalho em grupo pode ser amplamente auxiliado pela Computação.

O I Plano Diretor de Tecnologia da Informação (PDTI) da Embrapa também considera estratégico implementar mecanismos que favoreçam um

trabalho integrado e colaborativo no atendimento às demandas de TI (EMBRAPA, 2010).

4.1.3 Computação e TICs

Outra característica fundamental da *e-Science* é o papel essencial da Computação e das TICs na pesquisa científica. Para essa área, considerada estratégica na 4ª CNCTI, foram sugeridos: formulação de programas de pesquisa e desenvolvimento estáveis, de longo alcance e em uma escala compatível com a do papel das TICs na sociedade brasileira; criação e fortalecimento de infraestrutura sofisticada de TICs avançadas no Brasil para estimular a inovação e a competitividade; promoção de ações voltadas à Integração Digital Global; investimentos em *e-Science*; lançamento de editais específicos para o desenvolvimento e infraestrutura em: a) processamento de alto desempenho; b) programação paralela; c) arquiteturas de processadores de alto desempenho e baixo consumo de energia.

Apesar de *e-Science* aparecer em separado, todos os outros itens citados relacionam-se diretamente a *e-Science*, cujos objetivos são sempre ambiciosos e visam justamente a prover infraestrutura para o avanço da Ciência e a integração entre pesquisadores. Para isso, é necessário, entre outros fatores, investir no aumento de poder computacional.

4.1.4 Grande volume de dados

Já está claro que a quantidade de dados disponíveis atualmente está muito além da capacidade de análise dos pesquisadores. É necessário, portanto, um esforço no sentido de tirar o máximo proveito desses dados, o que demanda cada vez mais o auxílio da Computação. A Figura 3.1, que apresenta o ciclo de vida dos dados, ajuda a estruturar e a pensar de forma integrada os diferentes laboratórios que compõem a Embrapa Informática Agropecuária, principal responsável por pesquisas em Computação na empresa. Essa abordagem é bastante adequada, uma vez que a ciência atual, conforme já exposto na seção 2, é baseada em dados. Na subseção 4.3, são explicadas as principais responsabilidades de cada um dos laboratórios presentes na unidade e de que forma eles se relacionam com o ciclo de vida dos dados.

4.2 Prioridades estratégicas

Na seção 2, *e-Science* também foi definida como a colaboração global em áreas chave da Ciência e a próxima geração de infraestrutura que vai possibilitá-la. Nesse caso, enfatizou-se a importância de áreas estratégicas para o desenvolvimento da Ciência.

A Embrapa atua de maneira direta na maioria das áreas e tecnologias consideradas, pela CNCTI (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010a), estratégicas para o desenvolvimento sustentável brasileiro, a saber: mitigação e adaptação aos efeitos das mudanças climáticas e preservação ambiental; uso sustentável da biodiversidade brasileira; mapeamento e gestão integrada dos recursos hídricos e minerais; desenvolvimento e difusão de tecnologias de uso e manejo do solo; fortalecimento do agronegócio, agropecuária, aquicultura e pesca; sistemas urbanos sustentáveis; pesquisa em saúde e fármacos; fontes alternativas de energia, bionergia e combustíveis fósseis; nanotecnologia; TICs e microeletrônica. Ganham destaque as áreas relacionadas à agricultura e às TICs (CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO, 2010b), focos da Embrapa Informática Agropecuária.

Além disso, o Plano Diretor da Embrapa (2008a) sugere a intensificação de PD&I em temas de ciência e tecnologia estratégicos para o Brasil, enfatizando, inicialmente, nanotecnologia, bioinformática, biotecnologia, biomateriais, bioeconomia, simulação, modelagem, automação, tecnologia da informação, agricultura e zootecnia de precisão e fundamentos da agricultura ecológica. A maioria destes temas é objeto de pesquisa na Embrapa Informática Agropecuária.

Os pesquisadores da Embrapa atuam na resolução dos problemas, fazendo parte dos chamados macroprogramas, que são atualmente divididos em 6 categorias. A primeira delas, formada pelos macroprogramas 1 (MP1), envolve “projetos de Pesquisa, Desenvolvimento e Inovação (PD&I) com base científica elevada, transdisciplinar e multi-institucional, abordando pesquisas de natureza básica, estratégica ou aplicada, mas de caráter estratégico, para atender aos grandes desafios nacionais, que exijam para sua execução arranjos institucionais complexos ou grandes redes, e aplicação intensiva de recursos” (EMBRAPA, 2008b).

Nota-se que, por suas características, os MP1 demandam plataformas de *e-Science* que facilitem o trabalho dos pesquisadores da Embrapa. E o próprio desenvolvimento de um programa nacional, nos moldes do conduzido no Reino Unido, por exemplo, poderia constituir um MP1, pois além de ser uma plataforma de pesquisa, *e-Science* é também objeto de pesquisa.

Antes de se propor um MP1 abordando um programa de *e-Science* amplo e que envolva tantos recursos, é possível desenvolver projetos que se enquadrem em macroprogramas 5 (MP5), que visam à promoção de avanços institucionais e ao aumento da capacidade de resolver problemas. Esse caminho de executar projetos MP5 com a possibilidade de utilizá-los como base para um MP1 permite que as pessoas envolvidas na organização passem a entender os conceitos e os benefícios relacionados a *e-Science*. Esse projeto pode nascer dentro da Embrapa Informática Agropecuária, tomar corpo, envolvendo outras unidades, e se expandir para instituições parceiras, inclusive internacionais, e até mesmo para o público geral.

4.3 Embrapa Informática Agropecuária

A Embrapa Informática Agropecuária é uma unidade de pesquisa temática da Embrapa cuja missão é “viabilizar soluções de pesquisa, desenvolvimento e inovação em tecnologia de informação para a sustentabilidade da agricultura, em benefício da sociedade brasileira” (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2008). Sua principal função é a de prover uma infraestrutura computacional para dar suporte às pesquisas desenvolvidas pela empresa e seus parceiros. Portanto, é natural que as iniciativas de *e-Science* partam da unidade.

A Embrapa Informática Agropecuária está organizada em laboratórios de pesquisa, que, nesta seção, serão associados às etapas do ciclo de vida dos dados (Figura 3.1) para que se permita uma visão mais integrada de sua estrutura. As descrições dos laboratórios, apresentadas a seguir, são baseadas no Relatório de Gestão 2005-2009 da Embrapa Informática Agropecuária (2009).

- O Laboratório de Bioinformática Aplicada (LBA) apoia os diversos projetos de melhoramento genético desenvolvidos na Embrapa e está envolvido na articulação da plataforma de bioinformática da

Embrapa, cujo objetivo é encontrar soluções em bioinformática para o desenvolvimento da biotecnologia focada na agropecuária. O LBA é o responsável pelo Laboratório Multiusuário de Bioinformática, apresentado em maiores detalhes mais adiante e cuja ideia é prover suporte computacional em todas as etapas do ciclo de vida dos dados, desde sua geração até a disponibilização de ferramentas e de informações relevantes.

- O Laboratório de Biologia Computacional (LBC) está voltado para pesquisas básicas, visando ao entendimento de regras fundamentais que regem o processo de dobramento tridimensional da estrutura e a especificidade da função das macromoléculas biológicas, com ênfase nos estudos sobre as relações entre sequência, estrutura e função das proteínas.
- O Laboratório de Modelagem Agroambiental atua na implantação da plataforma de mudanças climáticas e agricultura tropical, em projetos focados na redução do impacto dos efeitos do aquecimento global sobre a agricultura brasileira. É importante destacar que pesquisadores da unidade têm participado de relevantes eventos relacionados a mudanças climáticas e agricultura, como as COPs, Conferências da Organização das Nações Unidas sobre mudanças climáticas, e representado o país nas discussões sobre o tema. No contexto computacional, esse laboratório está ligado ao desafio proposto pela Sociedade Brasileira de Computação (2006) sobre modelagem computacional de sistemas complexos.

Esses três laboratórios foram montados para se resolver problemas específicos de grande porte. Portanto, necessitam de plataformas completas e demandam suporte em todo o ciclo de vida dos dados. Enquanto nestes a computação é aplicada em larga escala para resolver grandes desafios, nos demais laboratórios, os principais objetos de pesquisa são relacionados à Computação, a fim de se obter novos conhecimentos e tecnologias para aplicá-los a uma variedade de problemas na empresa.

- O Laboratório de Geotecnologias (LabGeo) tem por objetivos: desenvolver, validar e aplicar técnicas de sensoriamento remoto em culturas de interesse econômico para o Brasil, usar geotecnologias e técnicas de processamento de imagens digitais e disponibilizar ferramentas para a construção de aplicativos espaciais. Esse laboratório serve inúmeros projetos na empresa, uma vez que fornece suporte para

um conjunto de problemas muito comuns na Embrapa: aqueles que demandam informações geográficas. Na 4ª Conferência Nacional de Ciência, Tecnologia e Inovação (2010a), destacou-se a importância do desenvolvimento de aplicações em sistemas de informação para mapeamento, monitoramento e análise da dinâmica de uso, e cobertura vegetal das terras e dos indicadores espaciais, temporais e de desempenho da atividade agrícola, integrando tecnologias de sensoriamento remoto e geoprocessamento em ambiente da web. Esse tipo de atividade é desenvolvido no LabGeo, que está envolvido em todo o ciclo de dados relacionados a essa natureza de problema.

- O Laboratório de Inteligência Computacional (LabIC) dedica-se à pesquisa de métodos, técnicas e algoritmos na área de inteligência computacional e está fortemente associado à análise de dados, etapa que talvez abrigue os maiores desafios da *e-Science*. Ferramentas de mineração de dados e de textos e técnicas de aprendizado de máquina, por exemplo, são desenvolvidas e aplicadas no laboratório e podem ser aplicadas a uma grande variedade de problemas.
- O Laboratório de Matemática Computacional (LabMaC) (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2011b), o mais recente, pesquisa, desenvolve e inova métodos, técnicas e algoritmos nas áreas de modelagem matemática, estatística aplicada, simulação e otimização de sistemas, a fim de prover ferramentas para análise de decisão e soluções para problemas complexos do setor agropecuário. Considerando a diversidade dos temas abordados nas diferentes unidades de pesquisa da Embrapa e a transversalidade da aplicação desse tipo de ferramenta, o laboratório visa a servir de incentivo, auxílio e referência a projetos de P&D que necessitem do uso da matemática computacional. O LabMaC está associado, portanto, principalmente, à análise de dados.
- O Laboratório de Organização e Tratamento da Informação Eletrônica dedica-se ao desenvolvimento de metodologias para organização, estruturação e tratamento da informação digital e desenvolve aplicativos para integração de dados e interoperabilidade de sistemas de informação. Percebe-se a importância desse trabalho quando conceituados pesquisadores afirmam que a heterogeneidade de plataformas, dados e aplicações é considerado um problema ainda mais difícil do que a escala dos recursos empregados (EMMOTT, 2006).

Além disso, muitas ferramentas são necessárias para que toda a literatura científica e todos os dados científicos fiquem online, e ainda haja interoperação entre eles (GRAY, 2009), o que corresponde a um dos objetivos da *e-Science*. Portanto, esse laboratório está associado, principalmente, à distribuição de dados e à preservação dos dados, com a criação de metadados, por exemplo, ou a migração para formatos adequados.

- O Laboratório de Software Livre (LabSoL) foca em ações de implantação de software livre, em conjunto com a oferta de soluções livres para a agricultura e a pecuária, viabilizando as metas da instituição na transferência de conhecimento e de tecnologias para a sociedade. Foi recomendação na Conferência Nacional de Ciência, Tecnologia e Inovação (2010b) adotar políticas de uso do software livre e de fomento ao desenvolvimento e uso de tecnologias abertas e interoperáveis. O LabSoL atua em todas as etapas do ciclo de vida dos dados, realizando a seleção e o desenvolvimento de software, além do oferecimento de treinamento aos usuários, de maneira a aumentar a produtividade na pesquisa, explorando o que há de melhor no mundo do software livre.
- O Laboratório de Novas Tecnologias (LabTec) tem a missão de pesquisar, avaliar e desenvolver soluções fundamentadas sobretudo em novas tecnologias que sirvam de suporte aos projetos de pesquisa da Embrapa Informática Agropecuária. Acompanha avanços e tendências em tecnologia da informação e provê infraestrutura tecnológica de apoio à atividade de pesquisa e de desenvolvimento dos projetos da unidade. Esse laboratório visa a contribuir para o avanço da fronteira do conhecimento e incorporar novas tecnologias, inclusive as emergentes, o que constitui desafios científicos e tecnológicos na Embrapa (2008a). Seu papel, portanto, é parecido com o do laboratório de software livre, no sentido de dar suporte a quem tiver necessidade, mas com a diferença de que não lida com software livre, mas com tecnologias emergentes.
- O atual Núcleo de Tecnologia da Informação (NTI), antes denominado Laboratório de Redes, Conectividade e Alto Desempenho, também provê suporte às pesquisas desenvolvidas na Embrapa Informática Agropecuária, mantendo os recursos físicos de rede e conexão e gerenciando a infraestrutura computacional. Esse laboratório tem um

papel fundamental na construção de uma plataforma de *e-Science*, uma vez que a maior parte do poder computacional e das capacidades de armazenamento e de comunicação fica sob sua responsabilidade. Em relação aos dados, suas principais contribuições são, além de oferecer poder de processamento e comunicação de dados, criar políticas de administração e compartilhamento, e armazená-los de maneira adequada.

Na Embrapa, muitos projetos visam ao desenvolvimento de uma infraestrutura científica para se resolver problemas específicos. Por exemplo, a Plataforma Nacional de Recursos Genéticos visa a melhorar o processo de gestão integrada de recursos genéticos de vegetais e evoluir a infraestrutura computacional do sistema Sibrargen para atender às melhorias incorporadas ao processo, e adotar tecnologias de informação modernas. Outro exemplo é a Plataforma de Recursos Naturais, cujo objetivo é prover uma infraestrutura que integre informação de recursos naturais dos biomas brasileiros, visando à geração de inteligência competitiva para a agricultura nacional (EMBRAPA INFORMÁTICA AGROPECUÁRIA, 2011a).

Muitas das tarefas realizadas em projetos diferentes são redundantes. Uma plataforma de *e-Science* genérica poderia aumentar a produtividade, unificando metodologias, reaproveitando resultados produzidos na solução de outros problemas e integrando fortemente as unidades da Embrapa e as divisões dentro das unidades. Somente para ilustrar, todos os projetos podem ter, por exemplo, uma política unificada de administração e compartilhamento de dados, ferramentas comuns que auxiliam na manutenção dos dados e na disponibilização de informações e ambientes de colaboração padronizados.

Apesar dos aspectos comuns, os diferentes desafios exigem recursos distintos. Plataformas dedicadas podem ser construídas para atender aos diferentes projetos, mas podem ter base em uma plataforma de *e-Science* genérica. Em um primeiro momento, pode ser considerada a construção de uma plataforma para a bioinformática, uma vez que já há um ponto de partida para tal iniciativa: o Laboratório Multiusuário de Bioinformática.

A Embrapa tem estimulado a criação de laboratórios multiusuários, para realizar testes e análises de alta complexidade científica, envolvendo equipe técnica multidisciplinar e equipamentos altamente especializados. Outro exemplo é o Laboratório Multiusuário de Química de Produtos Naturais.

A experiência com a bioinformática pode ser, posteriormente, estendida à química de produtos naturais e outros campos em que a Embrapa atua, seja com auxílio de laboratórios multiusuários ou não. A partir destas experiências, será mais fácil identificar o que deve constituir uma plataforma científica genérica e implantar novos laboratórios multiusuários. Assim, será possível atender melhor às demandas por e-infraestrutura da empresa toda.

4.4 Laboratório Multiusuário de Bioinformática

O Laboratório Multiusuário de Bioinformática visa a contribuir com trabalhos não só de centros de pesquisas da Embrapa distribuídos pelo País, como também de instituições parceiras públicas e privadas. A ideia é atuar principalmente no suporte ao processamento de dados genômicos e transcriptômicos dos programas de melhoramento genético. Sediado na Embrapa Informática Agropecuária, possui equipamentos de computação de alto desempenho, grande capacidade de armazenamento e ainda uma equipe de pesquisadores experientes.

O Laboratório Multiusuário de Bioinformática atende a demandas do País e da empresa, que possui pesquisadores de bioinformática em várias unidades. Tais demandas são expressas tanto nos planos diretores da Embrapa quanto nas recomendações da 4a CNCTI.

Do Plano Diretor da Embrapa (EMBRAPA, 2008a), destacam-se as seguintes estratégias:

- Assegurar o uso compartilhado de equipamentos, laboratórios, informações e campos experimentais entre técnicos, Unidades Descentralizadas (UDs), pesquisadores e parceiros para assegurar a manutenção, a atualização e a melhor utilização da infraestrutura laboratorial, de tecnologia da informação e dos campos experimentais.
- Estimular o compartilhamento de infraestrutura em laboratórios multiusuários de excelência.
- Articular o ambiente de cooperação em rede.
- Articular redes cooperativas, produtivas e sociais, com base em modelos de gestão ágeis e flexíveis para a construção de plataformas tecnológicas, preconizadas pela Empresa.

Os três últimos itens estão associados ao estímulo de uma cultura de cooperação e organização de alianças estratégicas, com a participação de agentes públicos e privados, no País e no exterior.

Além das recomendações, já apresentadas neste documento (seção 4.1.1) a respeito da ampliação e modernização da infraestrutura, a 4ª CNCTI fez recomendações relacionadas à agricultura e à bioinformática, como por exemplo:

- Fortalecimento da infraestrutura de C,T&I para o setor agrícola.
- Apoio ao desenvolvimento de atividades de P&D em nanotecnologia, biotecnologia e bioinformática para agricultura.
- Desenvolvimento de pesquisas de melhoramento genético.

Portanto, está muito claro que o laboratório constitui um recurso muito importante para o desenvolvimento de pesquisas consideradas estratégicas na Embrapa e no Brasil. Além disso, uma plataforma de *e-Science* construída para o Laboratório Multiusuário de Bioinformática pode ser modelo para se determinar uma plataforma genérica para outros problemas da Embrapa. Com isso, será possível personalizar a infraestrutura computacional para atender às necessidades específicas de cada projeto.

A capacidade de armazenamento é uma questão importantíssima no laboratório, uma vez que a quantidade de dados genômicos com que se trabalha atualmente é enorme, e só tende a aumentar. Artigo da *Wired* (ANDERSON, 2008) mostra a evolução dos dispositivos de armazenamento: *kilobytes* eram armazenados em disquetes, *megabytes* são armazenados em disco rígidos e *terabytes* em *disk arrays*. *Petabytes* na nuvem. Como a pesquisa científica já caminha em direção à escala dos *petabytes* (BELL et al., 2005; SZALAY; GRAY, 2006), seu futuro passa pela computação em nuvem.

É possível comparar custos de conectividade, armazenamento e processamento (GRAY, 2003) para determinar o que é melhor: levar a computação aos dados ou levar os dados à computação. Em relação ao Laboratório Multiusuário de Bioinformática, em princípio, os dados podem ser centralizados em máquinas localizadas fisicamente em apenas um lugar. No futuro, porém, devido à quantidade de dados e ao fato de dados serem coletados em diferentes pontos geográficos, provavelmente será melhor enviar computação aos dados em vez de levar os dados à computação, o que corresponde a uma das regras informais que Gray formulou como

princípios para projetar os sistemas do futuro baseados em grande volume de dados (SZALAY; BLAKELEY, 2009).

Mas toda essa capacidade de armazenamento também influencia o desenvolvimento de software. Devido à necessidade de processamento de dados em escala de ordem superior a *terabytes*, novos algoritmos mais eficientes também precisam ser desenvolvidos (BELL et al., 2005). O gerenciamento de computação distribuída também levanta novas questões de segurança, acesso livre a dados públicos e custos (SZALAY; GRAY, 2006).

Outra questão no futuro é a ampliação de parcerias, pois nenhuma organização detém todas as competências necessárias. Os desafios enfrentados atualmente são complexos e exigem composição de esforços, transdisciplinares e multi-institucionais, para gerar soluções para a sociedade (FONSECA, 2010). Então, deve-se prever também que a plataforma de *e-Science* evolua de maneira a incluir muitos parceiros, nacionais e internacionais.

Artigo da Nature Reviews Microbiology de 2008 (CRADDOCK et al., 2008) apontava que já havia mais de 3.000 bases de dados e ferramentas de bioinformática disponíveis na *Web Services*⁸ e destacou a importância dos *workflows*, que permitem a especificação e a execução de experimentos *in silico* sem a necessidade de habilidades em programação. Esse artigo conclui que a ênfase, até então, estava em prover poder computacional a análises de dados escaláveis, mas que os desafios futuros estavam no desenvolvimento de sistemas que gerem conhecimentos relevantes aos biólogos.

Portanto, um dos primeiros desafios do Laboratório Multiusuário de Bioinformática da Embrapa é selecionar e disponibilizar adequadamente os recursos computacionais que melhor atendam às necessidades da empresa. Também é importante permitir que os usuários utilizem com facilidade ferramentas de *workflow* para garantir maior padronização, melhor comunicação e, conseqüentemente, maior produtividade. Deve-se ter em vista a criação de novas ferramentas de análise que possibilitem o avanço do conhecimento e a criação de novas tecnologias, a fim de guiar o melho-

⁸ O próprio artigo de Craddock et al. (2008) define Web services como interfaces que permitem a disponibilização de recursos computacionais de uma maneira uniforme.

ramento genético das espécies pesquisadas na Embrapa. Tudo isso pode ser disponibilizado em um ambiente virtual de pesquisa, que além de oferecer os recursos associados especificamente aos problemas da bioinformática, deve prover aos pesquisadores um conjunto de outras ferramentas já discutidas na seção 3.

A infraestrutura necessária para o bom funcionamento de um laboratório multiusuário como esse deve contar com o auxílio de muitos profissionais. A Embrapa Informática Agropecuária pode oferecer muitos dos recursos necessários, envolvendo, idealmente, todos os laboratórios da unidade. Pode-se ter como referência a organização do programa de *e-Science* do Reino Unido (REDFEARN, 2007; RESEARCH COUNCILS UK, 2009), que, além de financiar inúmeros projetos relacionados a problemas específicos - mais de 100 só nos primeiros cinco anos - tem um núcleo responsável por: e-infraestrutura compartilhada, desenvolvimento de ferramentas genéricas, interoperabilidade, disseminação de boas práticas e transferência de tecnologia.

4.5 Formação de recursos humanos para *e-Science*

Já se reconhece amplamente a necessidade de que cientistas da computação e cientistas de outras áreas devam trabalhar colaborativamente, e não em relações de subordinação (HEY et al., 2009). Porém, a dificuldade é justamente implementar medidas que propiciem uma cultura de colaboração e interdisciplinaridade, uma vez que nem a Universidade prepara adequadamente os profissionais nesse aspecto.

A estrutura rígida em torno de disciplinas e departamentos pode ser barreira para a inovação e para a necessidade de se educar estudantes para um ambiente de trabalho mais colaborativo. Assim, podem ser incentivados trabalhos interdisciplinares baseados em grandes desafios, e não em campos delimitados ou com pouca conexão com o mundo real (HENNESSY, 2010; EMMOTT, 2006).

Além disso, a estrutura interna de gratificação das universidades e as revisões por pares favorecem projetos de pesquisa que são de interesse da comunidade científica, mas que não atendem necessariamente à comunidade fora da Academia. Um resultado disso é a desconexão entre

o conhecimento científico gerado e a demanda por esse conhecimento por parte dos setores governamental e privado (ABBOTT, 2009).

Outro ponto relevante é o desenvolvimento do chamado pensamento computacional nas inúmeras áreas do conhecimento. A própria Conferência Nacional de Ciência, Tecnologia e Inovação (2010a) recomendou a ampliação de programas de qualificação profissional na área de TICs, com ações voltadas à qualificação da força de trabalho para as habilidades da sociedade da informação, citando explicitamente o termo *computational thinking*. Recomendou inclusive a introdução de disciplinas específicas nos ensinamentos fundamental e médio. Assim, da mesma maneira que a matemática está no currículo escolar desde o começo da educação formal, os conceitos fundamentais da computação também devem fazer parte da formação, o que possibilitará às pessoas uma melhor abordagem para se resolver problemas complexos.

As universidades têm que avançar as fronteiras do conhecimento e educar estudantes (HENNESSY, 2010). A Embrapa pode ajudar nisso, uma vez que é seu dever avançar as fronteiras do conhecimento e a empresa conta com um corpo de colaboradores altamente capacitado, que tem um enorme potencial para educar pessoas. Mais especificamente, a Embrapa Informática Agropecuária (2008) estimula a intensificação de parcerias com instituições de pesquisa e ensino, do Brasil e do Exterior, por intermédio de promoção de cursos e visitas técnicas, com o objetivo de ampliar e intensificar os esforços de formação e aperfeiçoamento de recursos humanos em PD&I e gestão empresarial.

Portanto, a Embrapa Informática Agropecuária tem tudo para contribuir nesse novo cenário, pois consegue facilmente levar grandes problemas para a Universidade a fim de buscar soluções com a colaboração de estudantes e pesquisadores de várias áreas. Desta maneira, pode criar pontes entre as demandas agropecuárias e o conhecimento gerado na Universidade, ajudar a desenvolver o pensamento computacional em diferentes áreas e ainda estimular, na Universidade, um ambiente mais propício à colaboração, à interdisciplinaridade e à inovação. Tudo isso pode começar a ser feito por meio de um curso de *e-Science* em parceria com instituições de ensino e pesquisa

5 Conclusão

A Computação já tem sua importância para a Ciência amplamente reconhecida. A aplicação de ferramentas computacionais é fundamental para qualquer área do conhecimento e simulações computacionais estão presentes em um amplo conjunto de problemas relevantes para a humanidade. No entanto, muitos avanços ainda são desejáveis. A formação de profissionais de todas as áreas dotados de capacidades relacionadas ao pensamento computacional (ver capítulo 2) ainda é um projeto de longo prazo, mas a melhoria da infraestrutura científica pela e-infraestrutura é urgente nesse cenário em que o conhecimento é gerado a partir de grande volume de dados.

A *e-Science* não é um modismo, mas um termo que sintetiza ideias que são fundamentais para a pesquisa científica. Também não é algo que se resume a um sistema computacional específico ou a um conjunto de recursos, mas um termo quase tão abrangente quanto 'Ciência'. A credibilidade dos pesquisadores envolvidos com o tema não permite a conclusão de que se trata de assunto passageiro ou irrelevante. É necessário dar muita atenção aos trabalhos que vêm sendo realizados na área e utilizá-los para direcionar as decisões que norteiam a pesquisa.

O termo *e-Science* talvez deixe de ter importância no momento em que a computação tornar-se ubíqua e pervasiva na Ciência, mas até isso ocorrer, o que levará bastante tempo ainda, deve estar na agenda dos pesquisadores de Computação e de outras áreas. Neste ponto, a computação deixará de ser preocupação para os pesquisadores, para constituir uma infraestrutura científica consolidada. Já aos cientistas da Computação caberá o desafio de encontrar novas formas de aprimorar o avanço da Ciência e surpreender a humanidade com mais soluções inovadoras. Porém, antes de se explorar todo o potencial da integração entre Computação e demais ciências, propósito da *e-Science*, ainda há muito a se fazer na chamada Ciência Computacional, mais relacionada à aplicação da computação aos problemas específicos presentes em todas as áreas do conhecimento.

Os benefícios de uma cultura de compartilhamento, cooperação e de interdisciplinaridade já foram identificados há muito tempo. No entanto, faltam atitudes concretas e políticas que estimulem essa cultura para potencializar os avanços técnicos-científicos. A *e-Science* estabelece o

compromisso de se atender, da melhor maneira possível, às necessidades computacionais dos pesquisadores, o que demanda o trabalho conjunto de profissionais de Computação com cientistas de todas as áreas para auxiliar na pesquisa e no desenvolvimento de soluções para seus problemas.

As direções apontadas pela *e-Science* estão alinhadas às diretrizes expressas nos documentos estratégicos da Embrapa e do País, além de acompanharem o que vem sendo recomendado pela comunidade científica internacional. Também estão fortemente associadas aos grandes desafios apontados pela SBC para a Computação no Brasil para o período de 2006 a 2016. Portanto, as práticas de *e-Science* podem auxiliar a Embrapa a resolver grandes problemas nacionais rumo à sustentabilidade. Pode-se dizer que elas constituem um pré-requisito para o desenvolvimento da ciência moderna, mas, infelizmente, o Brasil está atrasado nesse ponto. É necessário que o país tome providências para se inserir cada vez mais nesse cenário e assumir um papel cada vez mais importante na ciência mundial.

Infraestrutura e recursos humanos estão, atualmente, entre os principais gargalos para o crescimento sustentável da economia brasileira. Por isso, o governo federal vem dando ênfase ao PAC, programa de aceleração do crescimento, que visa ao desenvolvimento sustentável do País pelos investimentos em infraestrutura. Em uma época em que a economia é baseada no conhecimento e na inovação, investir em infraestrutura científica é fundamental.

Hardware e conectividade têm ficado cada vez mais baratos, apesar da crescente demanda, mas recursos humanos na área de Computação são escassos para atender com qualidade às necessidades atuais. Os profissionais são os responsáveis por desenvolver software, otimizar o uso do hardware disponível, criar as estruturas organizacionais adequadas, prover os recursos de dados e informações e promover uma cultura multidisciplinar de colaboração, e se adaptar a ela. Porém, a questão educacional tem sido um problema para o desenvolvimento do País e é dever de todos participar ativamente na solução desse grande desafio nacional, principalmente das instituições ligadas à geração de conhecimento, que demandam profissionais altamente qualificados e têm capacidade de oferecer treinamentos e auxílio na formação de pessoas, caso da Embrapa. A Embrapa Informática Agropecuária pode contribuir oferecendo cursos relacionados a *e-Science* na Universidade.

e-Science pode ser tema de vários projetos de pesquisa na Embrapa que visem ao desenvolvimento institucional, e ainda tem potencial para ser foco de um grande desafio nacional, assim como foi no Reino Unido, por exemplo. É necessário agir agora, uma vez que as primeiras iniciativas já datam de mais de uma década e as principais potências científicas também já desenvolvem os seus programas de *e-Science*.

O desenvolvimento de projetos de *e-Science* na Embrapa Informática Agropecuária ajuda a consolidar a identidade da unidade, cuja missão envolve prover recursos computacionais aos pesquisadores da empresa. Talvez, *e-Science* seja o termo que melhor caracteriza as atividades desenvolvidas na unidade. Mais do que informática, a unidade faz computação (ver discussão no capítulo 2) envolvendo o trabalho cooperativo de pesquisadores de vários segmentos. Ao compreender *e-Science*, cientistas e gestores da Embrapa Informática Agropecuária e de outras unidades podem perceber o tamanho dos desafios que se têm pela frente e que a Embrapa Informática Agropecuária é fundamental para buscar as soluções necessárias.

O Laboratório Multiusuário de Bioinformática da Embrapa vem de encontro às necessidades da empresa e da nação por melhoramento genético na agropecuária. A bioinformática é considerada estratégica no país e tem concretizado as ideias defendidas pela *e-Science* em todo o mundo. Portanto, o laboratório pode ser um modelo a ser seguido pela empresa toda, pelo País, e até mesmo pela comunidade internacional. Um bom ponto de partida é a criação de um ambiente virtual de pesquisa, que possa atender às demandas de bioinformática, mas que também possa constituir a base para um *framework* comum para ambientes virtuais de pesquisa em qualquer área.

É importante destacar que a cooperação entre os laboratórios da Embrapa Informática Agropecuária é essencial para a construção de uma plataforma de *e-Science*, devido à sua complexidade. Por isso, procurou-se mostrar o papel de cada laboratório e uma visão integrada da unidade.

Este documento procurou explicar as principais ideias relacionadas à *e-Science* e contextualizá-las no Brasil e na Embrapa, no presente e no futuro. Especificamente para a Embrapa Informática Agropecuária, também aponta caminhos que podem auxiliar na tomada de decisão.

6 Referências

ABBOTT, M. R. A New Path for Science? In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p. 111-116.

ALLAN, R.; ALLDEN, A.; BOYD, D.; CROUCHLY, R. **Roadmap for a UK Virtual Research Environment**. 2004. Report of the JCSR VRE Working Group. Disponível em: <http://www.jisc.ac.uk/uploaded_documents/VRE%20roadmap%20v4.pdf>. Acesso em: 31 jan. 2011.

ANDERSON, C. The end of theory: the data deluge makes the scientific method obsolete. **Wired Magazine**, 23 jun. 2008. Disponível em: <http://www.wired.com/science/discoveries/magazine/16-07/pb_theory>. Acesso em 25 jan. 2011.

ATKINSON, M.; CROWCROFT, J.; GOBLE, C.; GURD, J.; RODDEN, T.; SHADBOLT, N.; SLOMAN, M.; SOMMERVILLE, I.; STOREY, T. **Computer challenges to emerge from eScience**. 2002. Disponível em: <<http://www.semanticgrid.org/documents/>>. Acesso em: 1 fev. 2011.

ATKINSON, M.; BRITTON, D.; COVENEY, P.; DE ROURE, D.; GARNETT, N.; GEDDES, N., GURNEY, R.; HAINES, K.; HUGHES, L.; INGRAM, D.; JEFFREYS, P.; LYON, L.; OSBORNE, I.; PERROTT, R.; PROCTER, R.; RUSBRIDGE, C.; TREFETHEN, A. A; WATSON, P. Century-of-Information Research (CIR): a strategy for research and innovation in the Century of Information. **Prometheus**, v. 27, n. 1, p. 27-45, 2009. Disponível em: <<http://dx.doi.org/10.1080/08109020802657479>>. Acesso em: 26 jan. 2011.

BAIR, J. H. Supporting cooperative work with computers: addressing meeting mania. In: IEEE COMPUTER SOCIETY INTERNATIONAL CONFERENCE, 34., 1989, San Francisco. **Proceedings ...** Los Alamitos: IEEE Computer Society, 1989. p. 208-217.

BALL, A. **Review of the state of the art of the digital curation of research data**. Bath: University of Bath, 2010. Project Report (ERIM Project Document erim1rep091103ab12).

BELL, G; GRAY, J; SZALAY, A. **Petascale Computational Systems: balanced cyberInfrastructure in a Data-Centric World**. 2005. Carta à NSF. Disponível em: <<http://research.microsoft.com/en-us/um/people/gray/>>. Acesso em: 08 fev. 2011.

BELL, G.; HEY, T.; SZALAY, A. Beyond the Data Deluge. **Science**, v. 323, n. 5919, p. 1297-1298, 6 mar. 2009. Disponível em: <<http://www.sciencemag.org/content/323/5919/1297>>. Acesso em: 26 jan. 2011.

BENIOFF, M. R. et al. **Computational Science: ensuring America's competitiveness**. Report to the President, President's Information Technology Advisory Committee. Washington, DC, 2005. Disponível em: <http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf>. Acesso em: 26 jan. 2011.

BIRNEY, E. et al. Prepublication data sharing. **Nature**, v. 461, n. 7261, p.168-170, 10 Sept. 2009. Disponível em: <<http://www.nature.com/nature/journal/v461/n7261/full/461168a.html>>. Acesso em: 28 jan. 2011.

BOWKER, G. C.; BAKER, K.; MILLERAND, F.; RIBES, D. Towards information infrastructure studies: ways of knowing in a networked environment. In: HUNSINGER, J.; ALLEN, M.; KLASRUP, L. (Ed.). **International Handbook of Internet Research**. Springer, 2007. p. 97-117.

BURTON, A; TRELOAR, A. Designing for Discovery and Re-Use: the 'ANDS Data Sharing Verbs' Approach to Service Decomposition. **The International Journal of Digital Curation**, v. 4, n. 3, 2009. p. 44-56.

CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O DESENVOLVIMENTO SUSTENTÁVEL, 4., 2010, Brasília, DF. **Consolidação das recomendações**. Brasília, DF: Ministério da Ciência e Tecnologia: Centro de Gestão e Estudos Estratégicos 2010a. Disponível em: <<http://www.cgee.org.br/publicacoes/livroazul.php>>. Acesso em: 31 jan. 2011.

CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O DESENVOLVIMENTO SUSTENTÁVEL, 4., 2010, Brasília, DF. **Livro azul**. Brasília, DF: Ministério da Ciência e Tecnologia: Centro de Gestão e Estudos Estratégicos, 2010b. Disponível em: <<http://www.cgee.org.br/atividades/redirect.php?idProduto=6820>>. Acesso em: 31 jan. 2011.

CRADDOCK, T.; HARWOOD, C. R.; HALLINAN, J.; WIPAT, A. e-Science: relieving bottlenecks in large-scale genome analyses. **Nature Reviews Microbiology**, v. 6, p. 948-954, Dec. 2008. Disponível em: <<http://www.nature.com/nrmicro/journal/v6/n12/full/nrmicro2031.html>>. Acesso em: 1 fev. 2011.

DOZIER, J; GAIL W. B. The Emerging Science of Environmental Applications. In: Hey, T.; Tansley, S.; Tolle, K. (Ed.). **The Fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p.13-19.

EMBRAPA. Departamento de Tecnologia da Informação. I **Plano Diretor de Tecnologia da Informação (PDTI) da Embrapa**: 2010-2011. Brasília, DF, 2010. 47 p.

EMBRAPA. Secretaria de Gestão e Estratégia. **V Plano Diretor da Embrapa**: 2008-2011-2023. Brasília, DF, 2008a. 44 p.

_____. **SEG – Sistema Embrapa de Gestão**. Edital 01/2008. 2008b.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Embrapa Informática Agropecuária**. 2011a. Disponível em: <<http://www.ctwiki.cnpia.embrapa.br/index.php/Apresentações>>. Acesso em: 1 fev. 2011.

_____. **Matemática Computacional – LabMac | Embrapa Informática Agropecuária**. 2011b. Disponível em: <<http://www.cnptia.embrapa.br/content/matematica-computacional-labmac.html>>. Acesso em: 1 fev. 2011.

_____. **IV Plano Diretor da Embrapa Informática Agropecuária: 2008-2011-2023**. Campinas, 2008. 123 p.

_____. **Relatório de gestão 2005-2009**. Campinas, 2009. 66 p. (Embrapa Informática Agropecuária. Documentos, 91). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/17873/1/relatorio-digital-reduzida.pdf>>. Acesso em: 31 jan. 2011.

EMMOTT, S. (Ed.) **Towards 2020 Science**. Cambridge: Microsoft Research, 2006. Disponível em: <<http://research.microsoft.com/towards2020science>>. Acesso em: 25 jan. 2011.

FONSECA, C. E. L. da. **O Sistema Embrapa de Gestão - SEG**. Trabalho apresentado na Reunião de Pesquisadores e Analistas do CNPTIA, Campinas, dez. 2010. 50 slides. Disponível em: <<http://www.ctwiki.cnptia.embrapa.br/index.php/Apresentações>>. Acesso em: 8 fev. 2011.

FRASER, M. Virtual research environments: overview and activity. **Ariadne**, n. 44, Jul. 2005. Disponível em: <<http://www.ariadne.ac.uk/issue44/fraser>>. Acesso em: 31 jan. 2011.

GETOV, V. e-science:the added value for modern discovery. **Computer**, v. 41, n.11, p.30-31, Nov. 2008.

GOODMAN, A. A.; WONG, C. G. Bringing the night sky closer: discoveries in the data deluge. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p. 39-44.

GOBLE, C.; DE ROURE, D. The impact of workflow tools on data-centric research. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. p.137-145.

GRAY, J. **Distributed computing economics**. Redmond: Microsoft Research, 2003. Relatório Técnico. Disponível em: <<http://research.microsoft.com/apps/pubs/default.aspx?id=70001>>. Acesso em: 1 fev. 2011.

_____. Jim Gray on eScience: A transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009, p. 17–31.

GROTH D. P.; MACKIE-MASON, J. K. Why an informatics degree?: Isn't computer science enough? **Communications of the ACM**, v. 53, n. 2, p. 26-28, Feb. 2010.

HENNESSY, J. L. 2020 Visions: universities. In: NORVIG, P; RELMAN, D. A.; GOLDSTEIN, D. B.; KAMMEN, D. M. **Nature**, v. 463, n. 7, p. 28-29, Jan. 2010.

HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm**: data-intensive scientific discovery. Redmon: Microsoft Research, 2009. Disponível em: <<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>>. Acesso em 25 jan. 2011.

HEY, T.; TREFETHEN, A. e-Science and its implications. **Philosophical transactions of the Royal Society of London**, Series A Mathematical, Physical and Engineering Sciences, v. 361, n. 1809, p. 1809-1825 Aug. 2003. Disponível em: <<http://rsta.royalsocietypublishing.org/content/361/1809/1809>>. Acesso em: 26 jan. 2011.

HODSON, S. Data-sharing culture has changed. **Research Information**, Dec. 2009/Jan.2010. Disponível em: <http://www.researchinformation.info/features/feature.php?feature_id=243>. Acesso em: 31 jan. 2011.

JISC: inspiring innovation. Disponível em: <<http://www.jisc.ac.uk/>>. Acesso em 31 jan. 2011.

LEHNING, M.; DAWES, N.; BAVAY, M.; PARLANGE, M.; NATH, S.; ZHAO, F. Instrumenting the earth: next-generation sensor networks and environmental science. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm**: data-intensive scientific discovery. Redmond: Microsoft Research, 2009. p. 45–51.

MACHADO, N. J. **Educação: projetos e valores**. 3 ed. Editora Escrituras, 2002. 168 p.

NORVIG, P. et al. 2020 Visions. **Nature**, v. 463, p. 26-32, 2010.

OINN, T.; ADDIS, M.; FERRIS, J.; MARVIN, D.; SENGER, M.; GREENWOOD, M.; CARVER, T.; GLOVER, K.; POCOCK, M. R.; WIPAT, A.; LI, P. Taverna: a tool for the composition and enactment of bioinformatics workflows. **Bioinformatics**, v. 20, n. 17, p. 3045-3054, 2004. Disponível em: <<http://bioinformatics.oxfordjournals.org/content/20/17/3045>>. Acesso em: 31 jan. 2011.

PARASTATIDIS, S. A Platform for all that we know: creating a knowledge-driven research infrastructure. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm**: data-intensive scientific discovery. Redmond: Microsoft Research, 2009. p.165-172.

PETTIT, C. J.; RUSSEL, A. B. M.; MICHAEL, A.; AURAMBOU, J.-P.; SHARMA, S.; WILLIAMS, S.; CHAN, P. C.; BORDA, A.; BISHOP, I. D.; ABRAMSOM, D. Realising an eScience Platform to Support Climate Change Adaptation in Victoria. IEEE INTERNATIONAL CONFERENCE ON E-SCIENCE, 6., 2010, Brisbane. **Proceedings...** IEEE Computer Society, 2010. p.73-80.

POTHEN, P. **Developing the UK's e-infrastructure for science and innovation**. 2006. Relatório do OSI e-Infrastructure Working Group. Disponível em: <<http://www.nesc.ac.uk/documents/OSI/>>. Acesso em: 28 jan. 2011.

PRATT, W. et al. Incorporating ideas from computer-supported cooperative work. **Journal of Biomedical Informatics**. Elsevier, v. 37, 2004. p.128–137.

REDFEARN, J. **Highlights from the UK e-science programme**. Engineering and Physical Sciences Research Council, 2007.

REGALADO, A. Brazilian Science: Riding a Gusher. **Science**, v. 330. n. 6009, 2010. p. 1306-1312. Disponível em: <<http://www.sciencemag.org/content/330/6009/1306>>. Acesso em 26 jan. 2011.

REPRODUCIBLE Research, **Computing in Science and Engineering**, v. 12, n. 5, p. 8-13, Sept./Ooct. 2010.

RESEARCH COUNCILS UK. **RCUK Review of e-Science 2009**. Report of the International Panel for the 2009 Review of the UK Research Councils e-Science Programme, 2009. Disponível em: <<http://www.epsrc.ac.uk/pubs/reports/Pages/internationalreviews.aspx>>. Acesso em: 08 fev. 2011.

_____. **About Research Councils UK**. Disponível em: <<http://www.rcuk.ac.uk/about/aboutRCUK/Pages/home.aspx>>. Acesso em 28/01/2011.

SACHS, J. 2020 Visions: global governance. In: NORVIG, P. et al. 2020 Visions. **Nature**, v. 463, 7 jan. 2010. p. 29.

SCIENCE in Brazil: go south, young scientist: an emerging power in research. **The Economist**, 6 Jan. 2011. Disponível em: <<http://www.economist.com/node/17851421>>. Acesso em: 26 jan. 2011.

SETZER, V. W. Dado, Informação, Conhecimento e Competência. In: _____. **Os meios eletrônicos e a educação**: uma visão alternativa. São Paulo: Escrituras, 2001. p. 239-275. (Coleção ensaios transversais, v. 10. Disponível em: <<http://www.ime.usp.br/~vwsetzer/dado-info.html>>. Acesso em: 31 jan. 2011.

SIPSER, M. **Introduction to the Theory of Computation**. 2nd ed. Thomson Course Technology, 2006.

SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Grandes Desafios da Pesquisa em Computação no Brasil – 2006-2016**, 2006. Disponível em: <http://www.sbc.org.br/index.php?option=com_jdownloads&Itemid=195&task=viewcategory&catid=50>. Acesso em: 14 jan. 2011.

SZALAY, A. S.; BLAKELEY, J. A. Gray's Laws: Database-centric Computing in Science. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The Fourth Paradigm**: data-intensive scientific discovery. Redmond, WA, USA: Microsoft Research, 2009, p. 5-11.

SZALAY, A.; GRAY, J. 2020 Computing: Science in an exponential world. **Nature**, v. 440, p. 413-414, 23 mar. 2006. Disponível em: <<http://www.nature.com/nature/journal/v440/n7083/full/440413a.html>>. Acesso em: 01 fev. 2011.

TAYLOR, J. **The UK E-Science Programme**, 27 jul. 2001. Apresentação feita no e-Science London Meeting.

THE LONDON E-SCIENCE CENTRE. **LESC - London e-Science Centre - what is e-Science?** Disponível em: <<http://www.lesc.ic.ac.uk/admin/escience.html>>. Acesso em: 26 jan. 2011.

UK DATA ARCHIVE. **UK Data Archive - About the Archive.** Disponível em: <<http://www.data-archive.ac.uk/about/archive>>. Acesso em: 28 jan. 2011.

VAN DEN EYNDEN, V.; CORTI, L.; WOOLLARD, M.; BISHOP, L. **Managing and Sharing Data: a best practice guide for researchers.** 2 ed. UK Data Archive, 2009. Disponível em: <<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>>. Acesso em: 28 jan. 2011.

WILBANKS, J. I Have seen the paradigm shift, and It Is Us. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery.** Redmond: Microsoft Research, 2009. p.209-214.

WING, J. M. **Computational Thinking.** 2005. Disponível em <<http://www.cs.cmu.edu/afs/cs/usr/wing/www/ct-paper.pdf>>. Acesso em 26 jan. 2010.

_____. _____. **Communications of the ACM**, v. 49, n. 3, p. 33-35. 2006.

WOOLEY, J. C. **Building a cyberinfrastructure for the Biological Sciences (CIBIO): 2005 and Beyond: a roadmap for consolidation and exponentiation.** A Workshop Report, 2003. Disponível em: <http://research.calit2.net/cibio/archived/CIBIO_FINAL.pdf>. Acesso em: 28 jan. 2011.



Informática Agropecuária

Ministério da
**Agricultura, Pecuária
e Abastecimento**



CGPE 9799