



**PESQUISA AGROPECUÁRIA NO CONTEXTO DA *E-SCIENCE*:  
MONITORAMENTO DE TEMAS E PLATAFORMAS DE *DATA SCIENCE***

***AGRICULTURAL RESEARCH IN THE CONTEXT OF E-SCIENCE:  
MONITORING DATA SCIENCE TOPICS AND PLATFORMS***

**Martha Delphino Bambini**

Doutoranda no DPCT/ IG/ Unicamp, Analista da Embrapa Informática Agropecuária  
E-mail: [martha.bambini@embrapa.br](mailto:martha.bambini@embrapa.br)

**Maria Beatriz Machado Bonacelli**

Professora Livre-docentado Depto de Política Científica e Tecnológica (DPCT/IG/Unicamp).  
E-mail: [bia@ige.unicamp.br](mailto:bia@ige.unicamp.br)

**Roberto Hiroshi Higa**

Pesquisador da Embrapa Informática Agropecuária  
E-mail: [roberto.higa@embrapa.br](mailto:roberto.higa@embrapa.br)

**Grupo de Pesquisa: GT13. Temas emergentes no agronegócio**

**Resumo**

A *e-Science* é um novo paradigma da exploração científica, que envolve a coleta, análise, armazenamento de grandes volumes de dados e requer uma infraestrutura computacional robusta e distribuída, para pesquisas colaborativas. A pesquisa agropecuária pode se beneficiar muito desta nova abordagem. Este artigo apresenta um mapeamento não-exaustivo de temas de pesquisa agropecuária intensiva em dados e um *benchmarking* de *facilities* nacionais e internacionais relacionadas ao assunto. Em um ambiente de grande velocidade de mudanças tecnológicas, estudos de monitoramento e de *benchmarking* são essenciais para acompanhar as tendências da fronteira do conhecimento e embasar políticas públicas com foco estratégico e de longo prazo. Entre os resultados, aponta-se a necessidade de que as organizações brasileiras de pesquisa agropecuária aparelhem-se em termos de recursos computacionais e competências, seja por aquisição ou estabelecendo parcerias com centros de pesquisa nacionais e internacionais. Ainda existem muitos desafios de cunho tecnológico, de aprendizado e institucionais relacionados à Ciência Intensiva em Dados mas, para que o Brasil possa manter sua liderança tecnológica em agricultura tropical, há que se abrir novas frentes de investigação para participar ativamente da revolução da *e-Science* agrícola.

**Palavras-chave:** Ciência de Dados; Agricultura; Pesquisa Agropecuária; Agropecuária; *e-Science*.

**Abstract**

*E-Science* is a new paradigm of scientific exploration, which involves collecting, analyzing and storing large volumes of data and requires a robust and distributed computing infrastructure for collaborative research. Agricultural research can greatly benefit from this new approach. This article presents a non-exhaustive mapping of data intensive agricultural research themes and a benchmark of national and international facilities related to the subject. In a fast-paced technological change environment, monitoring and benchmark studies are essential to keep up with trends in the frontier of knowledge and to support public policies with strategic and long-term focus. Among the results, it is pointed out the need for Brazilian agricultural research organizations to prepare themselves in terms of



*computational resources and skills, either by acquisition or by establishing partnerships with national and international research centers. There are still many technological, learning and institutional challenges related to Data Science, but in order for Brazil to maintain its technological leadership in tropical agriculture, new research fronts must be opened to actively participate in the e-Science revolution in agriculture.*

**Key words:** *Data Science; Agriculture; Agricultural Research; Agriculture; e-Science*

## **1. Introdução**

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE,2016) apresenta alguns desafios a serem enfrentados pelo Sistema Global Alimentar nas próximas décadas. Existe uma necessidade de grande elevação da produção de alimentos, destinada a uma população crescente, que tem perspectivas e aumento de renda e consumo. Além disso, o setor agroalimentar poderá contribuir para o desenvolvimento e o crescimento econômico, para a redução da pobreza e para a geração de empregos no campo. Outros desafios envolvem a competição com outros setores pelo uso dos recursos naturais (como terra e água), a preservação da biodiversidade, a restauração de ecossistemas frágeis e a mitigação dos efeitos das mudanças climáticas globais.

A aplicação de novas tecnologias às atividades produtivas e o oferecimento de serviços tecnológicos destinados ao setor agropecuário têm um grande potencial para incrementar rentabilidade, reduzir perdas e desperdícios e contribuir para o aumento da qualidade dos produtos e dos processos produtivos e logísticos, bem como para minimizar os prejuízos decorrentes de infestações de pragas, manifestações de doenças e eventos climáticos (GOYAL; GONZÁLES-VELOSA, 2013; THE ECONOMIST, 2016). O estudo sobre o futuro da agricultura apresentado pela revista Economist (2016) destaca que, para atingir um nível de crescimento adequado à expectativa de elevação da população mundial, o setor agropecuário deve adotar algumas práticas vigentes no setor industrial, como o monitoramento e o controle de variáveis que influenciam sua produtividade, de forma a otimizar sua performance.

Ramos (2007) ressalta a forte dependência que o setor agropecuário possui em relação às condições naturais seja do ponto de vista estrutural (como a disponibilidade de terra e fertilidade do solo, a oferta de água e o clima predominante da região) quanto do ponto de vista conjuntural, relacionado ao meio ambiente e às variações climáticas. Pode-se dizer, assim, que as boas práticas produtivas agropecuárias envolvem o gerenciamento de uma ampla matriz de variáveis que incluem: dados sobre tempo e clima, nutrientes e umidade do solo; informações sobre ervas daninhas, pragas e doenças; o planejamento de atividades e o levantamento dos custos relacionados às atividades de produção. Se este gerenciamento é efetivo, pode-se maximizar rendimentos e lucros e minimizar custos e riscos associados à atividade.

A Agricultura Digital, ou Agricultura 4.0, se refere ao emprego de métodos computacionais de alto desempenho, redes de sensores, comunicação máquina para máquina



(M2M), conectividade entre dispositivos móveis, computação em nuvem, métodos e soluções analíticas para processar grandes volumes de dados e construir sistemas de suporte à tomada de decisões de manejo, conforme Massruhá e Leite (2017). Estão incluídas no escopo da agricultura digital, a agricultura e a pecuária de precisão, a automação e a robótica agrícola, além de técnicas de *Big Data* e a Internet das Coisas. Este novo contexto tecnológico da agropecuária se insere no paradigma da quarta revolução industrial (SCHWAB, 2016), caracterizado por uma terceira fase da Revolução das Tecnologias de Informação e Comunicação (TIC), apresentada por Cowhey, Aronson e Abelson (2009).

Cowhey et al. (2009) apontam que esta fase é marcada por um ponto de inflexão, no qual as novas arquiteturas de TIC estruturam-se em um conjunto bastante distintos de capacidades e organização de mercado do que nas fases anteriores. Estas arquiteturas consideram três tendências que se inter-relacionam: (i) a disponibilidade de redes de Internet de banda larga, a custo acessível; (ii) a integração de um armazenamento massivo e barato de informação, associada a uma nova arquitetura de redes e serviços; e (iii) a emergência de sistemas de computação virtual que controlam e se utilizam de muitos computadores, incluindo supercomputadores, de forma a atender demandas dos usuários e aumentar eficiência e produtividade.

No contexto desta nova fase da revolução das TIC, o volume de geração de dados de pesquisas em áreas de interesse agropecuário vem crescendo rapidamente e as necessidades de armazenamento de processamento de grandes volumes de dados geraram o que se chamou de “quarto paradigma da exploração científica” ou Ciência Intensiva em Dados, *e-Science*, conforme o pesquisador Jim Gray (HEY, TANSLEY e TOLLE, 2009). Os autores destacam que, no século XXI, a grande maioria dos dados científicos é capturada por instrumentos, de forma contínua, ou então são gerados a partir de modelos computacionais. Posteriormente, estes dados passam por processamento por softwares específicos, com o armazenamento da informação e do conhecimento resultante. Os dados podem ser armazenados a longo prazo, para ficarem acessíveis de forma pública, curados e gerenciados de forma a garantir sua análise continuada e a geração de novas teorias científicas derivadas.

Appel et al. (2016) destacam que este novo contexto de pesquisa científica requer uma infraestrutura computacional robusta e escalável, que possa ser organizada ou acessada de forma distribuída, de forma a que pesquisadores possam colaborar remotamente de qualquer parte do planeta. Os autores ponderam que, entre outros fatores que impulsionam a *e-Science*, destaca-se uma vertente científica com enfoque computacional estabelecida em algumas áreas do conhecimento, especialmente aquelas relacionadas à simulação de fenômenos complexos.

Vale dizer que a pesquisa intensiva em dados já vem sendo desenvolvida há várias décadas. No entanto, a partir de meados dos anos 2000, plataformas computacionais e ferramentas de software mais adequadas se tornaram mais acessíveis às equipes de pesquisa. Algumas áreas do conhecimento, como a pesquisa agropecuária, já vem estudando



fenômenos complexos que envolvem o estudo de um grande volume de dados envolvendo vários tipos de variáveis a partir do paradigma da Ciência Intensiva em Dados.

No Brasil, ações de pesquisa agropecuária, financiadas principalmente com recursos públicos, vêm sendo desenvolvidas desde a década de 1980 no sentido de monitorar e analisar computacionalmente as variáveis que influenciam a produtividade agropecuária, seja no âmbito da propriedade quanto para direcionar políticas públicas, de forma a promover transformações no campo. Lopes (2014) menciona iniciativas de ciência de dados desenvolvidas no Brasil pelo Instituto Nacional de Pesquisas Espaciais (INPE) e pelo Instituto Nacional de Meteorologia (INMET), correlacionando dados gerados a partir de imagens de satélites e medições de estações meteorológicas para melhorar a precisão da previsão do tempo e políticas públicas. A integração de dados gerados por satélites com dados cartográficos, aliado à utilização de técnicas de geoprocessamento, permitiu o desenvolvimento de ações de monitoramento de queimadas, desmatamentos e uso do solo a partir da estrutura de pesquisa pública no país.

Souza et al. (2017) destacam uma política do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) implementada em 1996, a partir de esforços de pesquisa desenvolvida em parceria com uma rede de organizações coordenada pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), chamada Zoneamento Agrícola de Risco Climático (ZARC). O ZARC foi desenvolvido a partir de um estudo do comportamento do tempo e clima para cada cultura, nos municípios brasileiros, considerando diferentes tipos de cultivares e solos (MAPA, 2018). A pesquisa envolve um trabalho de modelagem de um histórico de dados de temperatura, chuva e vento, solo para todo o território nacional com os comportamentos fenológicos de cada cultura, de forma a identificar a melhor época de plantio das culturas em cada localidade, criando um instrumento para gestão de riscos climáticos. O trabalho de Romani et al. (2016) apresenta esta iniciativa em detalhes, destacando o desenvolvimento de um sistema de informações agrometeorológicas (Agritempo) a partir de uma robusta base de dados estabelecida a partir de uma rede de parceiros formada para compartilhamento de dados meteorológicos.

Mesmo que algumas iniciativas de ciência baseada em dados já venham sendo realizadas por institutos de pesquisa brasileira, tendo em vista o intenso desenvolvimento e evolução das ferramentas computacionais disponíveis neste ponto de inflexão das TIC (COWHEY et al., 2009), há que se estabelecer um novo patamar de pesquisa agropecuária no país, direcionado por: disponibilidade e acesso a *facilities* computacionais de alto processamento e grande capacidade de armazenamento de dados; criação de novos modelos de coleta de dados (envolvendo sensores e simuladores); emprego de novas metodologias de análise para grandes volumes de dados; fortalecimento de competências para lidar com novas ferramentas computacionais; formação de equipes interdisciplinares para explorar problemas utilizando *data science*; habilidades de comunicação entre cientistas de computação e especialistas de domínio agro; promoção de pesquisas colaborativas com compartilhamento de equipamentos, dados e conhecimentos.



O paradigma da *e-Science* envolve a construção de um grupo de recursos físicos, computacionais e de competências compartilhados entre equipes localizadas em diversos pontos do globo. Estas iniciativas se tornaram possíveis com a redução de custos e os avanços do poder de processamento e armazenamento de dados, bem como pelas possibilidades de conectividade garantidas pela banda larga. Seja através de estruturas e competências próprias ou através de parcerias e compartilhamento de infraestruturas e recursos humanos, há que se buscar caminhos para inserir a pesquisa agropecuária brasileira no contexto de uma Ciência Intensiva em Dados.

Para manter seu papel preponderante em fortalecer a agricultura tropical e desenvolver-se no sentido de enfrentar seus desafios futuros, as organizações brasileiras de pesquisa agropecuária necessitam adequar-se e aparelhar-se ao contexto da *e-Science*.

Este trabalho buscou identificar oportunidades para as organizações brasileiras de pesquisa agropecuária no que se refere ao paradigma da *e-Science* ou Ciência Intensiva em Dados. Em um ambiente de incerteza e alta velocidade de mudança tecnológica, é relevante desenvolver ações periódicas, ou mesmo contínuas, de monitoramento tecnológico para apoiar o planejamento de ações de Ciência, Tecnológica e Inovação, incluindo priorização de temas e linhas de pesquisa e investimentos em equipamentos e plataformas de pesquisa. O artigo apresenta um estudo de monitoramento que identificou as principais áreas e temas estratégicos de pesquisa intensiva em dados, complementado por um estudo não-exaustivo de *benchmark* de plataformas de e-Science já estabelecidas, em âmbito internacional.

A próxima seção elenca aspectos conceituais relacionados à Ciência de Dados e ao paradigma da *e-Science*. Na sequência, são apresentadas as seções de Metodologia e Resultados e Discussão. Por fim, estão descritas as conclusões do estudo.

## **2. Aspectos conceituais da Ciência de Dados e o quarto paradigma da ciência (e-Science)**

A ciência intensiva em dados, ou *Data Science*, em inglês, se utiliza de processos científicos e computacionais para extrair conhecimento, padrões, tendências e *insights* a partir de conjuntos de dados de vários formatos, estruturados ou não-estruturados (BELL et al., 2009). Provost e Fawcett (2013) destacam que o objetivo final da Ciência de Dados é aprimorar a tomada de decisão a fim de entender fenômenos por intermédio da análise automatizada de dados. Esta Ciência se formou a partir da convergência de muitos campos de estudo tradicionais como: análise causal; estatística; métodos de visualização de dados, etc.

Gray (2009) utilizou o termo quarto paradigma da ciência (empírico, teórico, computacional e direcionado para os dados) para se referir a um novo contexto da pesquisa científica. Este novo paradigma é colaborativo, em rede e fortemente direcionado a dados. Chamado de *e-Science*, envolve a aplicação intensiva de Tecnologia de Informação e Comunicação (TIC) aos processos científicos, vem provocando mudanças e transformações nunca imaginadas, de acordo com o autor. Hey et al. (2009) destacam que, tendo em vista que a ciência intensiva em dados envolve técnicas e tecnologias bastante específicas e



distintas, pode ser considerada como um novo paradigma da exploração científica (diferente da ciência computacional pura e simples).

Hey et al. (2009) descrevem um novo modelo científico marcado por uma crescente geração de dados - seja a partir de instrumentos de captura (sejam sensores, satélites, câmeras, etc) ou então por intermédio de simulações computacionais - associada a etapas de tratamento, processamento e armazenamento de dados, seja em computadores ou via *cloud computing* (nuvem). Para analisar estes grandes volumes de dados, novas ferramentas tecnológicas e novas competências se fazem necessárias. Existe a demanda por um novo perfil profissional – um cientista de dados – que deve ser treinado nas habilidades necessárias para fazer descobertas a partir de um grande conjunto de dados (DAVENPORT e PATIL, 2012).

Provost e Fawcett (2013) abordam também o contexto tecnológico marcado por grandes quantidades de dados disponíveis, tornando impossível sua análise de forma manual e algumas vezes excedendo a capacidade de bases de dados computacionais mais usuais. Ao mesmo tempo, computadores vem se tornando mais poderosos, a conexão em rede vai se tornando ubíqua, e os algoritmos vem sendo aprimorados de forma a possibilitar a conexão de bases de dados e análises mais profundas do que se poderia imaginar. Os autores indicam que a convergência destes movimentos levou a uma ampliação da aplicação da chamada *Data Science* a vários setores econômicos.

Bell (2009) descreve três atividades básicas da computação intensiva em dados: captura, curadoria e análise de dados. Os dados são coletados em escalas e formatos variados, podendo envolver resultados de experimentos e de observações individuais, laboratórios únicos, grupos de laboratórios parceiros, redes de laboratórios internacionais em larga escala bem como conjuntos de dados pessoais de indivíduos e pesquisadores. O autor destaca que, mesmo que um enorme potencial tenha sido identificado, a ciência intensiva em dados tem se desenvolvido ainda de forma lenta, talvez pela falta de entendimento da comunidade científica em relação a este tema e também pelo desconhecimento de ferramentas práticas para criar, gerenciar e utilizar estas bases de dados.

Ramaswamy (2016) considera que o *Data Science* pode levar a descobertas transformadoras, com a integração entre campos científicos distintos. No caso da agropecuária, o autor aponta a importância de uma abordagem sistêmica envolvendo o setor produtivo, considerando a adoção de vários tipos de TIC como: o uso de sensores para a captura de dados e informações; técnicas de integração de dados; robôs e drones, entre outros. O autor aponta que a pesquisa agropecuária deve utilizar os dados gerados por este novo modelo, chamado *Smart Farm*, para entender fatores que afetam a produtividade e a geração de perdas no âmbito das cadeias produtivas. Algumas aplicações no campo de estudo da agropecuária, elencadas pelo autor, incluem: genômica, fenômica e agricultura/pecuária de precisão, a serem discutidos na seção de resultados.



Em relação à comunidade científica agropecuária, Reardon et al. (2018) destacam que, além de focar o segmento produtivo (dentro da porteira), os pesquisadores necessitam considerar, em suas estratégias de pesquisa, o sistema alimentar como um todo bem como as transformações pelas quais vem passando. Indústrias agroalimentares, mercado urbano, cadeias produtivas devem ser analisados, pois influencia m processos de adoção tecnológica e lucratividade a serem obtidas a partir de novas tecnologias.

A próxima seção apresenta a metodologia de pesquisa conduzida.

### **3. Metodologia de Pesquisa**

O objetivo deste trabalho foi identificar oportunidades para as organizações brasileiras de pesquisa agropecuária quanto ao paradigma da *e-Science* ou Ciência Intensiva em Dados. Tendo em vista a elevada velocidade das mudanças relacionadas às TIC e da geração de dados de pesquisa empírica e de simulações, foi desenvolvida uma ação de monitoramento, conforme descreve Porter et al. (1991), de temas de pesquisa agropecuária que vem sendo tratados sob o paradigma da pesquisa intensiva em dados. As bases de dados consultadas foram: (i) a base da *Web of Science* (para busca de documentos científicos) (ii) o *Google Scholar* para complemento das buscas. As palavras-chave utilizadas foram “*Data Science*” e “*Agriculture*”. Considerando a grande ocorrência do termo “*Big Data*” associado aos documentos encontrados, este também foi incluído na pesquisa. A busca foi conduzida em duas etapas: no segundo semestre de 2016 e posteriormente no primeiro semestre de 2018. Foram encontrados poucos documentos, considerando a novidade do tema; em 2016, foram selecionados para revisão estruturada apenas 6 artigos científicos e, em 2018, 8 documentos. Vale notar que alguns trabalhos encontrados foram elaborados pelo mesmo grupo de pesquisa.

Adicionalmente, foi conduzida uma busca não-exaustiva para identificar plataformas computacionais de *e-Science* aplicada a agropecuária atualmente em operação. As fontes de informação pesquisadas foram documentos, estudos, relatórios e artigos em *websites* de organizações relacionadas ao domínio agropecuário, orientadas para pesquisa como: a *Food and Agriculture Organization of the United Nations* (FAO) e o *United States Department of Agriculture* (USDA), entre outros.

### **4. Resultados e Discussão**

Pospiech e Felden (2012) enumeram a vasta distribuição de disciplinas às quais vêm sendo aplicadas tecnologias de *Data Science*, como: ciências sociais, *e-government*, finanças, medicina, ciências comportamentais, análise forense, construção civil, bioinformática, astrologia, climatologia. Adicionalmente, Diebold (2012) aponta que o *Big Data* vem se afigurando como uma iniciativa importante de colaboração interdisciplinar, com grandes chances de sucesso.



Ramaswamy (2016) destaca que o tema perpassa toda a cadeia produtiva agropecuária, desde os insumos, a produção na fazenda, a logística e o transporte, a refrigeração, os dados de venda e consumo e a questão de descarte e resíduos.

Em relação a problemas e demandas da pesquisa agropecuária foram identificados alguns grupos de aplicações, a partir da pesquisa documental efetuada no website do *United States Department of Agriculture (USDA)* <sup>1</sup>:

- Modelagem e simulação usando dados meteorológicos e ambientais;
- Trinômio genótipo + ambiente + GXE (*genotype–environment interaction*);
- Análise de Dados Genômicos;
- Metagenômica;
- Relação entre Alimento, Água e Energia (Nexus Food, Energy, Water);
- Agricultura de Precisão e Internet das Coisas: detectores e sensores, geotecnologias (sensoriamento remoto e sistemas de informação geográfica), engenharia.

A descrição e um detalhamento dos problemas de pesquisa e casos de aplicação associados a estes temas são apresentados a seguir.

#### **4.1 Meteorologia e Estudos do Clima**

O trabalho do *Agricultural Research Service (ARS,2013)*, vinculado ao USDA, enfatiza pesquisas conduzidas no sentido de aumentar a resiliência dos sistemas produtivos por intermédio da investigação e compreensão de respostas regionais e locais a variáveis ambientais estressoras considerando as interações entre culturas, animais, solo, água, tempo e clima. Pode-se obter um melhor entendimento destas interações com o aumento da integração de grandes quantidades de dados coletadas em diferentes locais. Geralmente estudos de impactos de condições ambientais levam a resultados conflitantes em virtude das complexas interações entre as variáveis, associadas a diferentes práticas de manejo e gestão.

Uma abordagem de pesquisa intensiva em dados possibilita que os resultados sejam analisados considerando-se diferentes anos e locais, a fim de entender e quantificar as interações entre as variáveis e os resultados obtidos. Com isso, é possível: aprimorar a identificação das melhores práticas produtivas para cada local, considerando as condições do tempo no longo prazo; inovar e explorar novas práticas de gestão; extrapolar resultados para outros locais; integrar os resultados relacionados a uma região; avaliar possíveis efeitos de mudanças climáticas futuras; desenvolver modelos teóricos e ferramentas preditivas. ARS (2013) destaca que análises comparativas entre resultados de diferentes locais poderão reduzir

---

1 Estes grupos foram selecionados a partir do conteúdo abordado no workshop “*Big Data and Computing*” promovido pelo ARS/USDA (ARS, 2013), e no evento “*Data Science in Agriculture Summit - Changing the Face, Place, and Space of Agriculture*” promovido em 2016 pelo *National Institute of Food and Agriculture (NIFA)* também vinculado ao USDA (RAMASWAMY, 2016).



a redundância de pesquisas que são conduzidas em muitos locais distintos, permitindo uma alocação de recursos mais eficiente.

Gustafson et al. (2015) descrevem as pesquisas de campo conduzidas no Meio Oeste americano para adaptação a mudanças climáticas globais. Boote (2016) apresenta o projeto *AgMIP - Agricultural Model Intercomparison and Improvement Project* (AgMIP, 2017), a maior iniciativa colaborativa para melhorar o estado das simulações agrícolas e entender melhor os impactos do clima sobre o setor agrícola em escalas global e regional.

Outra iniciativa envolvendo o emprego de *Data Science* à Meteorologia Agrícola é o projeto *Global Open Data for Agriculture and Nutrition* (GODAN) que possui mais de 457 parceiros dedicados à defesa do compartilhamento de dados abertos para a pesquisa como meio para a geração de soluções inovativas para aumentar a segurança alimentar na África (GODAN, 2018). Os parceiros pertencem a diferentes categorias: Governos, organizações não-governamentais e empresas privadas. Esta iniciativa foi lançada em 2012, em uma reunião do grupo de líderes mundiais G8, com financiamento parcial do Governo dos Estados Unidos. O objetivo da ação é incentivar o crescimento dos dados gerados por novas tecnologias a fim de resolver problemas antigos do setor e beneficiar os produtores agropecuários e a saúde dos consumidores.

No Brasil, temos as experiências de pesquisa intensiva em dados em agrometeorologia descritas por Romani et al. (2016) e Souza et al. (2017), anteriormente mencionadas

#### **4.2 Interação genótipo + ambiente + GXE**

A palestra de Pat Schnable (2016) no evento do *National Institute of Food and Agriculture* (NIFA/USDA) enfocou a interação genótipo + ambiente (GXE - *Genotype-Environment Interaction*) e o desenvolvimento de modelos preditivos para aumentar o conhecimento biológico, envolvendo o fornecimento de recomendações e previsão de safra. O trabalho deste pesquisador busca uma maior compreensão entre as características manifestadas pelas plantas, em ambientes diversos, e seus genomas, a partir de grandes volumes de dados. A utilização de robótica neste tipo de pesquisa incrementou fortemente as possibilidades para coleta de dados experimentais de forma conjunta às abordagens automatizadas de fenotipagem, associadas à redução do custo de sequenciamento e genotipagem, e ao grande aumento na velocidade de execução destas atividades. Este novo contexto vem gerando dados em grande volume e velocidade. Aliando dados, tecnologias para fenotipagem como sensores, robôs e VANTs (Veículo Aéreo Não-Tripulado), e novos algoritmos para extrair características a partir de imagens, é possível: estimar rendimentos, fenótipos e safra; e aumentar o conhecimento sobre genética, ambiente e GXE no fenótipo. É um trabalho multidisciplinar reunindo conhecimentos em ciências de plantas, engenharia e computação.



### 4.3 Análise de Dados Genômicos

Os laboratórios de estudos genômicos do ARS (2013) enfrentam vários desafios computacionais. A principal dificuldade se refere à estimativa de muitos parâmetros com relativamente poucas observações. Exemplificando: tem-se milhões de dados genéticos para seres humanos e poucos milhares de medidas de hábitos alimentares e estilo de vida e de fenótipos relacionados a manifestação de doenças. Neste caso, modelos estatísticos complexos são requeridos para tentar prever os efeitos dos marcadores para milhares de fenótipos considerando os milhares de observações.

Um exemplo citado é o ARS *Nutritional Genomics Laboratory* localizado na Universidade de Tufts, que estuda nutrição humana e sua relação com doenças comuns. Um de seus objetivos é caracterizar como os hábitos alimentares e o estilo de vida modulam a relação entre os fatores genéticos e as doenças, tentando eventualmente prever como a nutrição afeta os riscos de adoecimento. Neste caso, a análise de diferenças genéticas necessita ser associada com hábitos alimentares, dados de estilo de vida, e histórico de doenças. O aprimoramento da estrutura computacional permitirá análises mais rápidas e aumentará o impacto dos resultados gerados por este tipo de pesquisa.

### 4.4 Metagenômica

A saúde e o bem-estar dos humanos, das plantas e dos animais pode ser promovido pelo conhecimento das ricas interações entre comunidades de micróbios por intermédio de *insights* obtidos pela análise dos genomas dos membros destas comunidades (ARS, 2013). O microbioma – comunidades de microrganismos existentes no trato intestinal de animais e humanos - contribui de forma importante para a nutrição, digestão, imunidade, fisiologia, resistência a doenças, e possivelmente até no comportamento do animal hospedeiro.

Antes do sequenciamento de DNA em *high throughput*, as ferramentas para entender estas relações eram rudimentares. A metagenômica e filogenética modernas transformaram estas análises. O entendimento do microbioma e suas interações, permite melhorar a saúde humana e animal pela formulação de melhores dietas, capazes de conferir um aumento da imunidade. Os conhecimentos derivados da metagenômica começam a ser explorados também para solo e recursos hídricos.

### 4.5 Relação entre Alimento, Água e Energia (*Nexus Food, Energy, Water*)

A segurança hídrica, segurança energética e segurança alimentar estão interligadas: a produção de alimentos demanda água; a extração de água, seu tratamento e distribuição demandam energia e a produção de energia requer água. A energia é utilizada para produzir fertilizantes, para a colheita, o transporte, para a irrigação e o processamento influencia o preço dos alimentos. As pressões ambientais e as mudanças climáticas assim como o



crescimento econômico e populacional intensificam as relações entre os três sistemas (MOHTAR e DAHER,2012). Este é um tema cuja grande quantidade de dados e complexidade associada, muito se beneficiaria de uma abordagem em *Data Science*.

#### **4.6 *Smart Farming*, Agricultura de Precisão e *Big Ag Data***

O conceito de *Smart Farming* (Fazenda Inteligente) vem sendo empregado para se referir à implementação de vários grupos de tecnologias para substituir técnicas produtivas consideradas morosas e pouco confiáveis. Envolve o monitoramento automatizado das variáveis de influência produtiva de forma precisa e barata e o suporte à decisão do produtor pelo processamento e análise destas variáveis. Diferentemente dos ambientes controlados da indústria, este paradigma está associado à existência de algumas dificuldades como: a implementação de sensores ao ar livre e em grandes distâncias; o acesso a redes de telefonia móvel e internet e à própria resistência do produtor em adotar estas novas tecnologias.

Um conceito muito próximo é o de agricultura/ pecuária de precisão, definida como um conjunto de tecnologias geolocalizadas gerando informação para apoiar a gestão da produção agrícola (COBLE et al., 2016).

Dentre as aplicações de *Data Science* ao contexto da *Smart Farming*, Souza et al. (2017) citam: análise de dados de redes de sensores de forma a identificar padrões e possibilidades de intervenção, inclusive com agentes cognitivos; aplicações em lógica de games para treinamento; integração de sensores, simuladores e atuadores para reproduzir comportamentos de sistemas reais; realidade aumentada; modelos de alertas para a ocorrência de pragas e doenças a partir de dados de clima; sistemas de recomendações de conteúdo e sistemas de rastreabilidade baseados em *Blockchain*.

Keith Coble (2016) destaca a questão do *Big Ag Data*, um grande volume de dados gerado com grande velocidade na interseção das Geotecnologias, da informação sobre a produção no campo, sobre tempo e clima e sobre o mercado. Segundo o autor, a inclusão de informações de diferentes estratégias de gestão, junto a outras variáveis de interesse, aumenta a profundidade da análise para definir estratégias de gestão. Atualmente, a evidência empírica sobre a agregação de valor dos dados coletados pelos produtores ainda é limitada. O conceito de *Big Ag Data* inclui os dados espaciais, específicos ao local e associados com agricultura de precisão (como as características do solo no local e o rendimento da fazenda); e os metadados sobre práticas de gestão e tecnologias (como profundidade de plantio, colocação da semente, cultivar, condição de uso de máquinas, datas de plantio, aplicação de fertilizantes). Junto a eles existem os dados de eventos climáticos e evapotranspiração, entre outros, que não são controláveis pelo produtor.

Boas práticas de gestão e uso de informações do ambiente podem afetar as estimativas de insumos, o preço dos produtos, e o comércio de produtos, insumos e da terra. O valor



agregado à aplicação de *Big Ag Data* depende da existência de um número representativo de produtores contribuindo com suas informações para conjuntos de dados agregados, a fim de aumentar sua robustez e poder preditivo, em um “efeito de rede”.

Coble et al. (2016) destacam que nos países em desenvolvimento, pequenos produtores vêm, cada vez mais, usando informações derivadas de análises de *Data Science* (acessadas por *smartphones*) para reduzir riscos e aumentar a eficiência só sistema produtivo. Este pode ser um fator de influência sobre o comportamento de mercados globais. As limitações de acesso à infraestrutura de redes no meio rural, de certa forma, acabam por dificultar esta inclusão de forma mais ampla. Entende-se que esta questão estrutural será bastante relevante para que se possa construir bases de informações relacionadas aos microdados das propriedades.

#### **4.7 Benchmarking: plataformas existentes de pesquisa agrícola intensiva em dados**

A pesquisa documental, não exaustiva, identificou algumas plataformas de instituições de pesquisa internacionais, as quais são analisadas a seguir.

Uma organização bastante relevante é a **Research Data Alliance (RDA)**, criada em 2013, pela *European Commission*, o braço executivo da União Européia; pelo *National Science Foundation* e *National Institute of Standards and Technology*, dos Estados Unidos, e pelo Departamento de inovação do Governo da Austrália. O direcionamento da RDA é conduzido pela comunidade de pesquisa, com o objetivo de construir estruturas sociais e técnicas para permitir o compartilhamento de dados abertos, entre fronteiras, considerando alguns temas. Em relação aos dados da agropecuária, estes vêm se tornando cada vez mais importantes para o desenvolvimento de estratégias nacionais e globais de desenvolvimento e sustentabilidade. É um campo caracterizado por heterogeneidade e complexidade de um grande volume de dados, que necessita de sistemas integrados de ciência de dados. O grupo de interesse de Agricultura da RDA (IGAD, 2018) foi criado em 2013 com a missão de promover boas práticas de pesquisa, criar políticas de compartilhamento, gestão e interoperabilidade de dados, provendo um fórum para discussão, compartilhamento de experiências e divulgação de pesquisas.

O *Agricultural Research Service (ARS)* / *USDA* estabeleceu recentemente a SCINet formada por recursos computacionais de alta performance, plataforma chamada **Ceres**, associada a uma rede de internet de alta velocidade. Seu escopo envolve atividades de geração e/ou processamento e armazenamento de grandes conjuntos de dados. Paralelamente, está sendo desenvolvido o *Ag Data Commons*, um catálogo e repositório de dados para disponibilização pública e curadoria de dados (PARR, 2017). Woodard (2016) elenca outras plataformas computacionais no campo da pesquisa agropecuária intensiva em dados nos Estados Unidos.



No âmbito do *Consultative Group on International Agricultural Research* (CGIAR), a maior rede de organizações de pesquisa agrícola do mundo, foram identificadas três plataformas de pesquisa agropecuária intensiva em dados. A primeira delas é o **GIAR Platform for Big Data in Agriculture** (<http://bigdata.cgiar.org/>) uma plataforma, lançada em 2017, que busca gerar resultados disruptivos e de impacto na pesquisa agropecuária pela utilização de abordagens de *Big Data* para resolver problemas de pesquisa de forma mais rápida, melhor e em maior escala (CGIAR, 2018). A plataforma, co-liderada pelo *International Center for Tropical Agriculture* (CIAT) e pelo *International Food Policy Research Institute* (IFPRI), é focada em dados abertos e em seu compartilhamento, de forma a tornar-se um *broker* de dados confiáveis em nível global. Várias parcerias foram estabelecidas no âmbito desta iniciativa visando compartilhar dados, métodos, conhecimentos e expertise a fim de ampliar as capacidades e horizontes da pesquisa do CGIAR. As parcerias envolvem pesquisadores especializados em problemas de países em desenvolvimento (situados na África, Ásia, América Latina e Caribe), incluindo os 15 centros de pesquisa do CGIAR e 70 instituições externas (desde universidades e institutos de pesquisa públicos e privados e empresas).

Outra iniciativa é a **CGIAR Genebank Platform** (<https://www.genebanks.org>) cujo escopo é a conservar e disponibilizar um grande volume de material genético de culturas e plantas no âmbito do *International Treaty on Plant Genetic Resources for Food and Agriculture* (FAO, 2018). A plataforma apoia as operações e atividades centrais do banco de genes a fim de melhorar sua eficiência, aumentar seu uso e garantir conformidade com as diretrizes da política internacional. Sua finalidade é salvaguardar uma base de germoplasma que representa uma importante biodiversidade e atender aos pedidos de usuários de mais de 100 países. O emprego de novas tecnologias, novos conhecimento e processos possibilita uma atuação mais eficiente e efetiva para explorar novas oportunidades, conservar mais materiais e atender a um número maior de demandas a custos mais controlados. São 11 bancos de genes do CGIAR em vários locais e 591,000 acessos entre 2012 e 2016, existindo uma expectativa de aumento de demanda por germoplasma nos próximos anos em função de novas tecnologias de indexação de doenças, sequenciamento em *high throughput* e fenotipagem e *screening* de dados.

O CGIAR possui ainda a **Excellence in Breeding Platform** (<http://excellenceinbreeding.org/>), ainda em fase de implementação, que tem a perspectiva de modernizar os programas de melhoramento genético destinados aos países em desenvolvimento, visando aumentar seu impacto na segurança alimentar e nutrição e nas iniciativas de adaptação e às mudanças climáticas globais (CGIAR, 2018). A partir de inovações geradas nos setores público e privado, a plataforma proverá acesso a: ferramentas, serviços e práticas da fronteira do conhecimento, bem como capacitação orientada à aplicação e assessoria técnica. Seu intuito é estruturar novas estratégias de programas de melhoramento e investimentos para apoiar melhoristas na redução da pobreza pelo desenvolvimento de novas plantas para os países em desenvolvimento. A plataforma oferecerá um módulo para



genotipagem e ferramentas e serviços de sequenciamento, a custos mais acessíveis e de forma mais rápida. Outro módulo se refere a ferramentas de bioinformática, biométrica e gestão de dados, visando promover a integração de dados genéticos e de fenótipo e sua conversão em informação valiosa a partir de ferramentas e softwares de bioinformática.

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) possui também um **Laboratório Multiusuário de Bioinformática** (LMB), iniciativa de compartilhamento de recursos e prestação de serviços desenvolvida para otimizar o emprego de plataforma computacional e de competências humanas, nas áreas de expressão gênica, marcadores moleculares, evolução, regulação da expressão, modelagem de sistemas biológicos, predição de estrutura proteica e interação molecular, entre outras. A missão do LMB é viabilizar soluções de Bioinformática para projetos de pesquisa, desenvolvimento e inovação na Embrapa e seus parceiros, por intermédio de um ambiente colaborativo. A atuação do LMB é apresentada por Bambini et al. (2015).

O *Institut National de la Recherche Agronomique* (INRA, 2018) possui um laboratório convergente chamado **DigitAg** (<http://www.hdigitag.fr/en/>), *Institute for the Convergence of Digital Agriculture*, localizado em Montpellier que busca prover ferramentas científicas para estimular o desenvolvimento da agricultura digital a partir de uma abordagem interdisciplinar. O instituto busca ser uma referência em agricultura digital no campo acadêmico e empresarial. Seus focos são: a agricultura digital centrada no humano, o entendimento da adoção de TIC e sua interação com a agropecuária e a gestão da fazenda incluindo o desenvolvimento de equipamentos, modelos e ferramentas para uso agropecuário em qualquer escala. Um aspecto original é a interação que provê entre comunidades e objetivos científicos e tecnológicos. Seus eixos de atuação envolvem: TIC e sociedades rurais, agricultura digital e inovação, incluindo propriedade intelectual; sensores e aquisição de dados; sistemas de informação, armazenamento e transferência de dados, incluindo *Big Data* e interoperabilidade; mineração de dados, *analysis* e descoberta de conhecimento; modelagem de sistemas produtivos em *smart agriculture*.

Dentre as universidades, temos o **Data Competence Center** (WDCC) da Universidade de Wageningen (<https://www.wur.nl/en/Expertise-Services/WDCC.htm>) criado para apoiar as atividades de *big data* e *data science* na *Wageningen University* (2018), universidade com grande direcionamento para estudos agropecuários. A atuação do WDCC envolve educação, pesquisa, criação de valor, infraestrutura computacional e gestão de dados, e permite o desenvolvimento de novas aplicações e pesquisas intensivas em dados, para aumentar a qualidade de vida. As atividades do WDCC reúnem pesquisadores da área computacional e de outros campos científicos e promovem trocas de experiências, capacidades e competências.

A Universidade de Cornell criou a plataforma **Ag-Analytics** (<https://ag-analytics.org/>) em parceria com a empresa John Deere, com a missão de empoderar os produtores no uso de dados de agricultura de precisão para tomar decisões mais rápidas e melhores. A plataforma agrega dados climáticos, de solo, de rendimento e imagens de satélite, que podem ser



consultadas pelos produtores. A parceria com o centro de operações da John Deere possibilita que os produtores integrem seus dados de forma segura e precisa para extrair inteligência a partir das ferramentas de gestão de risco da plataforma. Uma descrição das bases de dados integradas nesta plataforma pode ser encontrada em Woodard (2016).

No Brasil, destacam-se algumas iniciativas de algumas universidades. Na Universidade Estadual de Campinas (UNICAMP, 2018), o **Laboratório Central de Tecnologias de Alto Desempenho (LaCTAD)** (<http://www.lactad.unicamp.br/br/sobre>) é uma estrutura de laboratório multiusuário, desenvolvida com apoio da Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP), que atua nas áreas de Biologia Celular, Genômica, Proteômica e Bioinformática oferecendo serviços de alta qualidade para toda a comunidade científica e empresas privadas. A implantação desse Laboratório tem como base a aquisição de equipamentos científicos, utilizados de forma continuada por uma equipe especializada, o que possibilita um melhor aproveitamento das funções de cada equipamento. O LaCTAD oferece assim, recursos de alta qualidade e boa relação custo-benefício à comunidade da Unicamp e outras Universidades e Institutos de Pesquisa e de Tecnologias Avançadas.

Na área da Saúde Pública, a Universidade de São Paulo (USP, 2018) implementou, em 2016, o **Laboratório de Big Data e Análise Preditiva em Saúde (Labdaps)** com o objetivo de identificar padrões no sistema de saúde por meio da análise de grandes bancos de informações. Outra iniciativa é o **Laboratório de Análise de Massiva de Dados (ANALYTICS)** que atua em pesquisas em ciência de dados (*Data Science*), em particular para dados que podem apresentar volume massivo, vir de diferentes fontes, apresentar diferentes formatos e ser gerados em fluxos. O laboratório trabalha com pré-processamento de dados, análise descritiva de dados e análise preditiva de dados e possui uma extensa rede de colaborações científicas nacionais e internacionais. Vale notar que a Universidade possui, desde 2013, uma rede de computação acadêmica de alta velocidade, iniciativa importante para enfrentar os desafios trazidos pelos novos paradigmas computacionais e vai ao encontro das crescentes demandas por largura de banda, tais como a computação em nuvem e o Big Data.

Outra iniciativa importante para a ciência intensiva em dados, empreendida pela FAPESP, o **Programa Equipamentos Multiusuários (EMU)** integrada ao Programa de Apoio à Infraestrutura de Pesquisa do Estado de São Paulo, mantido desde 1995, As chamadas referentes aos EMUs, iniciadas em 2010, envolviam recursos para a compra de equipamentos de última geração e para a sua disponibilização racional ao maior número de pesquisadores de São Paulo, do Brasil, da América Latina e de outros países, mediante critérios definidos para compartilhamento. Iniciativas de laboratórios multiusuários em andamento nas universidades do estado de São Paulo estão apresentadas em FAPESP (2018).



## 5. Conclusões

Este trabalho identificou oportunidades de temáticas de pesquisa que tem potencial para serem abordadas e analisadas a partir do paradigma da *e-Science*. Os temas e problemas de pesquisa tendem a crescer com a evolução tecnológica e o aumento das competências das equipes de pesquisa. A pesquisa documental evidenciou que a Embrapa já tem iniciativas de pesquisa intensiva em dados no campo da agrometeorologia e uma plataforma compartilhada em bioinformática. Outros campos podem ser explorados a partir desta perspectiva.

Entende-se que a ciência de dados pode levar a descobertas transformadoras, conforme indica Ramaswamy (2016), tendo em vista a possibilidade de integração entre campos científicos distintos, a partir de uma abordagem sistêmica considerando dados e informação gerados pela etapa produtiva, mas também a pré-produção e a pós-produção. Conforme destacava o pesquisador Jim Gray, a ciência está se modificando totalmente em função da revolução das TIC (HEY et al., 2009). Pesquisas experimentais, teóricas e computacionais estão se alterando em função do chamado “dilúvio de dados”.

Um grande valor da pesquisa intensiva em dados é a possibilidade de estabelecer relações entre variáveis, o que antes não era realizado especialmente por limitações técnicas, relacionadas ao processo de coleta ou a dificuldades de processamento ou armazenamento de dados. Este novo paradigma permite a análise dos fenômenos complexos, a partir do estabelecimento destas relações. Para enfrentar seus desafios futuros da agricultura tropical, considerando a evolução das tecnologias de informação, as organizações brasileiras de pesquisa agropecuária necessitam aparelhar-se com recursos computacionais e competências, seja pela aquisição de equipamentos próprios através de ações de fomento, como o programa EMU/ FAPESP, ou então através de parcerias com universidades ou centros de pesquisa nacionais e internacionais, públicos ou privados. Em um contexto de *e-Science*, compartilhamento de *facilities* e intercâmbios de dados serão frequentes. Neste contexto, o avanço científico agropecuário dependerá, cada vez mais, de uma visão sistêmica aliada a um esforço multidisciplinar de pesquisadores em vários campos (economia, engenharia agrícola, agronomia, computação, entre outras disciplinas).

Com o aumento das ferramentas e plataformas computacionais disponíveis para acesso livre, será cada vez mais relevante que especialistas de domínio agro se capacitem em novos métodos e técnicas da computação científica para analisar seus dados de pesquisa. Não só os pesquisadores, mas gestores agropecuários, extensionistas e empresários agrícolas necessitarão de uma capacitação técnica em dados e software no contexto de *Data Science*, a fim de favorecer a adoção de agricultura/ pecuária de precisão, controles de processos produtivos e logísticos, a gestão de risco a partir de dados de diversas naturezas. É um novo patamar de gestão agropecuária que se configura.

Muitas pesquisas só poderão ser conduzidas caso os cientistas tenham acesso a dados controlados por uma terceira parte, sejam dados de satélites, meteorológicos ou hídricos. No



caso de dados produtivos, pesquisadores necessitarão estabelecer acordos de uso de dados com os produtores, com uma contrapartida de sigilo dos dados. No contexto da Ciência de Dados no Brasil, há que se definir políticas, mercados e questões contratuais relacionadas com propriedade dos dados gerados e padrões de privacidade e segurança, assim como envidar esforços para capitalizar sinergias entre dados públicos e privados.

Ainda existem muitos desafios tecnológicos, de aprendizado e institucionais para a pesquisa agropecuária intensiva em dados. Mas, a fim de manter a liderança brasileira na geração de conhecimentos e tecnologias para a agricultura tropical, há que se integrar conhecimentos já existentes e abrir novas frentes de investigação para participar ativamente da revolução da *e-Science* agrícola. Ressalta-se que estudos de monitoramento tecnológico e de *benchmarking* são essenciais para acompanhar as tendências da fronteira do conhecimento, para embasar políticas públicas com foco estratégico e de longo prazo, integrando as temáticas da sustentabilidade e da inclusão social às pesquisas em Agricultura Intensiva em Dados e à Agricultura 4.0.

## 6. Referências

AGMIP - *Agricultural Model Intercomparison and Improvement Project*. Disponível em: <<http://www.agmip.org/about/>> Acesso em: 12.fev.2017.

APPEL, A. L., MACIEL, M. L.; ALBAGLI, S. A e-Science e as novas práticas de produção colaborativa de conhecimento científico. *Revista Internacional de Ciencia y Sociedad*, v. 3, n. 1, 2016.pp. 41-52.

ARS - Agricultural Research Service. United States Department of Agriculture – USDA. *Big Data and Computing Building a Vision for ARS Information Management*. Workshop Summary Feb. 5-6, 2013. Disponível em: <[https://www.ars.usda.gov/ARSPUserFiles/20800500/bigDataReport\\_Mar-7-2013.pdf](https://www.ars.usda.gov/ARSPUserFiles/20800500/bigDataReport_Mar-7-2013.pdf)> Acesso em: 13.fev.2017.

BAMBINI, M. D., GIACHETTO, P. F., FALCAO, P. R. K., & de OLIVEIRA, F. S. Emerging biotechnologies: bioinformatics services applied to agriculture. In: *Congresso Latino-Iberoamericano de Gestão da Tecnologia, 16., 2015, Porto Alegre. Inovação para além da tecnologia: anais*. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2015.

BELL, G. Foreword. In: HEY, T., TANSLEY, S. , TOLLE, K. M. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009. pp. xi-xiv.

BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge. , v. 323, n. 5919, p. 1297-1298, 2009.

BOOTE, K. *Modeling Climate and Environment Effects on Food Security* . Data Summit: Changing the Face, Place, and Space of Agriculture. 10.Oct.2016. Rosemont, Illinois: Big 10 Conference Center/NIFA/USDA, 2016. (Comunicação Oral)



CGIAR - *Consultative Group on International Agricultural Research*. Disponível em: <http://www.cgiar.org/> Acesso em: 20 mar. 2018.

COBLE, K. *Data Challenges as Seen Through the Public Lens. Data Summit: Changing the Face, Place, and Space of Agriculture*. 10.Oct.2016. Rosemont, Illinois: Big 10 Conference Center/NIFA/USDA, 2016. (Comunicação Oral)

COBLE, K., GRIFFIN, T., AHEARN, M., FERRELL, S., MCFADDEN, J., SONKA, S., & FULTON, J. Advancing US Agricultural Competitiveness with Big Data and Agricultural Economic Market Information, *Analysis, and Research* (No. 249847). Council on Food, Agricultural, and Resource Economics (C-FARE), 2016. 17p.

COWHEY, P.F., ARONSON, J.D., ABELSON, D. Transforming Global Information and Communication Markets: The Political Economy of Innovation. Cambridge, USA: Massachusetts Institute of Technology, 2009. 341 p.

DAVENPORT, T. H.; PATIL, D. J. Data scientist. *Harvard business review*, v. 90, n. 5, p. 70-76, 2012.

DIEBOLD, F. X. "On the Origin(s) and Development of the Term "Big Data". PIER Working Paper 12-037. Philadelphia, PA: Penn Institute for Economic Research, 2012.

FAO - Food and Agriculture Organization of the United Nations. International Treaty on Plant Genetic Resources for Food and Agriculture. Disponível em: <http://www.fao.org/plant-treaty/overview/en/> Acesso em: 21 Mar.2018.

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. Programa Equipamentos Multiusuários (EMU). Disponível em: <http://www.fapesp.br/emu/> Acesso em: 22 Mar.2018

GODAN - Global Open Data for Agriculture and Nutrition. Disponível em: <http://www.godan.info/> Acesso em 21.Mar.2018

GOYAL, A.; GONZÁLES-VELOSA, C. Improving agricultural productivity and market efficiency in Latin America and the Caribbean: how ICTs can make a difference? *Journal of Reviews on Global Economics*. Washington, D.C., v.2, p. 172-182, 2013.

GUSTAFSON, D., HAYES, M., JANSSEN, E., LOBELL, D. B., LONG, S., NELSON, G. C., PAKRASI, H. B., RAVEN, P., ROBERTSON, G. P., ROBERTSON, R., WUEBBLES, D. Pharaoh's dream revisited: an integrated US Midwest field research network for climate adaptation. *BioScience*, 66(1), 2015. pp.80-85.

HEY, T., TANSLEY, S. , TOLLE, K. M. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009. 252p.

HEY, A. J.G., TREFETHEN, A.E. The data deluge: An e-science perspective. In: BERMAN, F., FOX, G., HEY, A. J.G. (Ed.). *Grid computing: making the global infrastructure a reality*. John Wiley and sons, 2003. pp 809-824.



HEY, T., TANSLEY, S. , TOLLE, K. M. The fourth paradigm: data-intensive scientific discovery. Redmond, WA: Microsoft Research, 2009. 252p.

IGAD – Agricultural Data. Research Data Alliance. Disponível em: <https://www.rd-alliance.org/rda-disciplines/rda-and-agriculture> Acesso em: 22 Mar 2018.

INRA - Institut national de la recherche agronomique. Disponível em: <http://www.inra.fr/en> Acesso em: 21 Mar 2018

LOPES, M. Artigo: A Pesquisa Pública e a Era Big Data. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/1578339/artigo-a-pesquisa-publica-e-a-era-big-data> Acesso em: 21 Mar 2018.

PARR, C. *Current and projected research data storage needs of Agricultural Research Service researchers in 2016. Ag Data Commons.* (2017). Disponível em: <http://dx.doi.org/10.15482/USDA.ADC/1346946> Acesso em: 20 Mar 2018

PORTER, A.; BANKS, J; ROPER, A. T.; MASON,T.; ROSSINI, F.; WIEDERHOLT, B. *Forecasting and management of technology.* New York, NY: John Wiley and Sons, 1991. pp. 114-132.

POSPIECH, M.; FELDEN, C. Big data—a state-of-the-art. ( July 29, 2012). *AMCIS 2012 Proceedings. Paper 22.* Disponível em: <<http://aisel.aisnet.org/amcis2012/proceedings/DecisionSupport/222012>> Acesso em: 24 Jan 2017.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. *Zoneamento Agrícola.* Disponível em: <http://www.agricultura.gov.br/assuntos/riscos-seguro/risco-agropecuario/zoneamento-agricola> Acesso em: 21 mar. 2018.

MASSRUHÁ, S. M. F. S., & LEITE, M. D. A. Agro 4.0 - rumo à agricultura digital. In In: MAGNONI JÚNIOR, L.; STEVENS, D.; SILVA, WTL da; VALE, JMF do; PURINI, SR de M.; MAGNONI, M. da GM; SEBASTIÃO, E.; BRANCO JÚNIOR, G.; ADORNO FILHO, EF; FIGUEIREDO, W. dos S.; SEBASTIÃO, I.(Org.). *JC na Escola Ciência, Tecnologia e Sociedade: mobilizar o conhecimento para alimentar o Brasil.* 2. ed. São Paulo: Centro Paula Souza, 2017.

MOHTAR, R. H.; DAHER, B. Water, energy, and food: The ultimate nexus.In: Encyclopedia of agricultural, food, and biological engineering. *CRC Press*, Taylor and Francis Group, 2012.

OCDE - Organização para a Cooperação e Desenvolvimento Econômico. *Alternative Futures for Global Food and Agriculture.* Paris: OECD Publishing, 2016.

RAMASWAMY, S. *Opening Remarks at the Data Science in Agriculture Summit - Changing the Face, Place, and Space of Agriculture. Data Summit: Changing the Face, Place, and Space of Agriculture.* 10.Oct.2016. Rosemont, Illinois: Big 10 Conference



Center/NIFA/USDA, 2016. Disponível em: <<https://www.youtube.com/watch?v=MsNcNPAPBxY>> Acesso em: 12.fev.2017. (Comunicação Oral)

RAMOS, P. Referencial teórico e analítico sobre a agropecuária brasileira. In: Pedro Ramos (org.) BUAINAIN, A.M., BELIK, W., REYDON, B. P., GUEDES, S.N.R., HOFFMANN, R., SILVA, J. G. DA, TAKAGI, M., ROMEIRO, A. R. , SILVEIRA, J. M. F.J. DA , BORGES, I. de C., FONSECA, M. da G. D. *Dimensões do agronegócio brasileiro – políticas, instituições e perspectivas*. Brasília: MDA, 2007. pp.18-52.

REARDON, T., ECHEVERRIA, R., BERDEGUÉ, J., MINTEN, B., LIVERPOOL-TASIE, S., TSCHIRLEY, D., ZILBERMAN, D. Rapid transformation of food systems in developing regions: Highlighting the role of agricultural research & innovations. *Agricultural Systems*, 26 Feb 2018, 2018.(In press, corrected proof)

ROMANI, L. A. S., BAMBINI, M. D., COLTRI, P. P., JUNIOR, A. L., KOENIGKAN, L. V., OTAVIAN, A. F., EVANGELISTA, S. R. M., ZULLO-JR., J. , PINTO, H. S., ASSAD, E. D. Sistema de Monitoramento Agrometeorológico-Agritempo: inovação em rede apoiando políticas públicas e a tomada de decisão agrícola. *Agrometeoros*, 24(1), 2016.

SCHWAB, K. *The fourth industrial revolution*. Cologny/Geneva: World Economic Forum®, 2016. 171p.

SCHNABLE, P. *Models for Big Data and Genomics. Data Summit: Changing the Face, Place, and Space of Agriculture*. 10.Oct.2016. Rosemont, Illinois: Big 10 Conference Center/NIFA/USDA, 2016. (Comunicação Oral)

SOUZA, K. X. S. de, TERNES, S., OLIVEIRA, S. D. M., MOURA, M. F., BARIONI, L. G., HIGA, R. H., FASIABEN, M. A prospective study on the application of Data Science in agriculture. In: *Congresso Brasileiro de Agroinformática*, 11., Campinas. Ciência de dados na era da agricultura digital: anais. Campinas, Editora da Unicamp: Embrapa Informática Agropecuária, 2017.

THE ECONOMIST. The future of agriculture. *Technology Quarter*. Jun 11th, 2016. Disponível em: <http://www.economist.com/technology-quarterly/2016-06-11> Acesso em: 18.Dez.2016.

UNICAMP - *Universidade Estadual de Campinas*. Disponível em: <http://www.unicamp.br/> Acesso em: 21 Mar. 2018.

USP - *Universidade de São Paulo*. Disponível em: <http://www5.usp.br/> Acesso em: 22 Mar. 2018.

WAGENINGEN University. Disponível em: <https://www.wur.nl/> Acesso em: 20 Mar 2018.

WOODARD, J. D. Data science and management for large scale empirical applications in agricultural and applied economics research. *Applied Economic Perspectives and Policy*, v. 38, n. 3, 2016.pp. 373-388.