|  |  |
| --- | --- |
| **Título:**  Unlocking the Power of Big Data at the National Institutes of Health | |
| **Local de Publicação:**  Bethesda, Maryland, Estados Unidos da América, Ano 2013.  <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5206693/> | |
| **Responsável pelo fichamento:**  Maurício Collaça Ramos | |
| **Autor principal (nome e link para o Lattes ou outra fonte):** | Meghan F. Coakley  <https://www.researchgate.net/profile/Meghan_Mccarthy3> |
| **Tema\Assunto:** | Oportunidades e obstáculos para alcançar todo o potencial do Big Data na área de Saúde |
| **Abordagem metodológica utilizada:** | Pesquisa exploratória com abordagem qualitativa na análise de documentos. |
| **Principais conceitos, terminologia e enfoques que sustentam a argumentação:** | Big data – quando a complexidade e a quantidade de dados são grandes o bastante para transformar a forma como os dados são analisados tradicionalmente.  NIAID - Escritório de Ciber-infraestrutura e Biologia Computacional do Instituto Nacional de Alergias e Doenças Infecciosas dos Estados Unidos da América.  NIH – Instituto Nacional da Saúde dos Estados Unidos da América.  NAR – Nucleic Acid Research é um jornal científico da Oxford University Press. |
| **Fatos históricos, dados e outros autores que apóiam a argumentação:** | 2012 – O Escritório de Ciência da Casa Branca lançou a iniciativa federal de Big Data.  O NIAD endereçou em seu terceiro festival annual de bioinformática os desafios atuais em ciência de dados e análise big data.  6/2/2013 – O NIH apresentou um simpósio ‘‘Data Science: Unlocking the Power of Big Data’’ com a participação de diversos grupos de espcialistas da academia e setores público e privado.  A Coleção de Bancos de Dados de Biologia Molecular do NAR aumentou de 300 bancos de dados em 2001 para 1512 bancos de dados em 2013. |
| **Idéias centrais, proposições ou teses defendidas:** | A biotecnologia avança a passos largos e com ela crescem os bancos de dados inerentes.  Entretanto, este disponibilidade de dados é vista também de forma mais oportunista do que científica.  Inovações em aprendizado de máquina caminham para encontrar maiores descobertas nestes bancos de dados mas convivem com a falta de infra-estrutura robusta e algotimos escaláveis para a integração de todas estas bases.  O pré-processamento e limpeza dos dados é um grande desafio imposto aos cientistas que uma vez superado irá adicionar valor aos dados atualmente analisados.  O desafio também está associado a falta de profissionais treinados em informática e ciências da vida, aptos a analisar, interpretar e compartilhar suas descobertas, que muita vez pode significar escrever código de programação, entretanto já existem aplicações web capazes de eliminar muitas tarefas de programação e estas estão sendo continuamente aperfeiçoadas.  O big data caminha mais lentamente nas áreas da saúde e pesquisa científica do que em entidades comerciais e setor privado por questões óbvias: Analisando grandes volumes de dados esta últimas descobrem tendências dos consumidores e incrementam as vendas.  O setor de pesquisa biomédica é bastante restringido, por questões de privacidade, em ter acesso a registro médicos específicos por paciente, muitas das vezes tendo apenas acesso a dados agregados. Outras vezes os desafios são integrar bases de dados de entidades distintas, as quais podem ter políticas diferentes de segurança e publicidade de suas informações.  Os custos de se implementar uma infra-estrutura de computação de alta-performance é um grande impeditivo, particularmente para ambientes de pesquisa acadêmica. |
| **Idéias secundárias, proposições ou teses que**  **complementam o pensamento do autor:** | Órgãos do governo federal, defesa e inteligência dos Estados Unidos já vem se preocupando bastante com a questão da segurança da informação e isto pode se tornar uma oportunidade para que órgãos de saúde possam se basear e desenvolver suas infra-estrutura e protocolos de segurança.  Os desenvolvimento da computação em nuvem começam a endereçar a questão dos custos mencionada acima, entretanto, a comunidade científica ainda não adotou a computação em nuvem efetivamente. |
| **Outros artigos e publicações do mesmo autor:** | The NIH 3D Print Exchange: A Public Resource for Bioscientific and Biomedical 3D Prints.  Coakley MF, Hurt DE, Weber N, Mtingwa M, Fincher EC, Alekseyev V, Chen DT, Yun A, Gizaw M, Swan J, Yoo TS, Huyen Y.  3D Print Addit Manuf. 2014 Sep 1;1(3):137-140. doi: 10.1089/3dp.2014.1503.  3D Printing in the Laboratory: Maximize Time and Funds with Customized and Open-Source Labware.  Coakley M, Hurt DE.  J Lab Autom. 2016 Aug;21(4):489-95. doi: 10.1177/2211068216649578. Epub 2016 May 19. |
| **Questões para discussão:** | Como encarar os desafios técnicos do big data na saúde como o volume, crescimento, disparidades, heterogeneidades e privacidade de diferentes bancos de dados.  Como capacitar cientistas de dados e bio-estatísticos em big data, necessários à pesquisa na área de saúde.  Como treinar e disponibilizar acesso a infraestrutura computacional que permita profissionais de saúde não experientes em tecnologia da informação serem capazes de interpretar resultados e extrair conhecimento para uso prático e imediato.  Como reunir especialistas interessados em colaborar na adoção de padrões tecnológicos e modelos de dados, aprendendo as melhores práticas e maximizando o potencial do conhecimento big data. |