



pdu

Plano Diretor da Unidade

2016 ■ 2020

Minuta

Julho de 2016

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES



PRESIDÊNCIA DA REPÚBLICA

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÃO E COMUNICAÇÃO.

Presidente da República

DILMA ROUSSEF

Vice-Presidente da República

MICHAEL TEMER

Ministro da Ciência, Tecnologia e Comunicações

GILBERTO KASSAB

Secretário Executivo do Ministério da Ciência, Tecnologia e Comunicações

ELTON SANTA FÉ ZACARIAS

Subsecretário de Coordenação das Unidades De Pesquisa

FÁBIO DE PAIVA VAZ

Coordenador Geral das Unidades de Pesquisa

KAYO JULIO CESAR PEREIRA

LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA

Diretor

AUGUSTO CESAR GADELHA VIEIRA

Coordenador de Administração

ANMILY PAULA DOS SANTOS MARTINS

Coordenador de Ciência da Computação

JAUVANE CAVALCANTE DE OLIVEIRA

Coordenador de Matemática Aplicada e Computacional

FREDERIC GERARD CHRISTIAN VALENTIN

Coordenador de Mecânica Computacional

MÁRCIO ARAB MURAD

Coordenador de Sistemas e Controle

CARLOS EMANUEL DE SOUZA

Coordenador de Sistemas e Redes

WAGNER VIEIRA LÉO

© 2016 Laboratório Nacional de Computação Científica

Av. Getúlio Vargas, 333

25651-075 - Petrópolis –RJ

Telefone: (24) 2233-6000 / Fax:(24) 2231-5595

[HTTP://www.lncc.br/](http://www.lncc.br/)

MCTIC – Ministério da Ciência, Tecnologia, Inovação e Comunicação

Subsecretária de Coordenação das Unidades de Pesquisa

Esplanada dos Ministérios, Bloco E

70067-900 - Brasília – DF

Tel: (61)-3317-7607

Fax: (61)-3317-7768

[HTTP://www.mct.gov.br](http://www.mct.gov.br)

LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA

SUMÁRIO

APRESENTAÇÃO DO PLANO DIRETOR DO LNCC	6
1. O LNCC	6
1.1. Missão	8
1.2. Visão	12
1.3. Princípios e Valores	13
1.4. Objetivos Estratégicos e Metas Institucionais	13
2. PILAR ESTRATEGICO I: PROMOÇÃO DAS PESQUISAS EM CIÊNCIA BÁSICA E TECNOLÓGICA	15
2.1 MODELAGEM COMPUTACIONAL	15
2.1.1 Modelagem de Sistemas Complexos	15
2.1.2 Métodos Numéricos e Algoritmos	19
2.1.3 Sistemas, Controle e Sinais	24
2.1.4 Ciência de Dados	27
2.1.5 Computação de Alto Desempenho	39
2.2 Aplicações da Modelagem Computacional em Áreas Estratégicas	42
2.2.1 Ciências da Vida	42
2.2.3 Medicina e Saúde	46
2.2.4 Energia e Recursos Naturais	51
3. PILAR ESTRUTURANTE II: MODERNIZAÇÃO E AMPLIAÇÃO DA INFRAESTRUTURA DE CT&I	55
3.1 SINAPAD - Sistema Nacional De Computação De Alto Desempenho	56
3.2 Supercomputador Santos Dumont	58
3.3 Labinfo	60
3.4 Inct-Macc	64
3.5 INCT-CiD	66
3.6 LIneA	67

3.7 Outros Projetos que Recebem Apoio do LNCC	68
4. PILAR ESTRUTURANTE III: FORMAÇÃO, ATRAÇÃO e FIXAÇÃO DE RECURSOS HUMANOS	69
4.1 Pós-Graduação em Modelagem Computacional	69
4.2 Divulgação Científica: Difusão do conhecimento	71
4.3 Intercâmbios	75
5. PILAR ESTRUTURANTE IV: PROMOÇÃO DA INOVAÇÃO TECNOLÓGICA NAS EMPRESAS	76
6. ORGANIZAÇÃO ADMINISTRATIVA	78

APRESENTAÇÃO DO PLANO DIRETOR DO LNCC

Este Plano Diretor estabelece as orientações para a atuação do Laboratório Nacional de Computação Científica (LNCC) no período de 2016 a 2020. O documento sintetiza e consolida as propostas discutidas por pesquisadores, tecnologistas, funcionários e especialistas no processo de Planejamento Estratégico durante a vigência do PDU 2011-2015 e as compatibiliza com as formulações decorrentes do planejamento estabelecido pelo Ministério da Ciência, Tecnologia, Inovação e Comunicação (MCTIC) através da Estratégia Nacional de Ciência, Tecnologia e Inovação (ENCTI 2016-2019).

Vale ressaltar que o processo de elaboração dos PDUs 2006-2010 e 2011-2015 contribuiu fortemente para a melhoria do desempenho institucional, trazendo maior mobilização e transparência sobre os valores e prioridades institucionais.

Em 2016, novos desafios estão postos em função do cenário de incertezas do país, dos avanços nas áreas de pesquisa, da redução do quadro de funcionários e da ampliação substancial da plataforma computacional do LNCC com a aquisição do supercomputador Santos Dumont (SDumont), em 2015, e início de sua operação em janeiro de 2016.

Os recentes avanços da Computação e áreas correlatas aperfeiçoam e inovam métodos de solução de problemas, ampliando ou criando novas áreas de pesquisa. Como exemplo, a digitalização quase universal das informações disponíveis aliada ao advento da Internet permitiu acesso a uma imensa quantidade de dados relativos a fenômenos naturais, sociais, econômicos e biológicos. São dados gerados por medições em experimentos científicos, na observação direta de fenômenos naturais e sociais, em simulações computacionais e, cada vez mais, de forma automática, por sensores acoplados a todo tipo de objeto, como na coleta de dados de saúde de indivíduos ou de informações georeferenciadas por GPS via celulares. A Internet, hoje ampliada com a comunicação entre sensores, constituindo a chamada “Internet das Coisas” (*IoT*), permite a agregação dessas informações digitalizadas que podem ser manipuladas por novas técnicas de processamento e análise de grande massa de dados, permitindo uma melhor compreensão e previsão do comportamento de fenômenos físicos, sociais e econômicos.

A busca de novos métodos e algoritmos para o uso eficiente de grandes bases de dados originou a *Ciência dos Dados*, que inclui as denominadas técnicas de *Big Data*. O uso dessas técnicas na pesquisa científica criou um novo paradigma de se fazer ciência, para o qual os métodos de computação de alto desempenho em supercomputadores são essenciais. Por outro lado, podemos citar vários outros desenvolvimentos recentes que merecem a atenção dos pesquisadores em Computação Científica: o advento de tecnologias de redes interligando em alta velocidade sistemas computacionais de grande

porte permitiu o surgimento dos “sistemas em redes” (*Networked Systems*) gerando problemas teóricos desafiadores em diversas áreas, entre as quais a Teoria de Contrôles; o algoritmo *PageRank* criado pelo Google deu origem a um intenso tópico de pesquisa sobre máquinas de busca (*search engines*) na Internet com critérios de classificação das informações; os projetos de veículos auto guiados (AGV – *Automated Guided Vehicle*), financiados por grandes empresas de TIC e automotivas, têm impulsionado a área de aprendizagem de máquinas (*Learning Machine*) e as pesquisas associadas aos algoritmos de aprendizagem.

O desenvolvimento de novos métodos numéricos se faz necessário para o uso eficiente de computadores cada vez mais potentes na solução, impossível com os atuais computadores, de problemas de mais alta complexidade. Hoje os supercomputadores alcançam velocidades de processamento de dezenas de petaflops – 10^{15} operações aritméticas em ponto flutuante por segundo –, com previsão de, em 10 anos, alcançarem a faixa exaflopica – 10^{18} flops – (flops é o acrônimo de *floating point operations per second*). A disponibilidade de recursos computacionais de mais alto desempenho também impulsiona o desenvolvimento e a aplicação da Computação Gráfica em áreas de grande interesse na Medicina, na Biologia, na Modelagem Molecular e nas Engenharias, permitindo uma melhor compreensão do comportamento de sistemas e fenômenos através da visualização dinâmica de sua representação gráfica.

Essas e outras áreas de pesquisa, que conjugam questões teóricas altamente relevantes com grande interesse em aplicações, são contempladas nas atividades de P&D do LNCC previstas nesse PDU.

Outro aspecto importante para o LNCC é sua atuação como infraestrutura computacional de alto desempenho disponível à comunidade científica e como instituição coordenadora do Sistema Nacional de Processamento de Alto Desempenho (SINAPAD), fortalecida com a aquisição do SDumont. Atualmente, este é o supercomputador com maior capacidade de processamento da América Latina e está entre os 300 maiores no mundo, ampliando substancialmente a oferta de processamento computacional do nó principal do SINAPAD com grande impacto para as pesquisas científicas nacionais que necessitam de computação de alto desempenho. Emblematicamente, a capacidade nominal de processamento do SDumont de 1.1 petaflops, mostrou-se extremamente atraente para a comunidade científica que submeteu 75 projetos no primeiro edital fechado em maio de 2016, originados de todas as regiões do país. A forte demanda evidencia a importância de se ampliar a capacidade computacional do país para atender nossas universidades e centros de pesquisa.

A formulação do Plano Diretor 2016-2020 reflete os desafios identificados pelo corpo técnico-científico do Laboratório, para a pesquisa, a formação de recursos humanos, a gestão de serviços computacionais, a cooperação científica, a inovação e a transferência de tecnologia para o setor produtivo.

O PDU está organizado segundo os pilares da ENCTI para facilitar a compreensão do papel da instituição na Estratégia Nacional de CT&I. Como conclusão do processo de planejamento, as linhas de ação da instituição foram priorizadas e alinhadas à sua missão e visão de futuro. Outra diretriz importante foi garantir a compatibilização com o Plano Plurianual 2016-2020.

O plano define as linhas de ação, objetivos e metas. As metas foram construídas tomando como base os projetos, os grupos de pesquisa compostos de pesquisadores e tecnologistas atualmente trabalhando em cada área e o histórico das realizações dos últimos cinco anos. Vale ressaltar que esse dimensionamento projeta uma atuação e importância crescente do LNCC como centro de excelência em pesquisa, formação de recursos humanos e oferta de computação de alto desempenho.

Essa atuação e a expectativa do sucesso da execução do Plano Diretor tem como premissa fundamental a recomposição dos recursos orçamentários a partir de 2017, considerando como referência o orçamento de 2012, valor histórico para a adequada manutenção de sua infraestrutura física e desenvolvimento de seus projetos, acrescido dos recursos necessários para a manutenção e operação da plataforma computacional do SDumont que representam um item significativo de custo não presente naquele orçamento anterior.

Outra premissa de grande relevância é a adequação da dimensão do corpo técnico-científico e administrativo em quantidade compatível com as atividades programadas e o pleno atendimento das metas assumidas. Durante o período coberto pelo Plano Diretor, deverão ser adotadas medidas para estabelecer um quadro de servidores, administrativos, tecnologistas e pesquisadores, de forma a garantir o pleno aproveitamento da maturidade científica e da plataforma tecnológica de alto nível consolidada pelo LNCC em seus 35 anos de existência.

1. O LNCC

Desde a sua origem, em 1980, o LNCC tem como missão a pesquisa, o desenvolvimento e a formação de recursos humanos em Computação Científica, o que lhe conferiu um papel fundamental na consolidação dessa então nova área de conhecimento no Brasil. Completando sua missão, o LNCC disponibiliza à comunidade científica serviços computacionais de alto desempenho.

Antes de alcançar *status* de Laboratório Nacional, ainda como “Laboratório de Computação Científica”, agregou diversos grupos com interesse em problemas originados, dentre outras áreas, na dinâmica social, em sistemas e controle, na física, nos fenômenos de transporte e nas engenharias. Em particular, as metodologias utilizadas na solução dos fenômenos de transportes e em problemas nas engenharias, governados por sistemas de equações diferenciais parciais, eram, basicamente, a Análise Matemática, para provar a existência, unicidade e regularidade da solução, e os Métodos de Discretização e Análise Numérica, para determinar a consistência, estabilidade, convergência e precisão dos algoritmos utilizados na solução das equações. O processamento numérico dos modelos era realizado pelos recursos computacionais muito limitados disponíveis nos anos 80.

Na formação de pesquisadores, mesmo não dispondo de programas de graduação ou pós-graduação até o ano 2000, o LNCC contribuía com a orientação de teses de doutorado e dissertações de mestrado em instituições de ensino nacionais e internacionais. Além disso, o LNCC sempre orientou bolsas de iniciação científica e participou de programas como o PIBIC.

Em outra vertente, pela promoção e participação em escolas temáticas, seminários, organização de congressos, *workshops* e outros eventos científicos, o LNCC potencializava o intercâmbio técnico científico em níveis nacional e internacional e contribuía para a disseminação de novas metodologias para formulação, análise e aplicações da Computação Científica.

No final da década de 90 e primeiros anos do novo milênio, o LNCC começou a atuar fortemente na área de Bioinformática, inaugurando o Laboratório Nacional de Bioinformática e a Unidade de Genômica Computacional Darcy Fontoura de Almeida, e em aplicações na Medicina com a implantação do laboratório Hemolab e, em 2009, do INCT Medicina Assistida por Computação Científica – MACC.

Com a criação do programa de pós-graduação em Modelagem Computacional no ano 2000, dois anos após a inauguração de sua nova sede em Petrópolis, o Laboratório passou a contribuir diretamente para a formação, multi- e transdisciplinar, de pesquisadores oriundos de diferentes áreas de conhecimento (Engenharia, Matemática, Computação, Biologia, Física e Ciências Humanas).

A Computação Científica e disciplinas correlatas foram potencializadas pela notável evolução dos equipamentos e da ciência da computação e das tecnologias de informação e comunicação, com o crescimento exponencial da capacidade de processamento e da velocidade das redes de transmissão de dados. O supercomputador de maior capacidade

da América Latina aliado à conectividade *multiGigabit* proporcionada pela Rede Nacional de Ensino e Pesquisa (RNP) permite ao LNCC disponibilizar uma infraestrutura de alto desempenho computacional que pode ser usada pela comunidade científica nacional na aplicação dos métodos avançados de Computação Científica em suas pesquisas.

Essa evolução da computação aliada à conectividade teve forte impacto não apenas no suporte à pesquisa científica, com o consequente crescimento na qualidade e amplitude dos resultados; influenciou sobremaneira a própria forma de se fazer ciência, de se fazer pesquisa. A Computação Científica, portanto, evoluiu do *status* de um contribuidor na criação de conhecimento para o de um componente essencial nos processos de inovação científica e tecnológica e de ganho de competitividade das economias. A modelagem e a simulação atualmente são peças estratégicas para a criação de novos conhecimentos e para o desenvolvimento de produtos e serviços inovadores.

Neste contexto, a Computação Científica como ramo do conhecimento que tem por objetivo criar modelos e métodos matemáticos e computacionais para compreender, analisar e resolver problemas científicos e tecnológicos constitui uma grande área de pesquisa interdisciplinar, fundamentada em conhecimentos científicos e metodologias advindos primordialmente da Matemática e da Computação. Busca avançar no desenvolvimento de modelos, métodos, algoritmos e técnicas para simular condições, testar hipóteses, controlar e prever a evolução de processos e fenômenos. É alternativa cada vez mais utilizada para técnicas e observações da ciência experimental, principalmente nos casos em que as medições são impraticáveis, de alto risco ou muito custosas. Encontra aplicações em inúmeras áreas científicas e tecnológicas, pelo que transcende o universo acadêmico e chega ao governo, à indústria, ao comércio, aos serviços e à sociedade.

A aquisição do supercomputador SDumont em 2015 representou um marco fundamental para a alavancagem da computação de alto desempenho no Brasil. Esse é um divisor de águas extremamente relevante para toda a comunidade científica do país que passa a dispor de uma plataforma capaz de lidar com os desafios de velocidade, volume e variedade que caracterizam o processamento de alto desempenho aplicado na solução de problemas de alta complexidade que envolvem grande número de cálculos numéricos e de manipulação de dados. O SINAPAD disponibiliza agora para o SNCTI uma capacidade computacional com velocidade petaflopica e armazenamento na escala de petabytes. Temas de grande impacto, como pesquisas relacionadas ao vírus da zika, a doenças coronarianas, ou ao gerenciamento de parques eólicos, já estão sendo processados no SDumont.

O LNCC, como um dos Operadores do Sistema Nacional de CT&I, contribui significativamente para o avanço da ciência por meio da realização de pesquisas científicas e desenvolvimentos tecnológicos em Computação Científica e suas aplicações, da formação de novos pesquisadores, da disponibilização e facilitação do uso da sua infraestrutura computacional de alto desempenho para o meio acadêmico e setor empresarial e da promoção e disseminação da ciência, em benefício da sociedade brasileira e do desenvolvimento do país.

MISSÃO

O LNCC orienta-se pelas perspectivas da relevância global e do alto valor estratégico da Computação Científica, bem como pelo seu mandato de atuar como um Laboratório Nacional disponibilizando uma infraestrutura de computação de alto desempenho para o uso compartilhado com toda a comunidade de pesquisa científica e tecnológica do país. Nessa qualidade, contribui ativamente para o desenvolvimento autônomo do País na área estratégica em que atua.

Tem como atividade precípua a realização de pesquisa e desenvolvimento em métodos matemáticos e computacionais e construção de modelos, algoritmos, técnicas e aplicações na solução de problemas em temas relevantes para a sociedade, para as ciências e para o desenvolvimento do país.

Alia à sua finalidade central a formação de novos pesquisadores em Modelagem Computacional, com elevado grau de qualificação e perfil interdisciplinar, em nível de pós-graduação. Promove, direta e indiretamente, por si ou em redes cooperativas, a realização de pesquisa científica avançada e inovadora e a atualização e o emprego de processamento de alto desempenho, junto às instituições componentes do Sistema Nacional de Ciência e Tecnologia e Inovação (SNCT&I).

Acrescenta ainda às suas atribuições a disseminação dos conhecimentos e saberes que compõem o seu capital intelectual, por meio da promoção e da participação em fóruns especializados e de divulgação da ciência.

Esses aspectos e atividades são assim sintetizados na declaração de Missão do LNCC:

- *Realizar pesquisa, desenvolvimento e formação de recursos humanos em Computação Científica, em especial na construção e aplicação de modelos e métodos matemáticos e computacionais na solução de problemas científicos e tecnológicos;*
- *Disponibilizar ambiente computacional para processamento de alto desempenho, tendo como finalidades o avanço do conhecimento e o atendimento às demandas da sociedade e do Estado brasileiro.*

VISÃO

Modelagem e simulação computacionais são estratégicas para todas as áreas do conhecimento, como instrumento de análise e ferramenta de projeto e de tomada de decisões. Diversas áreas demandam continuamente modelos mais complexos e mais refinados, bem como novos modelos estocásticos, acoplados e multiescalas, nas suas dimensões espacial e temporal, criados a partir da matemática, física, química, biologia e computação, dentre outras disciplinas.

Coloca-se nesse contexto a visão de futuro para o LNCC. São muitos os possíveis caminhos. A opção é fazer melhor e aceitar novos desafios em pesquisa, desenvolvimento, inovação e formação de recursos humanos, como tem realizado desde sua criação. Assim, a visão institucional está associada aos seguintes aspectos:

- Às perspectivas de progresso teórico e de ampliação da gama de aplicações dos métodos e técnicas englobados na Computação Científica;
- Às áreas de pesquisa consolidadas no LNCC, o trabalho nelas realizado ao longo de sua existência e a perspectiva de sua continuidade e ampliação;
- Ao avanço tecnológico da computação de alto desempenho;
- Às infraestruturas física e computacional do LNCC disponíveis e projetadas para disponibilidade e capacidade crescentes;
- Ao programa de pós-graduação do LNCC e as possibilidades de seu aprimoramento como programa dinâmico que possa atender a novas demandas;
- Ao predomínio das tecnologias da informação no dia-a-dia das pessoas, na oferta de serviços governamentais, na implementação e monitoramento de políticas públicas se as decorrentes demandas por solução de problemas em todos os campos da ciência e tecnologia.

Formula-se, sobre esse pano de fundo, a visão de futuro do LNCC:

- Fortalecer seu papel como centro estratégico e de excelência em Computação Científica, atuando na fronteira do conhecimento nas suas atividades de pesquisa e desenvolvimento;
- Formação de recursos humanos altamente qualificados;
- Apoio ao Estado Brasileiro, às instituições de CT&I e às empresas através dos conhecimentos gerados e da sua infraestrutura computacional.
- Integrar-se no sistema de inovação nacional, interagindo com instituições e empresas e produzindo inovação e conhecimento fundamentados na excelência da pesquisa, na capacidade de processamento, na atuação na fronteira do conhecimento e na qualificação de recursos humanos.

PRINCÍPIOS E VALORES

Os princípios e valores estabelecidos para orientação da ação institucional refletem a tradição do LNCC como unidade de pesquisa atuante na fronteira do conhecimento, atenta a seu papel perante a comunidade científica e acadêmica e a sociedade. Tem a Ética como padrão essencial de conduta, e norteia-se pela seguinte declaração de princípios:

- Excelência e respeito ao mérito e aos valores científicos,
- Estímulo à criatividade, à colaboração e à cooperação com outras instituições acadêmicas, de CT&I e empresas,
- Estímulo à formação avançada de recursos humanos e à capacitação contínua de seu corpo funcional;
- Dedicção e eficiência na execução das suas atividades, com responsabilidade pública e social e transparência característica de uma instituição aberta à sociedade;
- Obediência aos princípios constitucionais de legalidade, impessoalidade, moralidade, igualdade, publicidade e probidade administrativa.

OBJETIVOS ESTRATÉGICOS E METAS INSTITUCIONAIS

A ENCTI 2016-2019 define como Eixo Estruturante do Sistema Nacional de CT&I a Expansão, Consolidação e Integração do Sistema Nacional de CT&I e estabelece cinco Pilares Estratégicos: i) Pesquisa; ii) Infraestrutura; iii) Financiamento; iv) Recursos Humanos e; v) Inovação.

Para cada um desses Pilares, o LNCC associou suas linhas de ação correspondentes desdobradas em metas que servirão de parâmetros para avaliação de desempenho da

instituição. A exceção é o pilar iii), uma vez que o Laboratório não realiza atividades de fomento.

No processo de construção do PDU foram identificados os seguintes objetivos estratégicos para o período 2016-2020:

- Produzir **pesquisa avançada** em Computação Científica e na aplicação de Modelagem Computacional em áreas estratégicas (Ciências da Vida, Medicina e Saúde, Energia e Recursos Naturais) que contribua para o avanço científico e tecnológico e o bem-estar da sociedade.
- Manter, reformular e expandir a **infraestrutura computacional** considerando as demandas de computação de alto desempenho projetadas para o Sistema Nacional de CT&I até 2020 atendendo as necessidades do SINAPAD e do LNCC.
- Manter a excelência da **formação de recursos humanos** no LNCC com foco na adequação ao mercado do perfil profissional e na evolução constante do conhecimento, apoiado nas competências consolidadas do LNCC em modelagem computacional e suas aplicações.
- Ampliar e aprofundar **as ações para o desenvolvimento e promoção da inovação tecnológica**, através do Núcleo de Inovação Tecnológica e da Incubadora de Empresas utilizando os mecanismos previstos em Lei;
- Coordenação e participação em redes e programas de **cooperação nacional e internacional** que permitam o avanço do conhecimento e o aperfeiçoamento das equipes nas linhas de ação do Laboratório;
- Desenvolver **projetos em colaboração com instituições e empresas públicas e privadas**, em particular, em projetos de modelagem e simulação computacional de reservatórios de petróleo e análise de dutos e *risers* de perfuração, em parceria com o Centro de Pesquisas da Petrobras (Cenpes) e outros centros de P&D, e em projetos nos programas Genoma Brasileiro, Rede Nacional de Bioinformática e Rede de Hemofilia;
- Manter, reformular e expandir a **infraestrutura física e administrativa** do LNCC garantindo a disponibilidade e a capacidade necessárias para dar sustentação às atividades do LNCC.
- Rever a **estrutura organizacional** da instituição para readequação ao cenário de expansão de áreas de pesquisa e da oferta de computação de alto desempenho ao Sistema Nacional de CT&I.
- Garantir a **formação continuada** dos servidores administrativos, pesquisadores e tecnologistas do LNCC.

- Alcançar as Metas ligadas aos Objetivos Específicos das linhas de ação descritas neste PDU.

Para evitar repetições, a seguir são apresentadas Metas gerais que se aplicam a todas as linhas de ação. A maior ou menor ênfase de cada meta será função das características e objetivos específicos dos temas, do nível de maturidade da pesquisa e da disponibilidade de recursos disponibilizados para cada área. Outras metas específicas serão descritas para cada linha de ação.

Metas Gerais:

- Publicação de artigos em periódicos e congressos científicos;
- Apresentação de trabalhos em eventos científicos;
- Formação avançada de recursos humanos com orientação de dissertações e teses de mestrado e doutorado;
- Oferta de cursos e minicursos de graduação e pós-graduação e de divulgação científica;
- Promoção e participação em eventos, redes e atividades de disseminação do conhecimento, no âmbito nacional e internacional;
- Fortalecimento e busca de novas parcerias entre o LNCC, instituições acadêmicas, grupos de P&D e empresas para promover a inovação;
- Fortalecimento das equipes multidisciplinares nas linhas de ação do LNCC;
- Oferta de serviços técnicos e uso dos equipamentos e instalações para entidades e empresas;
- Registro de propriedades intelectuais e participação nos direitos de exploração.

2. PILAR ESTRATEGICO I: PROMOÇÃO DAS PESQUISAS EM CIÊNCIA BÁSICA E TECNOLÓGICA

As grandes linhas de pesquisa do LNCC são Modelagem Computacional e Aplicações de Modelagem Computacional em Áreas Estratégicas.

2.1 MODELAGEM COMPUTACIONAL

2.1.1 MODELAGEM DE SISTEMAS COMPLEXOS

MODELAGEM DE INCERTEZAS:

As simulações numéricas têm se tornado ferramentas cada vez mais importantes na modelagem de problemas em diversas áreas do conhecimento, visando a previsão de

quantidades de interesse e no auxílio à tomadas de decisões. Entretanto, este é um processo que por definição é uma aproximação da realidade que se deseja modelar sendo baseado em hipóteses, suposições, simplificações e dependente de dados. Conseqüentemente, é um processo que vem acompanhado de incertezas e erros. Estas características são muitas vezes agravadas pela precariedade ou ausência de informações sobre parâmetros necessários para caracterizar a fenomenologia que se deseja modelar. Tais incertezas exercem forte impacto sobre a capacidade preditiva dos modelos. Uma das alternativas para contornar esta dificuldade consiste em incorporar dados de campo diretamente na simulação dos processos, visando reduzir as incertezas associadas e melhorar a previsibilidade dos modelos.

Diversas teorias podem ser utilizadas para a representação destas incertezas tais como o *Worst Case Scenario*, *Logica Fuzzy* e Teorias da Possibilidade, Evidência e Probabilidade e Métodos Bayesianos.

É importante ressaltar que na maioria dos casos, devido à quantidade de dados disponíveis e associadas qualidades, é necessário utilizar conjuntamente diferentes teorias.

Este é um tema multidisciplinar, fundamentado e desenvolvido a partir de combinações de conceitos, métodos e princípios que frequentemente abarcam várias áreas da mecânica, matemática, ciência de computação e outras áreas afins. Problemas de áreas como sócio educacionais também se caracterizam pelas incertezas de multifatores. Através de abordagens matemáticas podem ser construídas abordagens para analisar causas e definir estratégias que auxiliem a tomada de decisão em questões como evasão e aprendizado

Objetivos específicos:

- Consolidar e expandir a área de quantificação de incertezas;

Metas:

- Desenvolvimento de métodos estocásticos computacionalmente eficientes e robustos para a resolução de problemas nos quais estejam presentes incertezas nos parâmetros físicos e dados de entrada.
- Desenvolvimento de modelos de incerteza de parâmetros e modelos utilizando as *Logica Fuzzy*, Teoria da Evidência e Probabilidade e procedimentos para a construção dos processos de calibração, validação e predição.

MODELAGEM MULTIESCALA E MULTIFÍSICA:

O LNCC é uma das instituições pioneiras a desenvolver a modelagem multiescala de sistemas físicos com alto teor de heterogeneidade. Neste contexto, o sistema é descrito em escala pequena. As leis constitutivas acuradas são desenvolvidas e a informação proveniente da “microescala” é propagada para uma “escala grossa”, de maior dimensão, onde o sistema heterogêneo pode ser descrito por equações com coeficiente regularizado. Com essa modelagem, simulações numéricas factíveis podem ser realizadas.

Diversos sistemas físicos heterogêneos podem ser descritos pela modelagem multiescala onde podemos salientar reservatórios de óleo e gás, aquíferos para prospecção de águas subterrâneas, solos argilosos e sistemas biológicos.

Atualmente a modelagem multiescala baseada em primeiros princípios tem sido utilizada pelo grupo do LNCC para obtenção de equações efetivas em diversas aplicações tais como meios porosos altamente heterogêneos. Os resultados obtidos na descrição desta classe de sistema têm sido altamente promissores e dado reconhecimento internacional ao LNCC.

As informações contidas nos modelos multiescala são enriquecidas comparadas com os modelos tradicionais puramente macroscópicos uma vez que apresentam correlações precisas entre a resposta efetiva do sistema heterogêneo e os fenômenos descritos nas múltiplas escalas mais finas do problema.

Objetivos específicos

- Desenvolver a área de modelagem multiescala de sistemas físicos com alto teor de heterogeneidade

Metas:

- Desenvolver novas técnicas de modelagem multiescala aplicadas a meios porosos fraturando sujeito a fenômenos de acoplamento hidrogeomecânico. Aplicação a problemas de reativação de falhas em reservatórios de óleo e gás.
- Desenvolver novas técnicas de *upscaling* para descrever movimento de gases em meios porosos com múltiplos níveis de porosidade.
- Desenvolver e utilizar de métodos de análise e construção de modelos matemáticos para a abordagem de problemas que envolvem complexidades, **não linearidades** e/ou acoplamentos e multidependências fenomenológicas/constitutivas em problemas de fenômenos de transporte.

MODELAGEM DE SISTEMA BIOLÓGICOS:

Sistemas biológicos se caracterizam por serem sistemas complexos por excelência, envolvendo grande número de objetos e interações em múltiplas escalas, onde interações em escalas menores criam e dão permanência a objetos que interagem nas escalas maiores, ou mais agregadas. Além do mais, tanto os entes biológicos como seus canais de interação se alteram ao longo do fenômeno tanto quanto os vínculos induzidos por estas alterações, independentemente de escala. Isto é exemplificado por analogias existentes entre a dinâmica intracelular e o comportamento de ecossistemas, mesmo que populações de células sejam componentes de ecossistemas. Contudo, nosso entendimento das interdependências entre escalas e dos condicionamentos causados por alterações nas possíveis interações permanece limitado, devido a restrições teóricas e experimentais.

A análise qualitativa e quantitativa das redes de interação e sua interpretação como sistemas dinâmicos reativos proveem uma coleção de técnicas poderosas para compreendermos fenômenos biológicos e ecológicos. Em geral, o primeiro passo da modelagem consiste na identificação e reconstrução da rede de possíveis interações, seguido da identificação e modelagem do comportamento dinâmico de aspectos e variáveis, egressos desta rede, que descrevem os eventos que ocorrem no fenômeno em estudo.

A reconstrução da rede de interações fornece uma compreensão apenas limitada da essência biológica presente nos fenômenos, visto que interações biológicas dependem do contexto e ocorrem em diferentes escalas de espaço e tempo, diferindo substancialmente em sua natureza de uma instância para a outra. Apesar da crescente evidência a respeito da hierarquia e organização dos sistemas biológicos e ecológicos, a natureza “plana” da representação em rede associada a sistemas dinâmicos não consegue levar em conta estas características. Por outro lado, a natureza adaptativa dos sistemas vivos torna a experimentação recorrente direta muito difícil e pouco confiável, convidando ao uso da experimentação virtual através da simulação computacional. Novos arcabouços de representação computacional e análise que abarquem variações com o contexto, bem como a natureza hierárquica e multiescala de fenômenos biológicos são, portanto, objetos de pesquisa necessários.

Objetivos Específicos:

- Desenvolver o estudo de sistemas biológicos, ecológicos e ambientais, tendo como paradigma a hierarquia e organização biológicas;

- Estabelecer capacitação e formação de recursos humanos na interface entre biologia e modelagem matemático-computacional, centrados no conceito de organização e em formalismos para representá-la.

Metas:

- Elaboração e implementação de novos métodos e procedimentos de modelagem multiescala, dependentes de contexto, que contemplem as características singulares da interação biológica e de sua dinâmica, bem como a hierarquia inerente à organização biológica;
- Desenvolvimento de modelos matemático-computacionais para fenômenos biológicos e ecológicos centrados na organização biológica, que respeitem a fluidez, plasticidade, e a hierarquia inerentes a seus componentes e interações.

2.1.2 MÉTODOS NUMÉRICOS E ALGORITMOS

Procura-se a excelência no desenvolvimento e análise matemática de novas gerações de métodos e algoritmos numéricos e de algoritmos de otimização para a simulação computacional de problemas da física, engenharias e ciências biológicas. Três temas definem as linhas de pesquisa em métodos numéricos e algoritmos:

- Métodos de Elementos Finitos, Diferenças Finitas e Volume Finitos;
- Algoritmos de Otimização;
- Análise Numérica e Adaptatividade.

Os Objetivos Específicos e as Metas se aplicam às 3 áreas.

Objetivos Específicos:

- Tornar-se um centro de referência no desenvolvimento e análise matemática de métodos e algoritmos numéricos adaptados às novas gerações de arquiteturas massivamente paralelas;
- Torna-se um centro de referência na formação de recursos humanos (nível de pós-graduação) no tema desenvolvimento, análise e implementação computacional da nova geração de métodos e algoritmos numéricos;
- Tornar-se um centro de referência na difusão do conhecimento em métodos e algoritmos numéricos, através da organização de encontros, congressos, escolas e cursos.

Metas:

- Desenvolvimento e análise matemática de novos métodos de elementos finitos híbridos estabilizados, estáveis, precisos e livres de oscilações espúrias, que preservem as propriedades conservativas dos modelos contínuos;
- Desenvolvimento e análise matemática de novos métodos de elementos finitos multiescalas para modelos altamente heterogêneos, e implementação na máquina massivamente paralela do LNCC;
- Desenvolvimento de estimadores de erro *a posteriori* baseados nos novos métodos numéricos, combinado com algoritmos adaptativos espaço-tempo;
- Desenvolvimento de metaheurísticas eficientes envolvendo o uso de metamodelos para substituição parcial dos modelos;
- Geração de novas metaheurísticas, eficazes e eficientes, para a resolução de problemas de otimização multinível e multiobjetivo que surgem em Pesquisa Operacional, Mineração de Dados, Engenharia e Modelagem Molecular de Sistemas Biológicos (em especial ligados à otimização e à identificação tanto de parâmetros quanto da estrutura de modelos que melhor expliquem os dados observados);
- Desenvolvimento teórico e aplicações da análise de sensibilidade topológica no contexto de otimização topológica, problemas inversos, processamento de imagens, síntese de novos materiais e modelagem mecânica de fenômenos dissipativos incluindo dano e fratura;
- Desenvolver algoritmos de segunda ordem, não iterativos e independentes de qualquer chute inicial, com potencial aplicação na detecção de fontes poluidoras em bacias hidrográficas, tratamento de câncer por hipertermia, medicina diagnóstica de detecção precoce de câncer, reconstrução de obstáculos imersos em um fluido e reconstrução de anomalias utilizando ondas sonoras;
- Desenvolvimento de métodos de volumes finitos de alta ordem para equações predominantemente hiperbólicas com o tratamento de condições de acoplamento entre domínios espaciais, e implementação dos algoritmos em clusters de alto desempenho computacional com o SDumont.

MÉTODOS DE ELEMENTOS FINITOS, DIFERENÇAS FINITAS E VOLUME FINITOS

Os métodos de elementos finitos, diferenças finitas e volumes finitos são métodos numéricos adaptados à resolução de modelos matemáticos baseados em equações diferenciais. Construídos sobre um sólido embasamento matemático, esses métodos aproximam a solução exata do modelo matemático original (forma forte), no caso do

método de diferenças finitas, ou da forma fraca do modelo original (formulação variacional) no caso dos métodos de elementos finitos e de volumes finitos. Tais abordagens necessitam de uma malha ou discretização do domínio para serem empregados. Os métodos são atrativos do ponto de vista do custo computacional por induzirem sistemas lineares esparsos, por suas qualidades de aproximação e pela flexibilidade em aproximar geometrias complexas (nos casos dos métodos de elementos finitos e volumes finitos, principalmente).

Neste contexto, novos e grandes desafios estão presentes na área de métodos numéricos. De fato, a última década foi marcada por um grande desenvolvimento de computadores com arquiteturas massivamente paralelas. As novas arquiteturas baseiam-se na multiplicação do número de processadores (agrupados em cores) com baixa/média velocidade e capacidade de estocagem. Esse novo paradigma implica na revisão do que se espera dos simuladores computacionais. Embora a precisão e robustez dos métodos numéricos continuem sendo atributos fundamentais para a qualidade das simulações computacionais, estes devem ser capazes de naturalmente incorporarem as características intrínsecas das novas gerações de arquiteturas, além de serem tolerantes a falhas. De fato, falhas durante simulações computacionais de grande vulto são uma certeza neste cenário. Consequentemente, os novos métodos numéricos precisam necessariamente induzir algoritmos assíncronos e que evitem ao máximo comunicações entre processadores.

Os métodos numéricos baseados na estratégia de “dividir para conquistar” preenchem tais imperativos arquiteturais, os quais não estão presentes na construção dos métodos numéricos clássicos. De fato, dividir a computação associada a simulações de altíssimo porte em um imenso conjunto de problemas independentes de pequeno tamanho apresenta-se como a estratégia desejável. Como resultado, limita-se o impacto das falhas de hardware e tira-se proveito da localidade temporal e espacial das simulações.

Este é o cenário desafiador em que os novos métodos numéricos de elementos finitos, diferenças finitas e volumes finitos desenvolvidos no LNCC estão inseridos. Em particular, estamos interessados em desenvolver novos métodos de elementos finitos híbridos e multiescalas para operadores elípticos, parabólicos e hiperbólicos com coeficientes heterogêneos por esses se adequarem perfeitamente ao novo paradigma descrito nos parágrafos anteriores. Os novos de elementos finitos deverão possuir as seguintes características:

- Estabilidade e alta ordem de aproximação dirigidos pela aproximação nas faces da malha grossa do domínio;
- Inclusão de processos de *upscaling* locais e independentes via funções de base;

- Campos localmente conservativos;
- Adaptatividade temporal e espacial via estimadores de erro *a posteriori*;
- Naturalmente adaptadas as novas arquiteturas massivamente paralelas.

Com relação aos métodos de volumes finitos, o foco está em desenvolver métodos de alta ordem para equações predominantemente hiperbólicas. Ênfase especial será dada ao tratamento de condições de acoplamento entre domínios espaciais assim como a fenomenologia de caráter elíptica nas equações, para o qual se necessitam formulações “hiperbolizadas”. Tais tipos de métodos numéricos são fundamentais para a abordagem numérica de problemas hiperbólicos altamente não-lineares e de grande porte e ainda, neste contexto, para a resolução de problemas inversos de identificação de parâmetros e problemas de quantificação de incertezas. Com efeito, estas metodologias permitem a paralelização massiva dos cálculos numéricos com altíssima eficiência. Cabe destacar que a alta ordem, conjugada com a robustez do método numérico, são vitais para a obtenção de um código eficiente do ponto de vista numérico e computacional. Ainda, pretende-se implementar estes algoritmos no âmbito de clusters de alto desempenho computacional como os existentes no LNCC.

ALGORITMOS DE OTIMIZAÇÃO E PROBLEMAS INVERSOS

A descrição matemática de um processo físico é caracterizada em geral por dados de entrada, parâmetros do sistema e dados de saída. Quando são conhecidos os dados de entrada e os parâmetros do sistema e deseja-se encontrar os dados de saída do processo, tem-se em mãos um problema direto, ou seja, a partir das causas, objetiva-se conhecer os efeitos. Por outro lado, quando se deseja encontrar os dados de entrada ou os parâmetros do sistema a partir de observações sobre os dados de saída, tem-se um problema inverso, isto é, a partir dos efeitos, objetiva-se determinar as causas. Tomografia computadorizada, sísmica e gravimetria são exemplos típicos de problemas inversos, cujas aplicações vão desde reconstrução de órgãos internos a partir de sondagem não invasiva, passando por detecção de falhas em componentes estruturais, indo até prospecção de reservas petrolíferas e lençóis freáticos. Uma forma bastante geral de abordar certas classes de problemas inversos consiste em reescrevê-los na forma de problemas de otimização. Em linhas gerais, o processo de otimização consiste na busca pela melhor alocação de um conjunto limitado de recursos, escolhendo a alternativa que maximize o lucro ou minimize o custo, dentre todas aquelas que satisfazem um conjunto dado de restrições. Definido um conjunto de variáveis de projeto, a função custo e o conjunto de restrições, a solução ótima está representada pelo conjunto de variáveis que minimiza o custo dentre todas aquelas que verificam as restrições. Devido a escassez de recursos e ao aumento da

competição tecnológica, o processo de otimização aparece naturalmente em diversas aplicações relevantes, tais como nas indústrias aeronáutica, nuclear e naval, bem como energia, petróleo e gás. Em particular, objetiva-se utilizar abordagens baseadas em meta-heurísticas e derivadas topológicas na resolução de uma vasta classe de problemas inversos e de otimização.

No que tange o escopo da abordagem via meta-heurísticas, existe a necessidade de maior realismo na modelagem computacional e a abordagem de problemas cada vez mais complexos nas mais variadas áreas da ciência e tecnologia traz severos desafios à área de otimização (mono- e multi-objetivo) e de identificação. Uma característica dos problemas de otimização e identificação é a necessidade de se realizar numerosas simulações computacionais com modelos complexos (via método dos elementos finitos, por exemplo), com alta demanda de recursos computacionais, especialmente quando se deseja explorar a generalidade das meta-heurísticas consideradas, que podem atuar sobre variáveis contínuas e/ou discretas e sobre gramáticas capazes de gerar estruturas mais complexas, como programas, expressões simbólicas, projetos estruturais e outras.

O conceito de derivada topológica tem sido aplicado com sucesso na resolução de certas classes de problemas inversos e de otimização. A derivada topológica é definida como o primeiro termo da expansão assintótica de um dado funcional de forma com relação ao parâmetro associado ao tamanho de uma perturbação infinitesimal singular - tais como furos, inclusões, termos fonte e até mesmo trincas - introduzida em um ponto arbitrário de um dado domínio geométrico. Ao longo da última década a análise de sensibilidade topológica tornou-se uma ampla área de pesquisa tanto do ponto de vista teórico quanto numérico, possuindo aplicações em diversas áreas da modelagem computacional, quais sejam: otimização topológica, problemas inversos, processamento de imagens, síntese de novos materiais e modelagem mecânica de fenômenos dissipativos incluindo dano e fratura. Objetiva-se, portanto, dar continuidade as atividades de pesquisa em análise de sensibilidade topológica e suas aplicações em otimização topológica e problemas inversos. E dessa forma consolidar a Análise de Sensibilidade Topológica como linha de pesquisa no LNCC, permitindo o desenvolvimento sistemático de novos modelos, métodos e ferramentas matemáticas, incluindo novos algoritmos de otimização e problemas inversos.

Em particular, diversas classes de problemas inversos são escritos na forma de equações diferenciais parciais sobre determinadas. Essa dificuldade pode ser contornada reescrevendo o problema inverso na forma de um problema de otimização topológica. A ideia básica consiste em minimizar um funcional que mede a diferença entre o dado lido e o calculado numericamente em relação aos parâmetros de interesse. O conceito de

derivada topológica é então utilizado. Em particular, o funcional objetivo é expandido e em seguida truncado até o termo de segunda ordem, o que resulta em uma forma quadrática e estritamente convexa em termos dos parâmetros de interesse. Finalmente, através de um processo trivial de otimização obtém-se um algoritmo de reconstrução de segunda ordem não-iterativo e independente de qualquer chute inicial. Como resultado, o processo de reconstrução torna-se extremamente robusto a presença de ruído. A metodologia ora descrita será aplicada a diversas classes de problemas inversos, quais sejam: tomografia por impedância elétrica, reconstrução de fontes poluidoras pontuais, reconstrução de anomalias utilizando ondas sonoras, projeto ótimo de antena em hipertermia e reconstrução de obstáculos imersos em um fluido.

ANÁLISE NUMÉRICA E ADAPTATIVIDADE

A resolução numérica de modelos oriundos das ciências e engenharias induz a inserção de incertezas e erros. Análise numérica é a área da Matemática Aplicada dedicada ao estudo da existência e unicidade de solução para métodos numéricos e do desenvolvimento de estimativas de erro entre a solução aproximada e a exata. Desta forma, a linha de pesquisa em análise numérica objetiva desenvolver conceitos e ferramentas matemáticas para inferir a qualidade e a confiabilidade dos métodos numéricos. Como consequência, desenvolve-se o conceito de adaptatividade, o qual consiste em uma técnica numérica que emprega as estimativas de erro *a posteriori* (oriundas da análise numérica) no direcionamento da melhoria da qualidade da aproximação com redução dos custos computacionais via (des)refinamento de espaços de aproximação no tempo e no espaço.

Apesar de consolidado no LNCC, esse tema possui novos e grandes desafios em virtude das novas gerações de métodos numéricos híbridos, estabilizados e multiescalas atualmente em desenvolvimento no LNCC. De fato, essas inovadoras metodologias numéricas induzem novos algoritmos computacionais. Estes demandam a introdução de novas técnicas de análise matemática para a certificação da qualidade da solução numérica, e para a obtenção de novos estimadores de erro *a posteriori* que dirijam o processo adaptativos de melhoria das soluções numéricas.

2.1.3 SISTEMAS, CONTROLE E SINAIS

A área de pesquisa de Sistemas e Controle tem como objetivo o estudo do comportamento de sistemas dinâmicos e o desenvolvimento de estratégias visando a que se obtenha um desempenho ou propósito desejados. Sistemas e Controle é uma área de conhecimento bastante abrangente e intrinsecamente interdisciplinar, tendo aplicações nas mais variadas áreas, tais comobiologia, ecologia, economia, sistemas de energia elétrica e robótica. A

teoria de controle tem tido um papel fundamental no desenvolvimento de sistemas de tecnologia avançados, tendo sido instrumental na automação industrial: aeronáutica, automotiva e eletrônica, e em projetos espaciais. Inúmeros produtos e equipamentos que usamos no dia-a-dia dependem de sistemas de controle para a sua qualidade e operação adequada.

O processamento de sinais e de imagens visa a investigar representações e propriedades de sinais/imagens com objetivos diversos: armazenagem eficiente; visualização de informações, realização de modificações controladas, análise, síntese, detecção e classificação de sinais/imagens, dentre outras tarefas que são demandadas por diversas aplicações tecnológicas modernas, desde comunicações móveis à bioinformática, passando por sistemas de informação, e medicina assistida por computação.

As pesquisas em Teoria de Sistemas, Controle e Sinais no LNCC são reconhecidas como referência nacional e internacional, sendo tal liderança evidenciada, por exemplo, através de publicações em periódicos internacionais de prestígio, citações de seus trabalhos científicos na *Web of Science* e honrarias de prestígio de sociedades científicas nacionais e internacionais de renome. São as seguintes as áreas de pesquisa consolidadas no LNCC em Sistemas, Controle e Sinais:

- Modelagem, Controle e Filtragem de Sistemas Estocásticos;
- Controle e Filtragem de Sistemas Determinísticos Sujeitos a Incertezas;
- Processamento Digital de Sinais e Imagens;
- Controle e Análise de Equações Diferenciais Parciais;

Em contrapartida, com base no conhecimento consolidado, atividades de pesquisas estão sendo também direcionadas objetivando a consolidação da nova área de pesquisa de:

- Modelagem Estocástica em Finanças.

Além disso, pretende-se incorporar às atividades de Sistemas, Controle e Sinais do LNCC a área estratégica de Robótica. A incorporação desta nova área de pesquisa é, entretanto, condicionada à contratação de pelo menos um pesquisador especialista em Robótica.

Objetivos Específicos:

- Expandir a liderança científica e explorar aplicações nas áreas de pesquisa consolidadas
- Consolidar a área de pesquisa de Modelagem Estocástica em Finanças.

Metas:

- Na área de Modelagem, Controle e Filtragem de Sistemas Estocásticos, expansão das pesquisas em:
 - Análise de tráfego pesado em redes;
 - Dinâmica de populações via processos de *Fleming-Viot*, incluindo aplicações no âmbito da genética populacional;
 - Análise de estabilidade, controle e filtragem de sistemas dinâmicos com chaveamentos Markovianos e incertezas de modelagem.
- Na área de Controle e Filtragem de Sistemas Determinísticos Sujeitos a Incertezas:
 - Desenvolvimento de técnicas avançadas de controle e estimação de sinais para sistemas dinâmicos sujeitos a classes de incertezas paramétricas e não-paramétricas de modelagem, incluindo sistemas com não linearidades e redes de sensores.
- Na área de Processamento Digital de Sinais e Imagens:
 - Expansão das pesquisas em processamento digital de sinais, explorando aplicações, incluindo:
 - Análise tempo-frequência para caracterização de distúrbios da rede elétrica;
 - Análise tempo-frequência para calibração de modelos computacionais de síntese de sons de impacto e de instrumentos musicais de membrana.
 - Desenvolvimento de pesquisas em processamento avançado de imagens, incluindo visualização, reconstrução tridimensional, segmentação e classificação de estruturas de relevância médica, bem como análise de imagens de faces e de texturas e suas aplicações.
- Na área de Controle e Análise de Equações Diferenciais Parciais:
 - Desenvolvimento de modelos de oscilações de grandes estruturas e estudo dos mecanismos utilizados para sua estabilização usando técnicas de Equações Diferenciais Parciais assim como de controle ótimo. Serão também desenvolvidos modelos de poluição ambiental com o objetivo de minimizar as consequências dos agentes poluidores no ar e em meios hídricos.
- Na área de Modelagem Estocástica em Finanças:
 - Investigação de modelos estocásticos de precificação de derivativos no mercado de renda fixa e de risco de crédito.
- Realização de cursos/seminários nas áreas de Processamento Digital de Sinais e Modelagem Estocástica em Finanças.

2.1.4 CIÊNCIA DE DADOS

O tratamento do imenso volume de dados que estão sendo produzidos pelas diversas áreas da ciência e por bilhões de usuários de serviços de Internet globais se apresenta como um dos grandes desafios para a sociedade do conhecimento. Apresentado de forma geral como um vetor de múltiplas facetas, o fenômeno ainda está sendo interpretado pelos cientistas e tem impulsionado iniciativas em diversas áreas. Nas ciências, o dilúvio aparece como a expressão de uma nova maneira de investigação, incitando biólogos, astrônomos, bioquímicos, e demais pesquisadores em diversas áreas científicas, a enfrentar problemas computacionais na chamada e-ciência, que se tornam barreiras para as suas descobertas. Na indústria, a interpretação de dados em grande volume se vale de técnicas estatísticas e de aprendizado no apoio à tomada de decisão a partir de várias frentes, tais como: criação de modelos preditivos, desenvolvimento de recomendações automáticas, identificação automática de perfis e padrões, apenas para citar algumas áreas de aplicação. No setor governamental, há oportunidades de se debruçar sobre imensas bases de dados do setor público com vistas a tornar o planejamento mais eficiente bem como novos serviços que possam melhorar o atendimento ao cidadão. Novas profissões especializadas no trato e, principalmente, na análise e interpretação de grandes volumes de dados, surgiram trazendo o método científico para o setor empresarial.

Do ponto de vista técnico-científico, constitui-se um desafio o estudo metódico para a extração generalizada e em escala de conhecimento relevante a partir de uma imensa massa de dados, em geral dinâmica. A abordagem a esse desafio com aplicações em diversas áreas no eixo ciência-indústria-governo emerge como uma nova espécie de ciência. A chamada Ciência de Dados incorpora elementos variados e se baseia em técnicas e teorias oriundas de vários campos em engenharia e ciências básicas, sendo assim intimamente ligada com muitas das disciplinas tradicionais bem estabelecidas, porém viabilizando uma nova área altamente interdisciplinar. Essa nova área se baseia em dados tendo por objetivo identificar os princípios, métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados heterogêneos, cuja incerteza também deve ser levada em conta e que possa também se apresentar com grandes demandas de tratamento em tempo limitado.

Em uma visão geral, pode-se compreender a Ciência de Dados como um conjunto de ações aplicadas a uma coleção de dados que conduz à descoberta de conhecimento (*i.e.*, tendências, relações e padrões subjacentes a esses dados). As ações vão desde a seleção dos dados até a extração de conhecimento e se organizam em quatro macro atividades: seleção e integração de dados, pré-processamento de dados, produção do modelo para extração do conhecimento e avaliação.

O processo pode ser compreendido como um caso particular de experimentação científica *in-silico* no qual os dados são volumosos, as estruturas de dados precisam ser bem definidas e os métodos de seleção, pré-processamento e de produção do modelo são computacionalmente intensivos.

Além do desenvolvimento de pesquisa científica e tecnológica na área de processamento e análise de grandes volumes de dados, descritos a seguir, o LNCC contribui e se fortalece com a formação de mestres e doutores em pesquisas interdisciplinares em modelagem computacional, fornecendo a base em recursos humanos para enfrentar os desafios impostos por esta nova área. Nesse contexto, o LNCC se posiciona com atividades de pesquisa, desenvolvimento e inovação em questões estratégicas de Ciência de Dados organizadas em torno das três áreas que serão apresentadas a seguir:

- Gerência e processamento de grandes volumes de dados;
- Modelagem e análise baseada em dados;
- Métodos para segurança de informação.

Os Objetivos Específicos e as Metas se aplicam às 3 áreas.

Objetivos Específicos:

- Realizar pesquisa básica em aspectos fundamentais de análise de dados em larga escala;
- Aplicar as metodologias desenvolvidas para análise de dados em larga escala em áreas interdisciplinares (como Saúde, segurança de informação, e-ciência, redes de comunicação);
- Consolidar as atividades de ciência de dados no LNCC com o objetivo de ser referência na área

Metas:

- Fortalecimento da integração das diferentes iniciativas ligadas a ciências de dados no LNCC
- Estruturação da formação de recursos humanos no nível de pós-graduação na área de ciências de dados

GERÊNCIA E PROCESSAMENTO DE GRANDES VOLUMES DE DADOS

Nesta seção, descrevem-se as atividades em sistemas de gerência de dados, sistemas de workflows científicos e de fluxos de dados, bem como sistemas multimídia.

Gerência de Dados

A gerência e análise de grandes volumes de dados vêm sendo tratadas pela comunidade de bancos de dados há muitos anos. No entanto, a disponibilização de instrumentos cada vez mais potentes, como telescópios na astronomia, sequenciadores de DNA na genômica e espectrômetros de massa, na química e bioquímica, apenas para mencionar alguns exemplos, modificou a escala em que se consideravam os grandes volumes. Ao fazerem parte da ciência, instrumentos dessa natureza passaram a participar do ciclo experimental, produzindo dados a cada etapa de validação de hipóteses científicas. Além de instrumentos de coleta de dados, sensores diversos monitoram o estado de variáveis de interesse, como volume pluviométrico, no contexto meteorológico, frequência cardíaca, no contexto da saúde e esporte de alto rendimento, e temperatura ambiental, no contexto de casas inteligentes. Em área como a pesquisa em Geofísica, o subsolo oceânico é varrido por hidrofones (*hydrophones*) que capturam ondas sonoras refletidas nas camadas tectônicas, por exemplo na área do pré-sal da costa brasileira. Em todos esses cenários, a grande quantidade de dados produzida precisa passar por um processo de limpeza e transformação para que o dado cru seja integrado com outras informações produzindo uma base de dados de alto valor para tomada de decisão na ciência ou na indústria.

Nesse contexto, o tema de pesquisa e desenvolvimento tecnológico no LNCC se concentra na gerência de dados multidimensionais, consideram-se as observações de fenômenos da natureza; captura de dados de sensores físicos; e dados produzidos em simulações computacionais. Em todos esses cenários, dados são registrados nas dimensões espaciais, temporais e demais relativas ao domínio do problema. Além das dimensões, variáveis de interesse têm valores associados a cada ponto do espaço multidimensional. Oferecer suporte ao armazenamento, indexação e processamento destes dados em grandes volumes, envolve técnicas de particionamento eficiente para distribuição de dados em sintonia com algoritmos eficiente para o tratamento de análise de fragmentos do espaço multidimensional. Mais recentemente, como parte do projeto em parceria com a comunidade europeia (HPC4E), pretendemos investigar o apoio a processos de visualização e análise de erro em tempo real de dados oriundos de simulação computacional em grande escala.

Workflows Científicos e de Fluxo de Dados

Experimentos científicos computacionais atuais são frequentemente dados pela composição, através do consumo e produção de conjuntos de dados, de várias tarefas computacionais e são frequentemente denominados de *workflows* científicos. Existem diversos desafios envolvidos no gerenciamento de *workflows* científicos em larga escala, tais como a identificação de oportunidades de execução paralela das atividades computacionais e o registro de dados de entrada, parâmetros e aplicações utilizadas para facilitar a rastreabilidade e reprodutibilidade.

A gerência manual de experimentos científicos computacionais em larga escala é trabalhosa e sujeita a erros pois diversos aspectos tem que ser levados em consideração, por exemplo: as atividades podem ter dependências de dados entre si, de modo que algumas atividades só podem iniciar a sua execução quando seus dados de entrada, produzidos por outras atividades que as precedem, estiverem disponíveis; eventualmente o fluxo de execução do experimento pode sofrer uma bifurcação para a execução de diversas atividades independentes entre si, tornando interessante a execução paralela das mesmas por questões de eficiência; o início da execução de uma atividade pode depender do fim da execução de diversas atividades disparadas após uma bifurcação, requerendo uma sincronização da execução do experimento, onde é necessário garantir que todas as atividades geradas pela bifurcação de fato terminaram a sua execução; caso o experimento utilize diversos recursos computacionais remotos, é necessário gerenciar a transferência de dados e monitorar a execução de atividades remotas. Esses aspectos são difíceis de controlar quando são utilizadas linguagens de *scripting* genéricas como Perl ou Python, populares em áreas como a bioinformática. A automatização destes experimentos pode ser apoiada por sistemas de gerência de *workflows* científicos, que permitem projetar, executar e analisar experimentos dados pela composição de diversas aplicações científicas, possivelmente distintas, que são encadeadas através da produção e consumo de dados.

Informações de proveniência, que reúnem detalhes sobre a concepção e a execução desses experimentos, descrevendo os processos e dados envolvidos na geração dos resultados dos mesmos, podem ser utilizadas para facilitar esta tarefa. Elas permitem a descrição precisa de como o experimento foi projetado, chamada de proveniência prospectiva, e do que ocorreu durante sua execução, denominada proveniência retrospectiva. Algumas aplicações da proveniência incluem a reprodução de um experimento para fins de validação, compartilhamento e reutilização de conhecimento, verificação de qualidade de dados e atribuição de resultados científicos. Um dos conceitos comumente capturados na proveniência é o de causalidade, que é dado pelas relações de dependência existentes entre

processos e conjuntos de dados. Estas dependências podem derivar, por transitividade, dependências entre conjuntos de dados e entre processos. O padrão PROV, para modelagem e representação de dados de proveniência, foi proposto recentemente pelo W3C (World Wide Web Consortium).

No LNCC, há projetos de colaboração com outras instituições (UFRJ, UFF e Universidade de Chicago) para o projeto e implementação de ferramentas para gerenciamento de *workflows* científicos, como o SciCumulus, para ambientes de nuvem, e o Swift, para *dataflows* massivamente paralelos. Tais ferramentas têm sido utilizadas na instituição em áreas como a bioinformática e modelagem molecular.

Sistemas Multimídia

Sistemas Multimídia são tradicionalmente uma fonte de grande quantidade de dados, bem como de uso intenso do sistema de distribuição de informação. Estudos recentes propõem o envio, em tempo real, de vídeo 3D com posicionamento livre do telespectador. Neste tipo de aplicação o telespectador pode mover-se livremente dentro da cena, observando-a de um ângulo específico por ele(a) escolhido. Tal aplicação permite, por exemplo, acompanhar um jogador específico em um evento esportivo, mesmo quando este jogador está ocluso pelo ângulo de visão padrão da cena. Este tipo de aplicação exige o envio de vários ângulos de visão tridimensional da cena, bem como o processamento em tempo real de um novo ângulo de visão interpolado daqueles existentes.

Sistemas Multimídia modernos precisam também ser ubíquos, i.e., capazes de ser explorados através de diversas plataformas, tais como celulares, *tablets*, *desktops*, etc. Tal característica também exige que customizações sejam realizadas em tempo real dependendo do tipo de cliente conectado.

Sistemas Multimídia tradicionais, com multiplexação de mídias básicas, a saber vídeo (inclusive 3D), áudio (inclusive 3D), texto, imagem (incluindo imagens 3D, bem como aquelas resultantes de câmeras plenópticas), objetos 3D, informações para representação de sentido tátil, olfativo ou mesmo gustativo, também são objeto de interesse, especialmente quando algumas das mídias menos tradicionais tem ainda grande potencial de exploração como tema de pesquisa.

MODELAGEM E ANÁLISE BASEADA EM DADOS

No núcleo de Ciência de Dados está a modelagem e análise baseada em dados. As atividades do LNCC nesse contexto se agrupam em (i) análise de grandes volumes de dados; (ii) mineração de dados e aprendizado de máquina; e (iii) ciência de redes, envolvendo a análise de redes complexas.

Análise de Grandes Volumes de Dados

A análise de dados diz respeito a um conjunto de técnicas, algoritmos e modelos que permitam ou facilitem a interpretação dos dados sobre o ponto de vista do investigador. Tem-se como objetivo representar os dados segundo modelo que ofereça expressividade semântica necessária para sua interpretação e cujo processamento possa se dar de forma eficiente. Neste sentido, retrata uma evolução da área de banco de dados para domínios mais complexos e com requisitos de dados refinados.

Neste sentido, a representação de dados pode se valer de modelos multidimensionais, multirepresentação, ou em grafos, conforme as características de dados a serem realçadas. Aspectos associados à representação que favoreçam estruturas de dados eficientes devem também ser considerados.

Do ponto de vista da tomada de decisão, a interpretação de dados em grandes volumes objetiva identificar novas informações que se tornam perceptíveis sobre a luz da integração de dados e exemplificação extensa da ocorrência de fenômenos nos dados. Neste contexto, avaliamos correlações e composições que possam revelar conhecimento novo e expliquem os fenômenos retratados. A identificação de correlações deve considerar as variações temporais e espaciais e tentar através de métodos probabilísticos estabelecer relações explicativas sobre eventos do mundo investigado. Esse tipo de análise envolve combinações exponenciais no número de dados envolvidos o que exige algoritmos eficientes para seu cálculo.

Uma característica que surge ao inferirmos sobre grandes volumes de dados é a qualidade da informação que substancia a tomada de decisão. Neste contexto, quantificar a incerteza associada às interpretações permite comparar alternativas. Em simulações numéricas, vários fatores de incerteza incidem sobre as predições e precisam ser representadas e interpretadas durante a análise de dados. Neste contexto, estamos interessados em oferecer uma plataforma para a publicação de dados de simulação numérica qualificados. Em uma plataforma dessa natureza, tomadores de decisão baseadas em dados de predição teriam resultados de diferentes modelos qualificados segundo métricas estabelecidas. Desenvolvemos técnicas para inferência de incerteza em predições produzidas por modelos de simulação e utilizamos sistemas de bancos de dados probabilísticos para o armazenamento e análise de dados de predição. Dessa forma é possível oferecer um ambiente para tratamento integrado de dados de predições pretendemos investigar cenários nos quais os modelos matemáticos não estejam disponíveis, sendo estes retratados nos dados de predição.

Finalmente, a análise de grandes volumes de dados precisa de estratégias eficientes e tolerantes à falhas. Consideramos estratégias baseadas em sistemas com amplo uso de memória e sistemas de arquivos distribuídos. O modelo baseado no princípio *Map/Reduce*, tem se mostrado um paradigma eficiente no qual os dados são particionados por nós do cluster e funções processam dados de forma paralela e independente. Estamos interessados em explorar vários aspectos associados a este modelo. Inicialmente, acreditamos que sistemas deste tipo devem suportar linguagens algébricas mais expressivas, como oferecido pelo sistema QEF desenvolvido pelo laboratório DEXL. Adicionalmente, acreditamos que para obtermos desempenho superior é preciso que haja uma relação próxima e eficiente entre a estratégia de execução dos *dataflows* e o esquema de particionamento de dados. Neste tema, pretendemos desenvolver técnicas e algoritmos que forneçam aos sistemas de execução de workflows a mesma qualidade e eficiência encontrada em sistemas de bancos de dados.

Mineração de Dados e Aprendizagem de Máquina

Uma das etapas fundamentais no processo de produção de conhecimento na Ciência de Dados é a análise dos dados e a construção de modelos que descrevam o fenômeno subjacente a estes. As características particulares dessa etapa do processo variam bastante, influenciadas não apenas pelos dados propriamente ditos como também pelo tipo de modelo que se pretende construir e pelas ferramentas técnicas usadas para tal. Não surpreende, portanto, que a fase de análise dos dados esteja associada a diferentes termos, como “aprendizagem de máquina”, “aprendizagem estatística”, “reconhecimento de padrões” e “mineração de dados”, entre outros. Embora cada um desses termos reflita peculiaridades das diferentes instâncias do processo de análise, todas essas áreas compartilham um objetivo comum: a construção de modelos que sintetizem algum aspecto relevante do processo gerador dos dados e que permitam alguma forma de generalização; ou seja, conclusões que extrapolem o escopo restrito dos dados. Tendo em vista esse denominador comum, e considerando a atuação do LNCC na área de modelagem matemática e computacional, pode-se referenciar genericamente a etapa de análise de dados como *modelagem baseada em dados*.

O objetivo da modelagem baseada em dados é extrair padrões inerentes a um determinado fenômeno usando observações coletadas a respeito do objeto de estudo. Por “observações” entenda-se, aqui, dados extraídos diretamente do fenômeno em questão e que dependam minimamente da interpretação do observador. Isso significa que, embora conhecimento prévio a respeito do fenômeno possa ser incorporado ao modelo, ele não é estritamente necessário. Como as técnicas da área não requerem uma descrição do processo físico subjacente aos dados, elas podem ser aplicadas mesmo que os detalhes do fenômeno de

interesse não sejam completamente compreendidos ou não possam ser facilmente simulados. Nesse sentido, a área contrasta e complementa a modelagem baseada em princípios fundamentais (ou “primeiros princípios”), cuja construção de modelos se apoia fortemente em um entendimento do fenômeno de interesse.

Nos últimos anos, a combinação das técnicas de modelagem baseada em dados com a enorme quantidade de dados e recursos computacionais atualmente disponíveis reconfigurou o cenário acadêmico e industrial. Em função da semelhança com o processo de aquisição de conhecimento, a construção de modelos baseada em dados é frequentemente chamada de aprendizagem de máquina ou aprendizagem estatística. A aprendizagem de máquina divide-se em três subáreas: a aprendizagem supervisionada, a aprendizagem não supervisionada e a aprendizagem por reforço.

Na aprendizagem supervisionada o objetivo é prever o valor de uma quantidade de interesse claramente identificada como a variável dependente do problema. Exemplos proeminentes de aplicações desse tipo são o diagnóstico médico, a detecção de objetos em imagens, a filtragem de e-mails, a classificação taxonômica de fragmentos de DNA e a predição da conformação espacial de complexos moleculares, entre muitas outras;

Na aprendizagem não supervisionada não há o conceito de variável dependente: nesse caso o objetivo é criar um modelo do processo que gerou os dados. Esse tipo de modelagem pode ser usado isoladamente ou como uma etapa preliminar em um problema de aprendizagem supervisionada. A aprendizagem não supervisionada pode ser usada em inúmeros problemas, como por exemplo: detecção de anomalias, separação de sinais sonoros e agrupamentos de objetos em categorias;

A aprendizagem por reforço trata de problemas de tomada de decisão sequencial envolvendo incerteza. A característica principal dos problemas desse tipo é o fato de as decisões envolvidas terem um efeito cumulativo, ou seja, as consequências de uma determinada decisão podem se estender por um intervalo indefinido de tempo. Esse arcabouço teórico é genérico o suficiente para englobar problemas provenientes de diversas áreas, como controle, alocação ótima de recursos e planejamento sequencial. A abordagem clássica para solucionar problemas de tomada de decisão sequencial é a programação dinâmica, uma disciplina madura, bem fundamentada teoricamente e exaustivamente testada na prática. Embora tenha sido desenvolvida de maneira independente da programação dinâmica, a

aprendizagem por reforço é atualmente reconhecida como uma abordagem estatística para esta última, em que amostras de transições e custos substituem um modelo explícito do problema de tomada de decisão sequencial. Como a aprendizagem por reforço é uma instância da modelagem baseada em dados – pressupondo apenas a possibilidade de extração de dados do problema, sem a necessidade de um modelo analítico –, ela aumenta consideravelmente o leque de problemas de tomada de decisão sequencial que podem ser abordados.

Ciência de Redes

Sistemas em rede encontrados em muitas áreas do conhecimento apresentam estruturas complexas que demandam avanços científicos e tecnológicos significativos para a melhor compreensão de sua estrutura e funcionalidade. Recentemente, tem-se adotado o termo Ciência de Redes para se referir ao estudo desses sistemas complexos em rede, tanto de sua estrutura quanto de sua funcionalidade, de forma ampla em um campo de atuação fortemente interdisciplinar.

O foco da Ciência de Redes é a caracterização, análise e modelagem de redes complexas, contribuindo para a área mais abrangente de ciência de dados. Uma rede, em uma definição ampla, é um conjunto de objetos interconectados. Esses objetos são denominados diferentemente segundo a área específica de estudo (vértices, atores, ou nós são exemplos usuais dessas denominações). As conexões entre esses objetos, também chamadas de arestas, representam a relação entre esses objetos. Sistemas em rede (também chamados de grafos na literatura matemática) estão presentes em muitas áreas do conhecimento. Apenas para nos mantermos no domínio de redes técnico-sociais, a estrutura de roteamento da Internet, redes sociais *online*, redes P2P, Internet das coisas (IoT) e redes de telefonia móvel representam alguns exemplos de redes complexas que impõem importantes desafios para a sua caracterização, análise e modelagem devido à sua larga-escala e o potencial dinamismo dessas redes. Em uma perspectiva mais ampla e multidisciplinar, exemplos similares são encontrados em biologia, física, sociologia e em diversas outras áreas que lidam com redes complexas de larga-escala na esfera de ciência-indústria-governo.

Em consonância com essa tendência, os objetivos das atividades de pesquisa, desenvolvimento e inovação envolvem o estudo de teoria e aplicações de ciência de redes em diferentes contextos. Um primeiro objetivo é a modelagem de redes complexas de alta ordem, por exemplo, multicamadas e variantes no tempo. Um segundo objetivo visa o estudo dos aspectos para melhoria do desempenho computacional no processamento e análise de redes complexas de larga-escala. Um terceiro objetivo é focar na

caracterização, análise e modelagem de redes na área de tecnologias de informação e comunicação (TICs), incluindo Internet, redes de comunicação móvel ou redes sociais *online*, abrangendo, portanto, redes de infraestrutura (acesso e roteamento), aplicações (p.ex., distribuição de conteúdo ou comunicação máquina a máquina) e usuários (tais como redes sociais *online*); ou mesmo a interdependência entre essas redes. Um quarto objetivo abrange a aplicação de conceitos de ciência de redes em contextos interdisciplinares com outras áreas do conhecimento além da área específica de TICs, sendo exemplos a aplicação de ciência de redes em áreas como saúde, e-ciência ou cidades inteligentes.

MÉTODOS PARA SEGURANÇA DE INFORMAÇÃO: CRIPTOGRAFIA

A terceira área de Ciência de Dados refere-se a métodos para segurança de informação, envolvendo preservação de privacidade, criptografia e segurança de protocolos de comunicação.

Privacidade

Tecnologias da informação e comunicação vêm proporcionando coletas e armazenamentos de grande quantidade de dados com um custo cada vez menor. Além disto, algoritmos de mineração de dados processam cada vez mais dados e permitem maior análise e extração de informações. Para protegê-las, os sistemas computacionais usam diversos algoritmos e protocolos, por exemplo, RSA e HTTPS. No entanto, o uso legítimo de tais dados permite os proprietários determinarem fatos relevantes da vida privada dos usuários de tais sistemas. Com efeito, algoritmos probabilísticos podem determinar a expectativa de vida de usuários. A proteção da privacidade é um problema cada vez maior. Certamente, a privacidade deve ser preservada, mas ainda não está claro quais dados e como eles devem ser preservados. Lembramos que dados anonimizados podem ser relacionados com outros dados anonimizados e o resultado pode ser uma desanonimização.

Com o avanço tecnológico, os sistemas computacionais se tornam cada vez mais ubíquos ao ponto que os usuários são praticamente obrigados a usá-los. Por exemplo, um usuário pode optar de não usar Internet Banking, mas normalmente, não pode optar de viver sem eletricidade. As redes elétricas são controladas por Sistemas de Supervisão e Aquisição de Dados (SCADA – Supervisory Control and Data Acquisition). Como se não bastasse, as redes elétricas estão se tornando inteligentes (*Smart Grids*). Neste cenário, um funcionário poderia determinar quando um usuário – neste caso consumidor – específico está em casa ou não, se está só ou acompanhado, que horas vai dormir e acordar, entre muitos outros dados privados. Algoritmos e protocolos que preservam a privacidade devem prover

garantias matemáticas que um funcionário não será capaz de invadir a privacidade dos usuários que normalmente são os clientes. Lembremo-nos do caso Snowden da Agência Nacional de Segurança dos Estados Unidos (NSA - National Security Agency). Somente com garantias baseadas nas limitações computacionais, as empresas protegem-se de escândalos e os usuários se sentirão seguros em usar os serviços.

Pesquisas em preservação da privacidade são naturalmente interdisciplinares e abrangem ciências humanas, biológicas e exatas. O nosso foco está em prover algoritmos eficientes que preservem a privacidade, por exemplo, que resultem em informações sobre uma população e mantenham os dados privados protegidos. Os algoritmos devem prover a informação necessária ao desenvolvimento tecnológico e preservar a privacidade.

Criptografia

A criptografia é o estudo de técnicas para comunicação segura na presença de terceiros, chamados adversários. A criptografia visa a construção e análise de protocolos que impedem a terceiros ou ao público a leitura de mensagens privadas. Vários aspectos de segurança da informação, tais como a confidencialidade dos dados, integridade de dados, autenticação e não-repúdio são fundamentais para a criptografia moderna, que nasceu do cruzamento das disciplinas de matemática, ciência da computação e engenharia elétrica. Aplicações de criptografia incluem cartões ATM, senhas de computador e comércio eletrônico. A criptografia moderna é fortemente baseada na matemática e na ciência da computação. Os algoritmos de criptografia são desenvolvidos usando a dificuldade computacional na resolução de certos problemas matemáticos. É teoricamente possível quebrar um tal sistema, mas é inviável de fazê-lo por quaisquer meios práticos conhecidos. Estes algoritmos criptográficos são, portanto, denominado computacionalmente seguro. Avanços teóricos, por exemplo, na melhoria dos algoritmos de fatoração de números inteiros e no avanço da velocidade dos computadores exigem que estas soluções sejam continuamente adaptadas. Existem protocolos teoricamente seguros e que, comprovadamente, não podem ser quebrados, mesmo com um poder de computação ilimitado, porém estes protocolos são mais difíceis de serem implementados.

Há dois tipos de chaves criptográficas: chaves simétricas (criptografia de chave única) e chaves assimétricas (criptografia de chave pública). A criptografia pós-quântica refere-se aos algoritmos criptográficos (geralmente algoritmos de chave pública) que são desenvolvidos para serem seguros contra um ataque de um computador quântico. Isso não é válido para os algoritmos de chave pública mais populares que podem ser eficientemente quebrados por um computador quântico. O problema com os algoritmos atuais é que a sua segurança depende de um dos três problemas matemáticos: o problema de fatoração de inteiros, o problema do logaritmo discreto ou problema logaritmo discreto

em curvas elípticas. Todos estes problemas podem ser eficientemente resolvidos em computador quântico através do algoritmo quântico de Shor.

Os principais protocolos da criptografia pós-quântica são: (i) Criptografia baseada em sistemas de equações quadráticas multivariáveis; (ii) Criptografia baseada em funções Hash, como assinaturas de Lamport e o esquema de assinatura Merkle; (iii) Criptografia baseada em código corretores de erros, tais como, a criptografia McEliece e Niederreiter; e (iv) Criptografia baseada em reticulados.

Os objetivos das atividades de pesquisa envolvem o desenvolvimento de protocolos de criptografia pós-quântica e análise de segurança destes sistemas. Desenvolvemos implementações computacionais usando processamento de alto-desempenho principalmente para testar a segurança dos protocolos pós-quânticos.

Segurança de Protocolos de Comunicação

O grupo de redes da Coordenação de Sistemas e Redes - CSR/LNCC e a equipe do Laboratório COMPASSO da COPPE Sistemas/UFRJ mantêm uma estreita parceria no desenvolvimento do projeto de ferramentas de monitoramento para aumentar a segurança da informação em LANs. O foco do projeto é o protocolo de comunicação Radnet-S, desenvolvido e já patentado nos EUA pela UFRJ. Em comparação com protocolos convencionais, Radnet-S usa um padrão de mensagem que omite os endereços dos dispositivos de origem e destino, permitindo transmitir os pacotes de forma indecifrável além do seu conteúdo criptografado. Radnet-S opera em uma rede estabelecida sem alterá-la, como um serviço *plug&play* na camada de rede. Um importante resultado é permitir proteger o servidor de Log de uma rede local de um eventual intruso que tente destruí-lo. Neste caso, o Radnet-S impedirá a ação do intruso e o servidor de Log permanecerá inalterado. Radnet-S implementa um novo protocolo de comunicação centrado em interesse, onde o interesse é um metadado, p.ex., uma palavra-chave, um nome de aplicação ou de um serviço, ou seja, pode ser qualquer conjunto de caracteres, os quais podem ser também criptografos.

Um resultado preliminar observado nas avaliações experimentais do Radnet-S, foi sua interoperabilidade, permitindo operar simultaneamente em segmentos de redes com protocolos IPV4 e IPV6 e em combinação dos dois protocolos em redes que estejam em transição. Observou-se também uma grande eficácia em subsistema de vigilância oculta para sistemas de Logs, entre outros.

Os testes realizados comprovaram o perfeito funcionamento do protocolo tanto em um único segmento de rede, como também entre segmentos distintos através de configurações específicas nos equipamentos de redes ou mesmo através da

Internet utilizado *gateways* ou de VPN's. Adicionalmente ao protocolo Radnet-S, o grupo da CSR desenvolve também um sistema de segurança denominado SGS (Sistema de Gerenciamento e Segurança), que consiste numa plataforma distribuída através de coletores e servidores repositórios para tratamento das informações coletadas, cujos os módulos utilizarão o protocolo Radnet-S para transportar informações sensíveis entre os coletores e os repositórios. Posteriormente, quando o SGS estiver integrado com o protocolo Radnet-S, irá monitorar além do ambiente computacional da sua sede, a infraestrutura do Ponto de Presença do LNCC-RJ, que se encontra localizada na cidade do Rio de Janeiro, na sua antiga sede.

Nessa instalação, o LNCC abriga, coordena ou gerencia importantes projetos de seu interesse, bem como projetos de outras organizações conveniadas com o LNCC, tais como o Ponto de Troca de Trafego PTT-RJ (ou do inglês IX *Internet Interchange Point*) do Núcleo de Informação e Coordenação do Ponto BR - NIC-BR, subordinado ao Comitê Gestor da Internet Brasileira (CGI-BR) o Ponto de Presença da RNP no estado do Rio de Janeiro (POP-RJ/RNP), a Infraestrutura computacional de Ensino à Distância do Exército Brasileiro, entre outros.

Objetivos específicos:

- Evoluir as tecnologias oriundas do RADNET-S no que tange a segurança de dados e disponibilizar para empresas nacionais.

Metas:

- Efetivar parceria com empresas para utilização comercial da tecnologia RADNET-S.

2.1.5 COMPUTAÇÃO DE ALTO DESEMPENHO

COMPUTAÇÃO MASSIVAMENTE PARALELA, DISTRIBUÍDA E QUÂNTICA-

O avanço dos ciberambientes de processamento de alto desempenho (PAD) tem modificado a forma de condução da investigação científica em diferentes áreas de pesquisa. Ao disponibilizar recursos com capacidade computacional da ordem de Petaflops e capacidade de armazenamento da ordem de dezenas a centenas de Petabytes, os ciberambientes atuais viabilizam a construção de aplicações de modelagem e simulação cada vez mais complexas, como as usadas nas áreas de bioinformática, medicina, meio ambiente, energia e governo.

Esse cenário induz o pensamento de que aumentar continuamente a capacidade computacional e de armazenamento dos ciberambientes é suficiente para as aplicações científicas tratarem, armazenarem e visualizarem massas de dados continuamente crescentes. Contudo, a habilidade dos usuários de empregar eficientemente os recursos computacionais dos ciberambientes atuais na construção dessas aplicações é pelo menos tão importante quanto o aumento da capacidade computacional e de armazenamento. A experiência de grandes centros de supercomputação em outros países (NCSA, LBNL, TACC e SDSC nos EUA, e BSC e CEA na UE, por exemplo) corrobora tal afirmação.

Nesse sentido, um desafio é o desenvolvimento de novos códigos e a otimização de códigos legados através da implantação de modelos de programação para plataformas de alto desempenho heterogêneas baseadas em arquiteturas de *processadores multicore*, *manycore* e *many-integrated core*.

Outro desafio nessa linha está na integração e virtualização de recursos (massivamente) paralelos e distribuídos, de recursos de armazenamento e tratamento de grandes massas de dados, de recursos de redes de sensores, e de recursos de visualização. Isso requer o desenvolvimento de *software* básico (sistemas operacionais, middleware) e desenvolvimentos de aplicações para computação em ambientes virtualizados massivos, seguindo modelo semelhante ao empregado em nuvens computacionais.

A despeito dos avanços dos ciberambientes atualmente disponíveis, a recente violação da Lei de Moore mostra que a computação baseada em *chips* de silício não admite mais melhorias significativas e que há necessidade de desenvolver novos paradigmas computacionais baseados em novas tecnologias. Em particular, a Computação Quântica tem se mostrado promissora em várias áreas de aplicação e recebido investimento sistemático no exterior. No Brasil, existem diversos grupos de pesquisa desenvolvendo protótipos de *hardware* quânticos baseados em sistemas físicos diversos, como Ótica quântica, Spintrônica, Ressonância magnética nuclear, entre outros, em diversas universidades e centros de pesquisas. Para funcionar adequadamente, o hardware quântico requer um tipo de *software* com leis de funcionamento radicalmente diferentes das do *software* clássico. Fica clara a necessidade prática do investimento também em *software* quântico.

Os exemplos acima ilustram como as áreas de computação massivamente paralela, distribuída e quântica estão interligadas entre si e servem de apoio às várias outras áreas de modelagem computacional, contribuindo para o ecossistema multidisciplinar do LNCC, seja no contexto das necessidades dos projetos em andamento no LNCC como no contexto da formação de recursos humanos.

Objetivos específicos:

- Desenvolver pesquisas e técnicas de computação de alto desempenho para problemas de computação científica.
- Desenvolver pesquisas e técnicas de computação distribuída e paralela para modelagem computacional.
- Desenvolver pesquisas e algoritmos em computação quântica.

VISUALIZAÇÃO CIENTÍFICA E AMBIENTES VIRTUAIS

A enorme quantidade de dados gerados diariamente pelos dispositivos de aquisição de imagens tem como consequência imediata a necessidade crescente de técnicas computacionais para armazenamento, transmissão, processamento, análise, extração e recuperação de informações relacionadas com os fenômenos e procedimentos de interesse. No caso particular da medicina e odontologia, imagens adquiridas em tomografias/ressonâncias podem ser processadas para que as estruturas de interesse sejam extraídas, ou segmentadas, e posteriormente visualizadas, auxiliando assim o diagnóstico de doenças e o planejamento de procedimentos terapêuticos, tais como cirurgias e radioterapias.

Por outro lado, técnicas e modelos em mecânica do contínuo vêm sendo aplicados com sucesso tanto em computação gráfica quanto em processamento de imagens. Por exemplo, equações de fluidos e modelos da teoria de elasticidade são utilizadas na geração de animações para a indústria cinematográfica e métodos em otimização topológica e análise funcional estão sendo aplicadas em processamento de imagens.

Neste contexto, pretende-se integrar métodos em processamento de imagens, computação gráfica e modelagem de meios contínuos para gerar temas multidisciplinares de pesquisa, com aplicações tecnológicas e científicas.

Objetivos específicos:

- Desenvolver pesquisas e técnicas de computação de alto desempenho para visualização de dados, animação de fluidos e de modelos de tecidos para aplicações gráficas.
- Aplicação de modelos massa-mola para visualização de tecidos flexíveis dentro de simuladores em ambientes de realidade virtual
- Desenvolver Visualização científica em ambientes de computação distribuída e paralela
- Desenvolver Visualização de imagens médicas tridimensionais

- Desenvolver técnicas de animação de fluidos via *Smoothed Particle Hydrodynamics* e Método de Lattice Boltzmann

2.2 APLICAÇÕES DA MODELAGEM COMPUTACIONAL EM ÁREAS ESTRATÉGICAS

2.2.1 CIÊNCIAS DA VIDA

BIOINFORMÁTICA

A Bioinformática é uma área de pesquisa multidisciplinar que combina a Biologia, a Ciência da Computação e Estatística para analisar e interpretar dados biológicos ou biomédicos. Desenvolveu-se nos últimos 30 anos, inicialmente para analisar dados gerados pelo sequenciamento de DNA e hoje é utilizada largamente em pesquisas relacionadas à biotecnologia, saúde humana e animal e agropecuária, dentre outras áreas de aplicação.

Objetivos específicos:

- Desenvolver e aplicar ferramentas, métodos e algoritmos para estudos de genômica comparativa, estrutural e funcional de organismos de interesse biotecnológico e agrícola, bem como relacionados à área da saúde humana, aplicando computação de alto desempenho, técnicas de inteligência e sistemas de gerenciamento de workflows científicos (SGWfC).
- Analisar, armazenar e disponibilizar os resultados de sequenciamento genômico de alta performance realizados na Unidade de Genômica Computacional Darcy Fontoura de Almeida.

Metas:

- Sequenciamento de genomas de diferentes espécies utilizando as plataformas de sequenciamento de alto desempenho da UGCDFa.
- Desenvolvimento e aplicação de Bancos de dados, SGWfC, SGBD, e ferramentas para Bioinformática e Biologia Computacional.
- Desenvolvimento de Ferramentas, métodos e algoritmos para análise de dados biológicos (montagem de sequências, predição e anotação de genes) e para estudos de genômica comparativa, estrutural e funcional.
- Montagem, anotação, comparação de genomas, transcritomas e metagenomas utilizando a ferramenta SABIA.
- Disponibilização de um portal web na área de Bioinformática e Biologia Computacional com acesso a *softwares* para comunidade científica.

- Expansão da inserção do LABINFO na comunidade científica nacional e internacional por meio de projetos de colaboração.
- Evolução de banco de dados biológicos e *workflows* científicos aplicados a Bioinformática e Biologia Computacional.

BIOLOGIA COMPUTACIONAL

Modelagem de sistemas moleculares

A modelagem de sistemas moleculares na biologia é marcadamente multidisciplinar. Envolve biologia, física, química, computação de alto desempenho e matemática aplicada. Ela se propõe a investigar propriedades físico-químicas de macromoléculas (por exemplo, proteínas, DNA) e modelar suas interações com pequenas moléculas orgânicas de interesse biológico e/ou farmacológico, utilizando métodos teóricos e técnicas computacionais. O Grupo de Modelagem Molecular de Sistemas Biológicos (GMMSB) do LNCC vem nos últimos anos investindo consideravelmente no desenvolvimento de novos métodos, algoritmos e programas nas áreas de desenho racional de fármacos e de predição de estruturas de proteínas. Na área dedesenho racional de fármacos, foram desenvolvidos no LNCC o programa DockThor (registrado no INPI) e o correspondente portal web (acessível em www.dockthor.lncc.br) de atracamento molecular receptor-ligante.

Na área de predição de estruturas de proteínas foram desenvolvidos no LNCC os programas GAPF (*Genetic Algorithm for Protein Folding*), GAHP, Profrager, todos registrados no INPI, e os portais web MHOLline (www.mholline.lncc.br) e Profrager (www.lncc.br/sinapad/Profrager).

Outro destaque foi a participação do GMMSB/LNCC na 12ª edição do CASP (“*Critical Assessment of Protein Structure Prediction*”, evento bianual para avaliação do estado da arte das metodologias computacionais para predição de estruturas de proteínas. O grupo do LNCC participou do CASP12 na categoria predições de estruturas, sem uso de moldes, com uso de servidor (acoplado ao supercomputador Santos Dumont), utilizando metodologias próprias desenvolvidas ao longo de mais de uma década de pesquisas no LNCC.

Nos próximos cinco anos a área de modelagem molecular continuará tendo uma atuação forte no desenvolvimento de novos métodos e algoritmos para o estudo de macromoléculas biológicas, principalmente em estudos associados à área de desenho racional de fármacos e engenharia de proteínas. A perspectiva é que as metodologias desenvolvidas no LNCC se tornem ainda mais robustas e com maior capacidade preditiva

e sejam disponibilizadas através de portais web, gratuitos para a comunidade científica brasileira, acoplados ao novo supercomputador Santos Dumont. É também esperado que pesquisas aplicadas, em colaboração com grupos experimentais da bioquímica, biologia molecular e química medicinal, ganhem ainda mais em importância.

Objetivos específicos:

- Desenvolver novos métodos, algoritmos e ferramentas computacionais capazes de obter resultados mais precisos e confiáveis nos seguintes tópicos:
 - Previsão de estruturas de macromoléculas;
 - Estudo da dinâmica e função de enzimas e receptores proteicos e de suas correlações com processos fisiopatológicos;
 - Desenho racional de moléculas candidatas a fármacos.
- Obter resultados em pesquisas aplicadas em engenharia de proteínas e nanobiotecnologia.

Metas:

- Desenvolvimento de novos métodos algoritmos e programas na área de desenho racional de fármacos baseado em estruturas.
- Desenvolvimento de novos métodos, algoritmos e programas na área de predição de estruturas de proteínas, com destaque para o desenvolvimento do programa de predição de proteínas GAPF e do portal web para modelagem comparativa em larga escala MHOLline.
- Desenvolvimento de estudos aplicados nas áreas de planejamento racional de fármacos e bioengenharia em parcerias com grupos de pesquisa do meio acadêmico e empresas de biotecnologia.
- Implementação de portal web DockThor-VS para triagem virtual em larga escala de moléculas candidatas à fármacos acoplado ao supercomputador Santos Dumont.
- Implementação do portal web para geração de estruturas de proteínas utilizando o programa GAPF utilizando técnicas com ou sem *templates* acoplado ao supercomputador Santos Dumont.
- Aprimorar o programa de *docking* receptor-ligante Dockthor em aspectos que permitam uma melhoria na sua capacidade de previsão tanto da geometria de interação receptor-ligante quanto da estimativa da constante de inibição.
- Aprimorar o programa GAPF para predição de estruturas de proteínas.
- Manter e aprimorar o portal web MHOLline (www.mholline.Incc.br) e migrá-lo para o parque computacional do CENAPAD/SINAPAD.
- Organizar a VIII EMMSB, em 2016, a XI EMMSB, em 2018, e a X EMMSB, em 2020.

Crescimento Tumoral e Terapias

A modelagem computacional tem contribuído substancialmente em quase todas as áreas da ciência, incluindo a biologia tumoral e oncologia. Em particular, a última década presenciou uma explosão na literatura de modelos matemáticos e computacionais do surgimento, crescimento e invasão de tumores em tecidos vivos.

É especialmente relevante o progresso na direção de tratamentos paciente-específicos, tornados possíveis através de previsões de simulações computacionais. Isto se deve à uma série de fatores dentre os quais o crescente entendimento pela comunidade médica dos principais mecanismos que levam ao surgimento e crescimento de vários tipos de câncer e do papel dos domínios ômicos na investigação das origens e crescimento do câncer e novas terapias para combatê-lo. Tais avanços têm levado ao desenvolvimento de novos modelos multiescala, que representam eventos que ocorrem numa variedade de escalas espaciais e temporais, desde o nível de tecidos e órgãos até os subcelulares.

Em particular, o gradual surgimento e crescente papel da ciência médica preditiva, fundamentada em análise matemática, teorias estatística e probabilística, ciências experimental e computacional, torna possível abordar em profundidade os conceitos de validade e previsão de modelos em presença de incertezas. Tais conceitos representam a vanguarda do ferramental indispensável para calibrar e validar modelos de crescimento de tumores que vêm sendo desenvolvidos recentemente.

Os avanços tecnológicos e numéricos em computação de alto desempenho, na análise e computação de grande massa de dados e de imagens, e de novos paradigmas em modelagem, que permitem integrar de forma híbrida modelos baseados em agentes, modelos em campo de fase e modelos estocásticos, dentre outros, têm propiciado um arsenal de novos ferramentais com grande potencial para contribuir de forma relevante no desenvolvimento de modelos de crescimento tumoral realistas, que possam ser efetivamente usados na prática médica.

Objetivos específicos:

- Consolidar e expandir a área de modelagem computacional do crescimento tumoral;

Metas:

- Promoção de encontro anual em Modelagem do Crescimento Tumoral no contexto do Programa de Verão do LNCC.

2.2.3 MEDICINA E SAÚDE

PLANEJAMENTO RACIONAL DE FÁRMACOS

Metodologias computacionais têm sido cada vez mais utilizadas pelas indústrias farmacêuticas e grupos de pesquisa da área de química medicinal para a busca e otimização de novas moléculas candidatas a fármacos. Neste cenário, a triagem virtual em larga escala utilizando métodos computacionais de atracamento molecular receptor-ligante tem se destacado como uma importante ferramenta para o sucesso e a redução de custos no planejamento racional de novos fármacos. Esta técnica utiliza informações sobre a estrutura molecular do receptor e consiste em analisar computacionalmente uma grande quantidade de ligantes com o objetivo de selecionar, de acordo com os modos de ligação e afinidades de ligação previstos, os compostos provavelmente mais ativos farmacologicamente para determinada doença.

Nos últimos quinze anos o Grupo de Modelagem Molecular de Sistemas Biológicos do LNCC (GMMSB/LNCC) tem se dedicado ao desenvolvimento de uma ferramenta de atracamento molecular receptor-ligante para a realização de estudos de planejamento racional de fármacos. O portal web DockThor (<http://www.dockthor.lncc.br>) foi lançado em nível nacional na 65ª Reunião Anual da Sociedade Brasileira para o Progresso da Ciência (SBPC) no ano de 2013, disponibilizando gratuitamente para a comunidade acadêmica o primeiro programa brasileiro (e possivelmente de todo o hemisfério sul) de atracamento molecular receptor-ligante.

O portal DockThor está acoplado às facilidades de computação de alto-desempenho do SINAPAD (Sistema Nacional de Processamento de Alto Desempenho – www.lncc.br/sinapad/) e desde o seu lançamento tem tido uma utilização bastante significativa tanto em número de *jobs* submetidos (~4000 *jobs* por ano) quanto em número de visitantes únicos (~3000 por ano). É importante ressaltar que o programa DockThor possui metodologias (i.e., algoritmos de otimização estocástica e metodologias para previsão da afinidade receptor-ligante) totalmente desenvolvidas em pesquisas realizadas no LNCC. Análises comparativas de desempenho com os programas de atracamento mais utilizados no mundo indicam que o programa DockThor tem potencial para ser altamente competitivo no cenário internacional.

Um dos grandes objetivos nos próximos anos é o desenvolvimento do portal web DockThor-VS para triagem virtual em larga escala de moléculas candidatas à fármacos acoplado à plataforma do supercomputador Santos Dumont. Este portal permitirá a disponibilização de um serviço de alto nível (com tecnologia própria e sem a necessidade de pagamentos de licenças de uso de *software*) em uma área estratégica para a

comunidade acadêmica e empresas ligadas à área de química medicinal e fármacos no Brasil. O desenvolvimento deste portal envolve uma forte colaboração com o INCT-INOVAR (Instituto Nacional de Ciência e Tecnologia de Fármacos e Medicamentos – <http://www.inct-inovar.ccs.ufrj.br/>), FIOCRUZ e unidades de pesquisa do MCTIC (e.g., LNBio, INPA).

No que se refere à pesquisa aplicada para o planejamento de novos fármacos o GMMSB/LNCC tem tido uma atuação bastante forte tendo como destaque as seguintes colaborações e projetos com grupos experimentais da química medicinal brasileira:

(i) participação no INCT-INOVAR, tendo sido encomendado um projeto de caráter prioritário intitulado “Desenvolvimento de plataforma de triagem virtual em larga escala utilizando o programa DockThor específica para alvos moleculares da classe de proteínas cinases, consideradas potentes alvos terapêuticos para o tratamento de doenças inflamatórias, câncer e doenças negligenciadas”;

(ii) colaborações envolvendo diversos grupos do Instituto Oswaldo Cruz – FIOCRUZ no âmbito de doenças negligenciadas (e.g., doença de Chagas e Leishmaniose) no sentido de se criar uma biblioteca de alvos moleculares e se viabilizar estudos de *inverse docking* (i.e., onde se tem uma molécula ativa conhecida mas não se conhece o alvo molecular);

(iii) colaboração com o grupo de química medicinal LASSBio-UFRJ no desenvolvimento de fármacos para o tratamento de doenças inflamatórias crônicas e câncer;

(iv) colaboração com pesquisadores do Instituto de Química Orgânica – UFRGS para o desenvolvimento de fármacos para o tratamento da doença de Alzheimer (colaboração que resultou no depósito em 2015 de uma patente de moléculas candidatas a fármacos inibidoras da enzima acetilcolinesterase);

(v) projeto de colaboração com pesquisadores da EMBRAPA para o desenvolvimento de fármacos para o tratamento de câncer de pele. A perspectiva para os próximos cinco anos é que estas colaborações se consolidem ainda mais e que possam ser planejadas e testadas diversas moléculas que possam ser alvo de depósito de patentes.

Objetivos específicos:

- Implementar plataforma para triagem virtual em larga escala de moléculas candidatas à fármacos acoplado ao supercomputador Santos Dumont.
- Desenvolver estudos aplicados nas áreas de planejamento racional de fármacos

em parcerias com grupos de pesquisa do meio acadêmico e empresas de biotecnologia.

Metas:

- Desenvolvimento do Portal Web DockThor-VS.
- Evolução do programa DockThor em aspectos que permitam uma melhoria na sua capacidade de previsão tanto da geometria de interação receptor-ligante quanto da estimativa da constante de inibição..
- Definição de uma política de uso nacional do portal DockThor-VS para grupos de pesquisa e empresas;
- Planejamento racional de compostos candidatos a fármacos para o tratamento de doenças negligenciadas, câncer e processos inflamatórios.
- Desenvolvimento de novas abordagens metodológicas para a triagem virtual em larga escala de moléculas candidatas à fármacos.
- Construção de bibliotecas virtuais de compostos associados a produtos naturais e sintéticos (i.e., bibliotecas públicas e pertencentes a grupos de pesquisa brasileiros) e de alvos moleculares de interesse para o desenvolvimento de novos fármacos para o tratamento de doenças negligenciadas e outros processos fisiopatológicos.
- Identificação e implementação de parcerias para o desenvolvimento das metodologias de triagem virtual e das bibliotecas virtuais de compostos e alvos moleculares.

MODELAGEM DO SISTEMA CARDIOVASCULAR HUMANO

O avanço da Computação Científica tem promovido um impacto profundo na área médica, permitindo: (i) a síntese do diagnóstico por imagem que, acoplada à modelagem e simulação, permite o desenvolvimento de novas técnicas terapêuticas, melhorando assim procedimentos e tratamentos médicos; e (ii) o desenvolvimento de modelos computacionais dos diversos sistemas que compõem o corpo humano, integrando anatomia, fisiologia, mecânica e biologia, e produzindo uma mudança de paradigma sobre a pesquisa cardiovascular.

Por exemplo, através das imagens médicas é possível reconstruir a geometria de estruturas anatômicas de relevância para diagnose de doenças assim como para alimentar modelos computacionais de maneira a simular o comportamento dos sistemas fisiológicos em condições normais ou alterados por doenças ou procedimentos médicos e suas aplicações no planejamento cirúrgico. Particularmente, a doença aterotrombótica arterial cardiovascular, cujas principais manifestações clínicas são a doença cardíaca isquêmica e

o acidente vascular cerebral, é responsável no Brasil, pelo dobro das mortes da segunda causa de óbitos: o grupo “Câncer”. Também é a principal causa de aposentadorias por doença, a principal causa de anos de vida perdidos ajustados por incapacitação, a segunda causa de internações e pelo principal custo destas internações devido principalmente ao uso de procedimentos de alta complexidade e alto custo. A modelagem computacional do SCVH mostra que podemos, através de modelos adequados, alcançar um melhor entendimento do processo de aterotrombose.

O grupo HeMoLab (Hemodynamics Modeling Laboratory, <http://hemolab.lncc.br>) vem desenvolvendo modelos computacionais e métodos de simulação numérica visando a representação de sistemas fisiológicos, mais especificamente o sistema cardiovascular humano (SCVH). As atividades de pesquisa estão orientadas para o desenvolvimento de modelos físicos de meios contínuos envolvendo múltiplas escalas mediante formulações variacionais, assim como também para o desenvolvimento de estratégias numéricas baseadas no método dos elementos finitos, método de volumes finitos e método de Lattice-Boltzmann. Modelos físicos para o escoamento sanguíneo, a interação fluido-estrutura, fenômenos de propagação de ondas, o processamento de imagens médicas, o comportamento de materiais ou a identificação de parâmetros via problemas inversos são alguns dos exemplos dos problemas de interesse no grupo. Em particular, cabe mencionar que no grupo HeMoLab foi desenvolvido um modelo computacional denominado ADAN (*Anatomically Detailed Arterial Network*) que incorpora todas as artérias reconhecidas pela literatura especializada (mais de 2000 vasos arteriais) para um paciente masculino de anatomia vascular padrão. Este modelo possui a geometria desenvolvida no espaço 3D, e ainda possibilita simular a propagação do pulso de pressão e a distribuição do fluxo sanguíneo ao longo dos territórios vasculares do corpo humano (<http://hemolab.lncc.br/adan-web>). Este modelo será a base para o desenvolvimento das atividades de pesquisa por trás dos objetivos detalhados a seguir.

Nos próximos quatro anos, a área de modelagem do sistema cardiovascular humano continuará mantendo um forte foco em atividades relacionadas com os problemas exemplificados no parágrafo anterior, visando o desenvolvimento incremental de tecnologia de modelagem computacional e simulação numérica computacionalmente eficiente e matematicamente robusta.

Objetivos específicos:

- Desenvolver um modelo anatomicamente detalhado do sistema arterial-venoso (modelo ADAVN) integrando dados anatômicos, conceitos fisiológicos e modelos

mecânicos visando à predição de estados fisiológicos e patofisiológicos gerados por doença ou alterações cirúrgicas.

- Desenvolver modelos funcionais da circulação coronária integrando dados de pacientes específicos e conhecimento anatômico e fisiológico com o objetivo de fornecer ferramentas computacionais complementares e não invasivas para a avaliação de risco de isquemia de miocárdio.
- Desenvolver modelos de multiescala para estudar processos de falha de material em tecidos biológicos, objetivando melhorar as ferramentas de análise de rompimento de tecidos arteriais.

Metas:

- Desenvolver simulador 1D do escoamento sanguíneo para o sistema ADAVN com a incorporação de todas as artérias e vasos caracterizados pela literatura médica.
- Desenvolver simulador 3D e 1D do escoamento sanguíneo no sistema coronariano e estabelecer índices computacionais robustos para a avaliação da severidade funcional de uma estenose.
- Desenvolver simulador 3D do escoamento sanguíneo e sua interação com as paredes arteriais em aneurismas cerebrais e estabelecer índices computacionais robustos para a avaliação da severidade funcional dos mesmos e o correspondente risco de ruptura.
- Desenvolver simulador 3D multiescala para o processo de dano e falha dos materiais constituintes das paredes arteriais.

Ações:

- Estudo e desenvolvimento de modelo computacional, anatomicamente detalhado e preciso desde o ponto de vista fisiológico e hemodinâmico (incluindo o desenvolvimento de novos e robustos métodos computacionais para sua solução) da parte venosa do SCVH, e sua incorporação no Modelo ADAN.
- Estudo e desenvolvimento de técnicas computacionais avançadas para a construção automática da topologia e geometria de leitos arteriais periféricos capazes de permitir (i) a incorporação de circulação colateral existentes nestes leitos e (ii) o estudo da microcirculação em tecidos e órgãos.
- Utilizando de dados obtidos por técnicas de processamento de imagens médicas de tomografia computadorizada (CT), realizar o estudo e desenvolvimento de modelo computacional avançado customizado ao paciente e anatomicamente detalhado e preciso desde o ponto de vista fisiológico e hemodinâmico da

circulação sanguínea coronariana, incluindo efeitos da circulação colateral existentes no miocárdio de maneira a permitir o cálculo computacional não invasivo da reserva de fluxo fracionado de miocárdio.

- Utilizando dados obtidos por técnicas de processamento de imagens médicas de ultrassom intravascular (IVUS), e angiografias ortogonais (AX), realizar o estudo e desenvolvimento de modelo computacional para a identificação das propriedades mecânicas dos tecidos possibilitando a histologia computacional *in-vivo* e conseqüentemente o desenvolvimento de índices hemodinâmicos associados com a estabilidade (ruptura) de placas de ateroma, auxiliando na diagnose e terapia.
- Estudo e desenvolvimento de modelos 1D+ (modelos com interpolação híbrida) na modelagem do escoamento sanguíneo em domínios tubulares capazes de competir com o mesmo grau de precisão que os modelos 3D de Navier-Stokes com custo computacional equivalente ao modelo 1D, de forma a acelerar tempos de cálculo de tensões de cisalhamento em paredes arteriais.
- Estudo e desenvolvimento de modelos Multiescalana simulação computacional do comportamento de tecidos biológicos. Empregando formulações variacionais multi-escala, teoria da mecânica dos sólidos em grandes deformações e análise de sensibilidade à mudança de parâmetros associados ao colágeno, elastina e músculo liso e a mudança de forma do distrito arterial, estes modelos permitem simular o surgimento, a evolução e instabilidade (ruptura) de aneurismas e ou placas de ateroma.

2.2.4 ENERGIA E RECURSOS NATURAIS

SIMULAÇÃO COMPUTACIONAL EM ENGENHARIA DE PETRÓLEO E GÁS

O LNCC, em parceria com a Petrobrás, tem desenvolvido projetos de interesse estratégico para o país relacionados com o desenvolvimento de novos modelos computacionais para descrever escoamento multifásico (água, óleo e gás) em reservatórios altamente heterogêneos com acoplamento geomecânico.

Recentemente a descoberta das formações geológicas do pré-sal brasileiro em águas profundas tem dado a origem a cenários altamente desafiadores para o desenvolvimento de uma modelagem computacional acurada e, conseqüentemente, predição das curvas de produção durante prospecção secundária. Neste contexto analisando as imagens sísmicas das formações do pré-sal com elevado número de fraturas a equipe do LNCC vem desenvolvendo novas metodologias numéricas para descrever o acoplamento

hidrogeomecânico em rochas carbonáticas fraturadas e efetuar o processo de mudança de escala para o cômputo de propriedades efetivas em células definidas em escalas mais grossas a serem utilizadas com parâmetros de entrada nos simuladores comerciais.

Além do planejamento inserido nestas atividades relacionadas com a exploração de petróleo no pré-sal brasileiro, o LNCC está desenvolvendo novos projetos com a Petrobrás na área de modelagem na escala de poros de processos de recuperação avançada de petróleo. Em particular a diminuição da salinidade da água de injeção tem mostrado aumento da recuperação de óleo devido a alteração da molhabilidade da rocha que compõe o reservatório.

A modelagem dos fenômenos eletroquímicos que governam as interações entre a água, óleo e a rocha tem sido objeto de estudo da equipe do LNCC. Nesta nova parceria com a Petrobrás o LNCC irá desenvolver uma metodologia multiescala inovadora para determinar a composição de salinidade e pH da água de injeção para maximizar a extração de óleo.

Além das duas atividades mencionadas o LNCC, também em parceria com a Petrobrás, desenvolve novas formulações de natureza multiescala para simular a extração de gás em reservatórios não convencionais de gás de xisto compostos por múltiplos níveis de porosidade. O modelo construído recebeu reconhecimento internacional e atualmente um simulador vem sendo desenvolvido para ser entregue a Petrobrás capaz de descrever a extração de gás com comportamento anômalo residente em nano poros.

Objetivo específico:

- Desenvolver novos modelos computacionais para descrever escoamento multifásico (água, óleo e gás) em reservatórios altamente heterogêneos com acoplamento geomecânico
- Desenvolver novas metodologias numéricas para descrever o acoplamento hidro-geomecânico em rochas carbonáticas fraturadas, em especial em formações do pré-sal com elevado número de fraturas
- Desenvolver modelagem multiescala de processos de recuperação avançada de petróleo com aplicações ao pré-sal brasileiro e aos arenitos da bacia de Campos.

Metas:

- Desenvolver dois pré-simuladores computacionais para descrever:
 - Processos de extração de petróleo por injeção de água e incorporando acoplamento hidro-geomecânico nas formações geológicas heterogêneas que compõem o pré-sal brasileiro.

- Processos de extração de gás em reservatórios não convencionais com múltiplos níveis de porosidade com aplicação a gás de xisto.

DUTOS, RISERS E FLEXÍVEIS

A modelagem e simulação computacionais em Mecânica dos Sólidos é uma das áreas fundadoras do LNCC. A partir daí foram desenvolvidas novas formulações variacionais e novos métodos de elementos finitos, algoritmos, metodologias e códigos computacionais para a abordagem de problemas em elasticidade, plasticidade, fluência, não linearidades geométricas, análise dinâmica, análise sísmica de estruturas, etc. Em alguns casos esses desenvolvimentos estiveram mais próximos à pesquisa básica e em outros estiveram ligados a aplicações como nos trabalhos desenvolvidos em cooperação para a Comissão Nacional de Energia Nuclear (CNEN). A modelagem em Mecânica dos Sólidos continua presente em várias fases da atividade de pesquisa no LNCC como, por exemplo, na modelagem multifásica de reservatórios de óleo ou gás.

Desde 1999 o LNCC vem cooperando com o Centro de Pesquisas da Petrobrás (Cenpes/Petrobrás) na análise estrutural e de estabilidade de dutos e *risers* de perfuração com configurações especiais, como o caso de um duto aquecido, enterrado, com configuração em zigzag, instalado no leito da baía de Guanabara ou de dutos contendo defeitos de corrosão. No caso de dutos corroídos a atividade se justifica em virtude do envelhecimento da ampla rede instalada e da necessidade de que sejam estabelecidos critérios que definam a necessidade de reparo ou de substituição de trechos danificados. Com a exploração em águas cada vez mais profundas, como no caso do pré-sal, surge a necessidade de conhecimento dos mecanismos de colapso da seção transversal de dutos corroídos utilizados nesta atividade. Em geral esses dutos possuem parede espessa e o colapso se dá no regime de comportamento elastoplástico. O estabelecimento de limites de utilização desses dutos se faz através de análise numérica, estrutural cujos resultados devem ser validados por experimentos realizados em laboratório.

Risers e flexíveis são também de grande importância na exploração submarina de petróleo e devemos dar continuidade a trabalhos iniciados visando estas aplicações.

Objetivos específicos:

- Desenvolver metodologia para a análise dinâmica de dutos flexíveis
- Ampliar conhecimento sobre os mecanismos de colapso de dutos corroídos utilizados na exploração de petróleo em águas profundas, como é o caso do pré-sal.

Metas:

- Desenvolvimento de regras que estabeleçam de forma rápida o limite de utilização de dutos corroídos utilizados na exploração de petróleo em águas profundas, em conjunto com o Cenpes/Petrobrás.
- Desenvolvimento de código computacional para a análise dinâmica de flexíveis.

SIMULAÇÃO COMPUTACIONAL EM SÍSMICA E NANOFOTÔNICA

O fenômeno de propagação de ondas baseado em sistemas de equações diferenciais parciais definidos em domínios altamente heterogêneos está na base de uma grande variedade de aplicações. Dentre estas, os problemas de propagação de ondas eletromagnéticas e de ondas sísmicas em meios heterogêneos tem atraído particular atenção de pesquisadores nos últimos anos por ainda apresentarem grandes desafios tecnológicos, principalmente do ponto de vista da simulação computacional realista de tais fenômenos.

Neste tema, estamos particularmente interessados na propagação de ondas eletromagnéticas, mais especificamente luz (ondas ópticas), em interação com estruturas em escalas nanométricas (nanofotônica). A nanofotônica é o campo da ciência que estuda a interação da luz com a matéria composta por nanoestruturas. Os domínios de aplicação são inúmeros, com particular ênfase nas áreas de biomedicina e eletrônica, para os quais a modelagem numérica tem complementado de modo fundamental os estudos experimentais. Os modelos matemáticos que estão na base do estudo de ondas eletromagnéticas em interação com nanoestruturas baseiam-se nas equações de Maxwell definidas em domínio temporal. A este normalmente associa-se modelos de dispersão como os modelos de Drude e de Drude-Lorentz.

A segunda motivação deste projeto é a modelagem computacional da propagação de ondas sísmicas em diferentes matrizes geológicas. A ideia é inferir as diferentes estruturas geológicas (rochas) consideradas contínuas, porém altamente heterogêneas (granularidades diferentes, por exemplo) separadas por longas e largas interfaces. A simulação computacional da propagação de ondas sísmicas em meios altamente heterogêneos demanda a resolução de modelos elastodinâmicos com coeficientes heterogêneos, acoplados a modelos de interfaces (condições de transmissão). A incorporação precisa a modelagem computacional das diferentes escalas e interfaces é fundamental, pois tais estruturas influenciam fortemente a determinação dos parâmetros físicos via coletas reversas de dados a partir dos captadores.

Finalmente, técnicas de problemas inversos em sísmica e de otimização topológica em nanofotônica são necessárias, onde a simulação direta do fenômeno de propagação de ondas constitui etapa fundamental em ambos algoritmos de problemas inversos e de otimização topológica.

Objetivos Específicos:

- Aplicar novos métodos e algoritmos numéricos multiescalas adaptados à resolução de modelos com múltiplas escalas ou grandes contrastes, - de forma precisa e que sejam naturalmente paralelizáveis nas modernas arquiteturas massivamente paralelas
- Desenvolver novos métodos de otimização topológica de dispositivos nanofotônicos e de problemas inversos em sísmica utilizando o conceito de derivadas topológicas
- Desenvolver modelos de propagação de ondas eletromagnéticas e elastodinâmicas em meios altamente heterogêneos, visando a resolução/simulação computacional de problemas das áreas de sísmica e de nanofotônica.

Metas:

- Desenvolvimento de métodos multiescalas para as equações de Maxwell no domínio da frequência e do tempo em meios altamente heterogêneos;
- Desenvolvimento de métodos multiescalas para as equações elastodinâmicas no domínio da frequência e do tempo em meios altamente heterogêneos;
- Desenvolvimento de novos métodos de otimização topológica de dispositivos nanofotônicos;
- Desenvolvimento de novos métodos de resolução de problemas inversos em sísmica.

3. PILAR ESTRUTURANTE II: MODERNIZAÇÃO E AMPLIAÇÃO DA INFRAESTRUTURA DE CT&I

O LNCC disponibiliza infraestrutura computacional para diversas iniciativas e demandas estratégicas do Sistema Nacional de CT&I portanto tem como Objetivo Específico:

Manter e ampliar as plataformas computacionais para atender as demandas de projetos e programas técnico-científicos do LNCC, do Sistema Nacional de CT&I e da comunidade de P&D nacional

3.1 SINAPAD - SISTEMA NACIONAL DE COMPUTAÇÃO DE ALTO DESEMPENHO

A supercomputação ou Processamento de Alto Desempenho (PAD) tornou-se um elemento essencial para a competitividade das economias, principalmente por dar suporte à inovação científica e industrial. No domínio da ciência, a simulação baseada nessa tecnologia é instrumental em várias áreas do conhecimento:

- Na climatologia: desenvolvimento de cenários climáticos futuros associados ao aquecimento global, às emissões de gases e a mudanças de uso da terra.
- Na engenharia: o projeto e o desenvolvimento de novos materiais, principalmente os advindos das novas áreas de nanoeletrônica e nanotecnologia, bem como de novas energias, como biomassa e eólica.
- Na saúde: a melhor compreensão dos mecanismos existentes nas células e organismos e também nas patologias e substâncias terapêuticas.
- Em Tecnologia da Informação (TI): o tratamento eficaz de grandes volumes de dados em setores da indústria e de serviços de TI, tais como, entretenimento, TV digital e governo eletrônico.
- Na segurança pública: o emprego de criptografia avançada e o processamento de dados de inteligência para o combate à criminalidade.

O país, portanto, precisa dispor das infraestruturas adequadas para utilizar uma ampla gama de recursos computacionais de alto desempenho de modo a atender à demanda proveniente de aplicações industriais, científicas e de serviços. Nesse contexto, o LNCC coordena o Sistema Nacional de Processamento de Alto Desempenho – SINAPAD, uma rede de centros de processamento de alto desempenho hospedados em universidades e centros de pesquisa distribuídos pelo Brasil: CESUP/UFRGS, CENAPAD-SP/UNICAMP, CPTEC/INPE, NACAD/UF RJ, CENAPAD-MG/UFMG, CENAPAD-PE/UFPE, CENAPAD-CE/UFC, CENAPAD-AM/INPA, e o próprio LNCC.

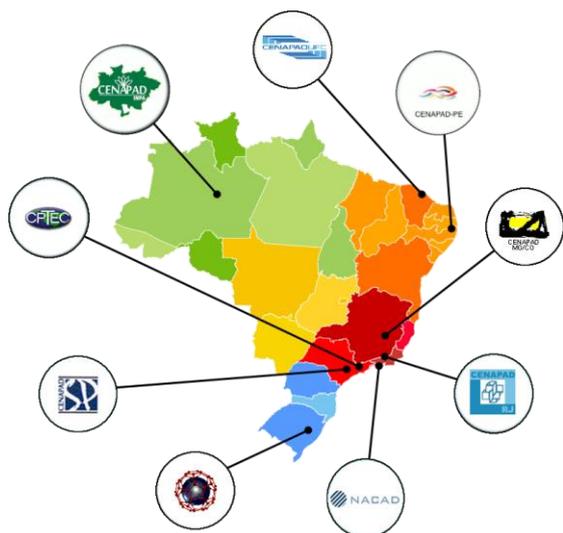


FIGURA 1: A REDE SINAPAD.

Um passo importantíssimo foi dado pelo país, através do LNCC, com a aquisição em 2015 do supercomputador Santos Dumont – SDumont, hoje o nó principal da rede SINAPAD. Na primeira chamada aberta de projetos para uso do SDumont, 72 projetos se apresentaram e muitas instituições estão se organizando para participar da próxima chamada demonstrando a importância de multiplicar cada vez mais a capacidade computacional do país.

Mais ainda, apontam para a necessidade premente de investimento em supercomputação no País, para de fato alcançarmos uma posição de liderança mundial em ciência, fatos como o forte investimento chinês na produção de seu primeiro supercomputador com tecnologia 100% nacional, e anunciado pela lista Top500 como o mais rápido do mundo em junho de 2016. Esse investimento não deve, portanto, se restringir a oferta de supercomputação para os setores de pesquisa, indústria e governo. Ele deve fomentar também o desenvolvimento de um ecossistema ao redor das tecnologias de supercomputação, com a difusão das tecnologias de simulação computacional de alto desempenho por um número muito maior de setores e a criação de um círculo virtuoso para a economia brasileira, destarte crescentemente intensiva em conhecimento e inovação.

O conjunto desses elementos então justifica que o SINAPAD tenha papel estruturante no contexto nacional devido à sua potencialidade de geração de conhecimentos em geral e, em particular, de contribuição para a disseminação de tecnologias de alto valor estratégico.

3.2 SUPERCOMPUTADOR SANTOS DUMONT

Em 2015 foi adquirido o supercomputador Petaflópico Santos Dumont (SDumont), que se integra ao programa SINAPAD como o maior de sua categoria na América Latina e entre os 300 de maior capacidade do mundo segundo a prestigiosa lista Top500 (<http://www.top500.org>). Sua capacidade de processamento de 1.1 PFlops corresponde a execução de 10^{15} operações de ponto flutuante por segundo, e sua capacidade de armazenamento de 1.7 PBytes corresponde ao armazenamento de aproximadamente 380.000 DVDs. Seu acesso é provido por meio dos enlaces multigigabits provisionados pela Rede Nacional de Pesquisa (RNP).

O SDumont potencializa o desenvolvimento das pesquisas em diversas áreas nos Institutos e Universidades atendidas pela RNP. Até o presente momento, 18,4% dos projetos em desenvolvimento no SDumont são da área de ciências biológicas, 15,8% da área de engenharias, 15,8% da área de química e 13,2% da área de física, entre outros em várias outras áreas como ciência da computação, geociências, ciências da saúde, ciências agrárias, astronomia, meteorologia, ciência dos materiais e ciências sociais. São abordados nesses projetos temas de grande impacto nacional, como pesquisas relacionadas ao vírus da zika, a doenças coronarianas, ou ao gerenciamento de parques eólicos, entre tantos outros.

O SDumont é dotado de diferentes arquiteturas de processamento, para atender necessidades específicas dos usuários. Por exemplo, alguns podem ser dependentes da disponibilidade de memória de grande porte, acessível por um número limitado de processadores, enquanto outros fazem melhor uso de arquiteturas de memória maciçamente distribuída. O apoio aos usuários da computação de alto desempenho é, portanto, um aspecto crítico para que o uso da computação petaflópica seja eficiente. É necessário prover uma equipe de especialistas em computação, com ênfase na arquitetura de supercomputadores, técnicas de otimização de uso do sistema e desenvolvimento de “softwares” nas principais áreas de utilização.

Outro aspecto importante relacionado à aquisição do supercomputador é que ele determina a transferência de tecnologia correspondente (uso e fabricação) de uma máquina petaflópica para o Brasil e abre a possibilidade do país buscar autonomia e independência na geração de produtos e prestação de serviços de PAD.

A mobilização da cadeia produtiva local em torno das tecnologias do Grupo BULL na área de computação de alto desempenho, estimula, de um lado, a autonomia no uso da tecnologia no Brasil, de outro, melhores condições de acesso do mercado brasileiro à oferta de supercomputadores, por parte da empresa francesa. A transferência de tecnologia

do Grupo BULL para um consórcio de atores foi aprovada pelo MCTI. O consórcio constituído envolve produtores locais, o LNCC, a RNP, e o Centro de Tecnologia da Informação Renato Archer (CTI), na forma de convênios e planos de trabalho específicos para cada atividade.

O acordo explicita, também, a disposição do Governo Francês de promover a capacitação e a transferência de tecnologia industrial e de processamento de dados ao Brasil na área de computação de alto desempenho com os seguintes objetivos:

- Criação de um centro nacional, baseado na RNP em parceria com o LNCC, para estimular, identificar e prover serviços altamente demandantes das tecnologias de informação e comunicação (TIC) para projetos de e-ciência acadêmicos e de inovação.
- Mobilização da cadeia produtiva local do Brasil na área de computação de alto desempenho estimulando, de um lado, melhores condições de transferência da tecnologia francesa e, do outro, a autonomia no uso da tecnologia por parte do Brasil.
- Identificar parceiros locais interessados na composição de consórcio, em parceria com o LNCC, a RNP, o CTI, para a transferência de tecnologia francesa assegurando ao governo brasileiro a autonomia e independência na geração de produtos e prestação de Serviços de PAD, em quaisquer circunstâncias nas quais lhe forem frustradas suas ambições de acesso a estes produtos e serviços. O consórcio de atores locais deverá envolver produtores locais na forma de convênios e planos de trabalho específicos para cada atividade.

Esta iniciativa, em sendo bem sucedida, elevará a qualidade da ciberinfraestrutura brasileira para PD&I a níveis comparáveis com os aqueles alcançados em economias mundiais semelhantes à brasileira.

Os resultados esperados que já estão começando a se materializar, são:

- Implementação no SINAPAD de uma nuvem computacional para atendimento adequado da demanda da comunidade acadêmica e do setor empresarial do País por serviços de simulação computacional e de análise de dados. O Sistema Brasileiro de Computação Petaflópica é o componente principal dessa nuvem, que deve satisfazer a evolução dos requisitos fundamentais de segurança da informação para atender as demandas dos usuários.
- Potencialização da atratividade do SINAPAD como parceiro de ICTs e organizações públicas e privadas na utilização da capacidade de processamento e

armazenamento ofertada sob a forma de nuvem computacional, em particular nas áreas estratégicas da ENCTI.

- Associação do Brasil ao esforço europeu na vanguarda tecnológica em PAD, tornando-o apto ao projeto, construção e operação de supercomputadores em pé de igualdade com as principais potências econômicas e industriais. Essa colocação se dará através da iniciativa do centro de pesquisa da BULL em Petrópolis e a associação com parceiros em ICT's.
- Mobilização da cadeia produtiva nacional, criando condições de autonomia para a produção de bens e serviços de PAD no Brasil, com respeito à propriedade intelectual da Bull, bem como respeito às pretensões estratégicas do Brasil
- Garantia da autonomia brasileira na gestão de segurança do sistema de PAD fornecido pela Bull.
- Atender a demanda de computação e armazenamento por aplicações avançadas e de altíssima complexidade computacional em áreas de PD&I estratégicas
- Assegurar ao Brasil uma paridade no acesso a recursos de PAD e redes de 100G para PD&I brasileiros
- Ampliar a penetração de serviços avançados de armazenamento e computação em áreas de conhecimento fortemente demandantes de TIC
- Permitir ao Brasil desenvolver pesquisas e produtos com as características básicas de confiabilidade, integridade, disponibilidade e autenticidade
- Mobilizar a cadeia produtiva brasileira em torno das tecnologias de PAD, garantindo o acesso do Brasil à tecnologia de PAD em qualquer circunstância.
- Aumentar a qualificação de professores e pesquisadores através do uso de aplicações e ferramentas de computação, simulação e uso de informações científicas e tecnológicas\
- Capacitar a indústria brasileira em tecnologias avançadas de empacotamento, montagem, integração, qualificação, aceitação e manutenção de sistemas de PAD

3.3 LABINFO

O Laboratório Nacional de Bioinformática (LABINFO), desde 2000, é responsável pela bioinformática de diversos projetos em genômica – sequenciamento e estudo de genomas de organismos –, realizados em redes de cooperação, com instituições acadêmicas nacionais e estrangeiras. O LABINFO analisa, armazena e disponibiliza os dados gerados por esses projetos. Desde 2008, a partir de iniciativa conjunta do MCTI e do Ministério da Saúde, coordena e opera a Unidade de Genômica Computacional Darcy Fontoura de

Almeida - UGCDFFA, que é uma *facilitie* multiusuária de sequenciamento de DNA de alta capacidade. Até o momento, os equipamentos da UGCDFFA foram utilizados para o sequenciamento de mais de 400 genomas, desde genomas humanos, de fungos, parasitos, vírus, bactérias de interesse na saúde humana, animal e vegetal, bem como vários metagenomas de interesse biotecnológico.

Estão em andamento cinco projetos de sequenciamento que envolvem várias amostras de todo o País do vírus da Zika e células de tecidos infectados e sadios (em parceria com a FIOCRUZ), além de um tripanossomatídeo (em parceria com a UFRJ), um fungo (colaboração com a UFRGS), duas amostras de aves (Rede Sisbioaves) e de pacientes com doenças crônicas (colaboração com o IFF/FIOCRUZ). Mais de 50 usuários, 35 departamentos de diferentes Instituições nacionais e internacionais e 14 Programas de Pós-Graduação utilizaram a infraestrutura oferecida pela UGCDFFA.

O LABINFO/UGCDFFA têm como competência as seguintes funções:

- i) Geração, processamento e análise de sequências de nucleotídeos (genomas, metagenomas, transcritomas, amplicons, exomas, pequenos RNAs e cDNA) geradas a partir do sequenciamento em grande escala no Brasil e fora do País;
- ii) Armazenamento e distribuição para a comunidade científica dos resultados da análise de sequências de nucleotídeos geradas pela UGCDFFA; iii) Gerenciamento de Redes Genômicas no Brasil;
- iv) Desenvolvimento de *softwares* de aplicação em Biologia Computacional, incluindo o primeiro *software* para montagem e anotação automática de genomas no Brasil, denominado SABÍÁ: *System of Automated Bacterial (genome) Integrated Annotation*;
- v) Desenvolvimento de *workflows* científicos em larga escala com a utilização de ambientes de processamento de alto desempenho (PAD) e técnicas de inteligência artificial, mineração de dados, aprendizado de máquinas para a exploração, tratamento e análise de grande volume de dados biológicos (*Big Data*);
- vi) Desenvolvimento de banco de dados biológicos de aplicação em Bioinformática;
- vii) Desenvolvimento de metodologias matemáticas e computacionais para estudos de genômica comparativa, estrutural e funcional de organismos de interesse biotecnológico, ambiental e agrícola, bem como relacionados à saúde humana;
- viii) Criação e manutenção da Rede Nacional de Bioinformática que disponibilizará um Portal na área;

- ix) Capacitação e orientação de dissertações de mestrado e teses de doutorado na linha de pesquisa em Bioinformática e Biologia Computacional no Programa de Pós-graduação em Modelagem Computacional do LNCC, no Departamento de Genética da UFRJ, na Universidade de Lyon1 e no Curso de virologia do Instituto Evandro Chagas;
- x) Formação de recursos humanos especializados em Bioinformática e Biologia Computacional (através da organização de cursos no Brasil e fora do País);
- xi) Desenvolvimento de projetos de cooperação científica com instituições nacionais e internacionais.

A (ENCTI 2016-2019) estabeleceu como áreas a Biotecnologia e as Ciências da Vida. Essas dependem significativamente da Bioinformática e da Biologia Computacional. Em consonância com o pilar **Modernização e ampliação da infraestrutura de CT&I**, o LABINFO e a UGCDA oferecem, respectivamente, estruturas computacionais e de sequenciamento abertas e multiusuárias em um ambiente robusto para o desenvolvimento de pesquisas com a geração e análise de grande quantidade de dados (*Big Data*) na área. Além disso, é responsável pela estruturação da primeira Rede Brasileira de Bioinformática e de Biologia Computacional, com apoio direto do MCTI.

Outro pilar fundamental da ENCTI com o qual o LABINFO apresenta forte aderência é a **Promoção da pesquisa científica básica e tecnológica**. Este Laboratório tem colaborado com diversos grupos de pesquisa no País, sendo o coordenador da Rede Genoma Brasileiro (que envolve 33 instituições de pesquisa e ensino) e possui ainda colaborações com outras instituições nacionais, tais como: FIOCRUZ (Rio de Janeiro, Bahia, Paraná, Minas), INMETRO, UERJ, Instituto Butantan, UFJF, UNESP São José dos Campos, UNIPAMPA, USP, UNIFESP. Entre os convênios de pesquisas assinados com parceiros internacionais podemos citar: Instituto Superior Técnico – Universidade Técnica de Lisboa (IST-UTL), Ludwig Institute for Câncer Research, Universidade de Lyon1, Laboratório redes Pathogènes Émergents (LPE) e a Rede GABRIEL da Fundação Mérieux, o Instituto Pasteur de Paris.

Além do fortalecimento das áreas relativas às ciências básicas, o LABINFO insere-se em alguns dos temas em CT&I propostos como estratégicos para o desenvolvimento, autonomia e soberania nacional (ENCTI 2016-2019). Vários projetos de pesquisa em que o LABINFO participa estão nas linhas temáticas estratégicas e de fronteira, tais como, **Água, Alimentos, Biomass e Bioeconomia, Economia e Sociedade Digital, Energia e Saúde**.

Dentro dos temas **água e alimentos** estudos em colaboração com uma rede de laboratórios e instituições de pesquisa dos Estados do RJ, SC e RN estão sendo realizados para o melhoramento do cultivo de ostras na costa fluminense e para o monitoramento da qualidade e comercialização do pescado. Ainda, dentro do tema **alimentos**, destacamos na área agropecuária, estudos de biodiversidade em parceria com a EMBRAPA SOJA, que visam contribuir no aumento da fixação biológica do nitrogênio em cultivos de leguminosas de importância econômica e social para o Brasil. E, em colaboração com a UFRGS e a Universidade de Lyon (França) estudos de biodiversidade de bactérias que vivem no trato respiratório de suínos visando à saúde animal.

Na área de **biotecnologia marinha**, o LABINFO participa da Rede Nacional de Pesquisa em Biotecnologia Marinha que compreende 23 laboratórios e institutos de pesquisa. Neste projeto estão sendo realizadas investigações de microrganismos marinhos e de organismos de ocorrência restrita ao Brasil que são fonte potencial para descoberta de novas moléculas e produtos biotecnológicos, promovendo desta forma, a valoração da biodiversidade endêmica brasileira. Ainda, o LABINFO participa da Rede SISBIOAVES (composta por 18 laboratórios nacionais e internacionais) onde estudos de **biodiversidade genômica** de aves brasileiras estão sendo conduzidos de forma a ajudar a responder questões centrais sobre a fisiologia e comportamento que envolve o aprendizado do canto e imitação da fala.

Nas áreas de **biotecnologia e energia**, o LABINFO atua em colaboração com a UFRN e o Instituto Técnico de Lisboa na identificação de espécies bacterianas, genes e rotas metabólicas relacionadas à degradação de hidrocarbonetos e biosurfactantes, entre outros, e que possuem potencial para o desenvolvimento de estratégias biotecnológicas aplicadas a produção de petróleo.

Vários projetos de pesquisa na **área da saúde** estão em desenvolvimento visando à prevenção, o diagnóstico e o tratamento de doenças crônicas não transmissíveis e de doenças infecciosas. Podemos destacar a busca de marcadores diferenciais para diagnóstico e de virulência e patogenicidade em tripanosomatídeos patogênicos e não patogênicos em colaboração com a UFRJ; USP, UFMG, entre outras instituições; estudos de infecção pelo vírus ZIKA e as manifestações neurológicas relacionadas visando à proposição de tratamentos, prevenção, inclusive o desenvolvimento de vacinas em conjunto com a FIOCRUZ; análise molecular das doenças genéticas crônicas visando soluções inovadoras na saúde pública com o IFF (Fernandes Figueira – FioCruz); caracterização da diversidade viral e bacteriana e na cocirculação com o vírus da gripe em vários países do terceiro mundo como membro integrante da REDE GABRIEL (Global Approach to Biological Research, Infectious diseases and Epidemics in Low-income

countries), composta por 16 laboratórios internacionais e; identificação de processos adaptativos em *Staphylococcus aureus* durante infecção osteoarticular em parceria com a UFRJ e o Centre National de Références Staphylocoques - CIRI-HCL-CNR.

Dentro do tema **Economia e Sociedade Digital**, o LABINFO atua no desenvolvimento de ferramentas de bioinformática e na criação do Portal de Bioinformática, este último juntamente com a UFMG e o LNBio, para gerenciamento e análise de dados biológicos usando plataformas de alto desempenho (HPC) para o processamento de grande massa de dados (*Big Data*).

3.4 INCT-MACC

Ao longo das últimas décadas, pesquisadores das áreas de modelagem computacional conjuntamente com profissionais da área médica começaram a desenvolver e aplicar, de forma cada vez mais frequente (e intensiva), ferramentas baseadas em modelos computacionais dentro de diferentes áreas da prática médica. Como consequência disso, os modelos computacionais têm evoluído significativamente com relação às capacidades de descrição e predição dos fenômenos mais relevantes que governam a resposta de um determinado sistema fisiológico e sua interação com os outros sistemas que integram o corpo humano, tanto em condições normais como em situações alteradas por doenças e/ou intervenções humanas (cirurgias, por exemplo). O anterior, somado ao vertiginoso aumento do desempenho dos computadores e das máquinas da aquisição de imagens médicas, tem dado lugar ao surgimento de um novo paradigma na medicina: *a medicina orientada a paciente específico assistida por modelagem computacional e simulação numérica*. Esta quebra de paradigma é o entendimento da ciência no campo da medicina como sendo construída de uma forma mais racional e baseada nos princípios fundamentais da mecânica, a partir da combinação de dados clínicos e de imagens médicas com dados obtidos dos sistemas/modelos computacionais desenvolvidos capazes de representar com alto grau de precisão os fenômenos subjacentes aos sistemas fisiológicos do corpo humano.

A Medicina Assistida por Computação Científica promove um campo de pesquisa altamente interdisciplinar ao permitir:

- Desenvolver modelos e simuladores precisos dos diversos sistemas do corpo humano e da sua inter-relação, que integram anatomia, fisiologia, propriedades biomecânicas, biologia celular e bioquímica, para aplicações terapêuticas, de pesquisa, de formação e de treinamento de recursos humanos;
- Realizar diagnósticos por imagem e complementar esses diagnósticos com

modelos matemáticos e simulação computacional, alavancando o desenvolvimento de novas técnicas terapêuticas para melhorar procedimentos e tratamentos médicos;

- Desenvolver modelos específicos para cada paciente, de maneira a servir como um laboratório *in silico* para diagnóstico, estudo de patologias e análise de outras informações médicas sobre o paciente baseando-se em modelos matemáticos; esses modelos virtuais dos pacientes permitem aumentar a comunicação entre paciente e médico e fornecem referência para exames, patologias e mudanças que acontecem com o passar do tempo;
- Utilizar modelos matemáticos e simuladores computacionais de alta precisão, para planejamento cirúrgico, treinamento e credenciamento médico, promovendo uma verdadeira interação do usuário com órgãos humanos simulados, com propriedades físicas e fisiológicas realísticas, úteis para educação e pesquisa e desenvolvimento de aplicações médicas.

Como indicador da tendência no nível nacional e internacional cabe mencionar o Instituto Nacional de Ciência e Tecnologia em Medicina Assistida por Computação Científica (INCT-MACC) criado em Novembro de 2008 e cujo novo projeto foi classificado em primeiro lugar pelo CNPq para o período 2016-2021.

Objetivos Específicos:

- Consolidar o reconhecimento do LNCC como um instituto de excelência em Computação Científica aplicada à medicina e como referência nacional e internacional em pesquisa e desenvolvimento, transferência de tecnologia, inovação e formação de recursos humanos altamente qualificados na área.
- Desenvolver ambientes computacionais de alto desempenho para que as aplicações médicas geradas fiquem à disposição das comunidades ligadas à pesquisa e à saúde e, como consequência, a serviço da população.
- Consolidar a atuação do LNCC como coordenador de uma rede de P&D em Medicina Assistida por Computação Científica.

Metas:

- Implantação de um sistema computacional para simulação do sistema arterial-venoso humano, em condições normais ou alteradas por doenças ou procedimentos médicos, incorporando assimilação de dados específicos de pacientes.

- Implantação de um sistema computacional para processamento avançado de imagens médicas de modalidades de tomografia computadorizada e de ultrassom intravascular, permitindo a reconstrução tridimensional de estruturas anatômicas.
- Implantação de um sistema computacional para diagnose de risco de doenças cardiovasculares por meio da fusão de imagens médicas, assimilação de dados, simulação computacional e computação distribuída de alto desempenho.
- Implantação de um sistema computacional para ambientes virtuais colaborativos de realidade virtual, aumentada e telemanipulação voltados para a diagnose a distância, planejamento de procedimentos médicos e treinamento e capacitação profissional.
- Implantação de um Ciberambiente capaz de disponibilizar via *web* os sistemas listados anteriormente para a comunidade médica e de saúde do País.

3.5 INCT-CiD

Alinhado com o estabelecimento de uma área de pesquisa básica e tecnológica em Ciência de Dados neste documento, o LNCC coordena a proposta de criação do Instituto Nacional de Ciência e Tecnologia em Ciência de Dados (INCT-CiD), recomendada para financiamento pelo CNPq em maio de 2016. O INCT-CiD propõe-se a atuar no eixo ciência-indústria-governo por meio de ciência de dados.

O desafio, e principal objetivo, do INCT-CiD é identificar os princípios, métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados, suplantando as dificuldades inerentes ao grande volume de dados em análise. Motivados pelos desafios anteriormente enfrentados pelas equipes de seus laboratórios associados em áreas tão diversas quanto astronomia, biologia, biodiversidade, defesa cibernética, educação, esporte, Internet, mobilidade urbana, petróleo & gás, saúde, segurança da informação, e comunicação móvel; o INCT-CiD propõe-se contribuir para a estruturação científica desta área tanto na formação de recursos humanos altamente qualificados quanto no desenvolvimento de pesquisa básica e aplicada na fronteira do conhecimento em ciência de dados.

Especificamente, o INCT-CiD conta com três linhas de pesquisa principais:

- (i) gerência de dados;
- (ii) análise de dados;
- (iii) análise de redes complexas.

O instituto será formado por algumas instituições de pesquisa e ensino localizadas no Nordeste, Sudeste e Sul do Brasil. Conta, igualmente, com laboratórios associados

privados visando transferência de conhecimento, e impulsionará seus resultados inovadores por meio de parcerias com *startups* incubadas em algumas das instituições parceiras.

O potencial tecnológico da proposta se evidencia ainda pela declaração de interesse nos resultados esperados do INCT-CiD por uma série de empresas privadas (de pequenas *startups* a grandes multinacionais) atuantes em diferentes mercados, setores do governo e instituições de pesquisa. A inovação visada pelo INCT-CiD não se furta da perspectiva internacional, espelhada em seu *International Advisory Board*, formado por cientistas renomados internacionalmente que aceitaram acompanhar e aconselhar estrategicamente o progresso do instituto, balizando sua atuação e contribuição de alto nível. O INCT-CiD pretende, portanto, posicionar o Brasil na direção da nova ciência baseada em dados preparando recursos humanos altamente qualificados e desenvolvendo o alicerce para sua projeção de forma relevante na sociedade do conhecimento.

3.6 LInEA

O Laboratório Interinstitucional e-Astronomia, apoiado pelo Observatório Nacional (ON), o Laboratório Nacional de Computação Científica (LNCC) e a Rede Nacional de Ensino e Pesquisa (RNP), tem por finalidade dar suporte à participação brasileira em levantamentos astronômicos gerando grandes volumes de dados. Para alcançar os objetivos científicos destes projetos, o LInEA gerencia toda uma infraestrutura de armazenamento, processamento, análise e distribuição de dados astronômicos e desenvolve tecnologia para lidar com os desafios de projetos envolvendo Big Data. Essa infraestrutura está abrigada na sede do LNCC. Participam do LInEA pesquisadores e técnicos dos institutos do MCTIC mencionados acima, além de professores de universidades.

Faz parte da missão do LInEA: é dar apoio logístico e financeiro para a participação de cientistas e estudantes em grandes projetos internacionais; Implantar e manter um Centro de Dados Astronômico para facilitar a exploração científica dos dados provenientes destes experimentos bem como de outras fontes que possam complementar estes dados; Desenvolver e manter um portal científico para a validação e mineração de dados e para conduzir de forma eficiente a análise científica deste grandes acervos; Ajudar na formação de jovens pesquisadores e na sua preparação para atuar pró-ativamente em grandes colaborações internacionais. Ajudar na formação de tecnologistas para lidar com projetos de *Big Data*; Contribuir na divulgação dos resultados de pesquisa para o grande público.

Pesquisadores e tecnologistas do LNCC contribuem na supervisão da operação da infraestrutura do LInEA montada no CPD do LNCC e atuam como consultores na

contratação das empresas de TI que prestam serviço ao projeto, na definição das máquinas a serem adquiridas, em questões relativas ao desenho de banco de dados, *workflows* e computação de alto desempenho, em particular, no uso do supercomputador SDumont.

3.7 OUTROS PROJETOS QUE RECEBEM APOIO DO LNCC

Historicamente, o LNCC mantém a vertente de serviços computacionais, seja com foco a disponibilização da sua plataforma computacional, seja na participação de projetos colaborativos e consultoria especializada para outros órgãos.

Nesse sentido, o LNCC mantém uma infraestrutura física (Centro de Dados e um conjunto de salas) na sua antiga sede, na cidade do Rio de Janeiro, denominada LNCC-Rio, para suportar projetos, de seu interesse e de organizações parceira, que necessitem de alto grau de conectividade à Internet.

No Centro de Dados, estão hospedadas importantes iniciativas:

- i) O Ponto de Presença da Rede Nacional de Ensino e Pesquisa (POP-RJ/RNP). Desde a criação do projeto RNP, grupo operacional do LNCC administra e opera este ponto, toda infraestrutura do Backbone até a interligação de instituições usuárias, incluindo a interligação da rede estadual (REDERIO) com outras redes Metropolitanas, a exemplo da Rede Metropolitana de Petrópolis (RMP) com a RNP.
- ii) A operação do Internet Exchange (IX.br), que é um projeto do Comitê Gestor da Internet Brasileira (CGI-Br), operado pelo Núcleo de Informação da Internet Brasileira (NIC-BR). O objetivo deste projeto é disponibilizar uma infraestrutura para interligação de Redes, que compõem a Internet Brasileira. A atuação do IX.br concentra-se nas regiões metropolitanas no País que apresentam grande interesse de troca de tráfego comerciais acadêmicos, sob uma gerência centralizada em área metropolitana.
- iii) A infraestrutura de Ensino a Distância do Exército Brasileiro, resultado de um longo convênio de cooperação técnica com o Departamento de Ensino e Cultura do EB (DECEX/EB).

O LNCC, através do seu grupo operacional, presta consultoria a outras organizações do Governo Federal, como o Jardim Botânico do Rio de Janeiro (JBRJ), por exemplo. O LNCC juntamente com o grupo de pesquisa, de museologia e do próprio grupo de TI do JBRJ, projetou e ajudou a implementar a atual infraestrutura de Redes de Alta Velocidade dentro do parque do JBRJ, bem como os servidores e unidades de armazenamento de

grandes volumes de dados. Além disso o LNCC hospeda (*hosting*) alguns servidores do JBRJ que recebem imagens de material Botânico de outros centros de pesquisa espalhados pelo mundo, dentro de um projeto de repatriação de informações da Flora Brasileira.

Através da Fundação de Apoio ao Desenvolvimento da Computação Científica (FACC), o LNCC elabora e executa projetos dentro da sua área de competência em parceria com organizações públicas e privadas.

As relações e parcerias do LNCC se estabelecem de forma direta ou via fundação, com empresas tecnológicas nacionais criando um canal de troca de conhecimento em áreas sensíveis e de ponta.

4. PILAR ESTRUTURANTE III: FORMAÇÃO, ATRAÇÃO E FIXAÇÃO DE RECURSOS HUMANOS

Desde a sua criação, o LNCC tem uma presença marcante na formação de recursos humanos, organizando escolas temáticas e programas sazonais e oferecendo cursos especializados. Essas atividades, fortalecidas com a criação dos programas de mestrado e doutorado de Modelagem Computacional em 2000, contribuem para a formação profissional num sentido amplo, bem como para a divulgação da Computação Científica e a inclusão social e digital. As áreas da fronteira do conhecimento nas quais o Laboratório desenvolve pesquisas são, e continuarão sendo no quinquênio 2016 a 2020, fonte de inspiração para futuros cursos e escolas.

A formação de recursos humanos no LNCC tem em vista a conjuntura em que evolui o perfil profissional requerido pelo mercado de trabalho, em função do progresso do conhecimento científico e tecnológico, especialmente em relação à Computação Científica, que abre continuamente novas frentes de atuação.

Apoiado nas competências consolidadas do LNCC em modelagem, análise matemática e numérica, sistemas e controles, simulação computacional e validação de modelos, os programas visam a acompanhar e a responder às novas demandas, originadas pelo uso crescente da Computação Científica na solução de problemas da engenharia, ciências da computação, física, química, biologia, saúde, meio ambiente, entre outras.

4.1 PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Iniciado em março de 2000, o programa tem conceito 6 da Capes, mantido na última avaliação (período 2010-2012). Os alunos veem de diferentes áreas do conhecimento. Em 2016, conta com 61 alunos de doutorado e 32 de mestrado. Já foram formados pelo programa até a presente data, 102 doutores e 137 mestres.

O programa está fundamentado nas competências do LNCC, sedimentadas ao longo dos seus 35 anos de existência, nas áreas de Modelagem Computacional e suas Aplicações com ênfase em Matemática Aplicada e Computacional, Mecânica Computacional, Sistemas e Controle e-Ciência da Computação.

A Pós-Graduação do LNCC está envolvida com o Programa e Difusão do Conhecimento, coordenado pelo Instituto Federal da Bahia e que tem como participantes as seguintes Instituições: SENAI/CIMATEC - LNCC - UNEB - UEFS (<http://www.difusao.dmmdc.ufba.br>)

A formação tem como perspectiva preparar alunos com sólidos conhecimentos em modelagem computacional e ciências correlatas, com visão interdisciplinar e capacidade para lidar com os desafios resultantes da permanente evolução da Computação Científica.

Adicionalmente, o programa conta com Ambiente Virtual de Aprendizado cujo objetivo é apoiar o corpo docente e discente no dia-a-dia do programa, bem como promover, na medida do possível, Cursos a Distância de interesse do próprio programa e de empresas, visando o aperfeiçoamento e a inovação em prol da sociedade Brasileira. Por sua vez, essa plataforma de Ensino a Distância (EAD) é baseada na ferramenta (aplicação) Moodle, (https://moodle.org/?lang=pt_br) customizada pelo próprio LNCC. Por ser uma plataforma livre, é regida pela licença GNU GENERAL PUBLIC LICENSE (https://docs.moodle.org/all/pt_br/Licença_do_Moodle), que permite a criação de cursos *online* no conceito de *e-learning*, administração de usuários e cursos, acesso dos alunos as salas de estudos; interação entre alunos e professores e administração dos cursos por parte dos professores.

Objetivos específicos:

- Manter a qualidade da pós-graduação do LNCC e fortalecer a formação interdisciplinar.
- Apoiar o intercâmbio de alunos com instituições internacionais e nacionais.
- Divulgar amplamente o programa de pós-graduação visando ao recrutamento de alunos do País e do exterior em diversas áreas do conhecimento.

Metas:

- Manutenção do número de doutores e mestres formados anualmente com relação a média do período 2011 a 2015.
- Ampliação da publicação de trabalhos em periódicos indexados gerados pelas teses e dissertações com relação à média do período 2011 a 2015.
- Atualização permanente do currículo do programa de pós-graduação incluindo temas na fronteira do conhecimento.

- Estímulo a elaboração de teses e dissertações em temas ligados aos projetos do LNCC aproveitando as bolsas vinculadas aos projetos.
- Atualização da plataforma de Ensino a Distância desenvolvida pelo LNCC e disponibilizá-la para os cursos da pós-graduação.
- Ampliação da produção de monografias didáticas, notas de aula e livros vinculados ao programa de pós-graduação.

4.2 DIVULGAÇÃO CIENTÍFICA: DIFUSÃO DO CONHECIMENTO

O LNCC tem por tradição utilizar a realização de eventos científicos para capacitação interna e como veículo de disseminação do conhecimento científico. As escolas temáticas realizadas são mecanismos eficientes que cooperaram para a consolidação das pesquisas desenvolvidas nas cinco grandes áreas de concentração do Programa de Pós-Graduação em Modelagem Computacional do LNCC. Com a realização de eventos promovidos pelo Programa, esperam-se reflexos positivos no desenvolvimento de teses doutorado e dissertações de mestrado em temas ligados às suas áreas de atuação e o estabelecimento de possíveis colaborações entre grupos do Brasil e do exterior..

São organizados pelos pesquisadores do LNCC Encontros Acadêmicos, Minicursos, Palestras e Jornadas. Também são oferecidas no LNCC atividades de extensão, tais como seminários, palestras, congressos, workshops, cursos especializados, escolas temáticas, visitas de intercâmbio científico e programas sazonais.

Objetivos específicos:

- Oferecer à comunidade acadêmica e profissional fóruns de aprendizado e/ou discussão de temas de grande relevância científica ou tecnológica.
- Oferecer à sociedade em geral a oportunidade de conhecer temas científicos ou tecnológicos que podem impactar positivamente em suas vidas.
- Despertar vocação científica e incentivar novos talentos potenciais entre estudantes de graduação.
- Contribuir para reduzir o tempo médio de titulação de mestres e doutores.
- Propiciar à instituição um instrumento de formulação de política de iniciação à pesquisa para alunos de graduação.
- Estimular uma maior articulação entre a graduação e pós-graduação.
- Contribuir para a formação de recursos humanos para a pesquisa.

- Contribuir para que o aluno de mestrado (ou doutorado) ao defender a dissertação (ou Tese) tenha um artigo aceito em congresso ou periódico indexado.
- Estimular pesquisadores produtivos a envolverem alunos de graduação nas atividades científica, tecnológica e artística-cultural.
- Proporcionar ao bolsista, orientado por pesquisador qualificado, a aprendizagem de técnicas e métodos de pesquisa, bem como estimular o desenvolvimento do pensar cientificamente e da criatividade, decorrentes das condições criadas pelo confronto direto com os problemas de pesquisa.

Metas:

A realização dos seguintes eventos regulares nacionais e internacionais no LNCC são as Metas de divulgação científica do LNCC. Os eventos/metasp, listados a seguir, tem coordenação e forte participação dos membros do corpo docente/discente do Programa de Pós-Graduação, pesquisadores visitantes e alunos de outras instituições de ensino trabalhando em projetos.

➤ **Programa de Verão do LNCC**

Esse programa acontece anualmente, sempre nos meses de janeiro e fevereiro, desde 2003. Oferece diversos cursos de diferentes níveis de forma aberta à comunidade. Seu caráter multidisciplinar gera intercâmbio científico e disseminação do conhecimento de base e de fronteira. As atividades do Programa de Verão são na sua maioria organizadas por pesquisadores, tecnologistas e colaboradores (pesquisadores visitantes, pós doutorandos e alunos da pós-graduação) do LNCC. Seu público-alvo é composto por alunos em final da graduação, alunos de Pós-graduação, professores, pesquisadores e profissionais de empresas e indústrias. As seguintes atividades fazem parte da programação básica do evento:

- Disciplinas obrigatórias e optativas do Programa de Pós-Graduação em Modelagem Computacional.
- Encontro Acadêmico em Modelagem Computacional, desde 2007, privilegia a participação dos alunos e ex-alunos do Programa de Pós-Graduação do LNCC. Esses últimos são convidados a apresentar suas pesquisas e as consequências de sua formação, servindo de estímulo e ampliação de conhecimento para os alunos em curso.
- Jornada de Iniciação Científica & Tecnológica, onde são apresentados os trabalhos premiados na avaliação do PIBIC e PIBICT do LNCC e colaborações

de alunos de graduação de diversas instituições de pesquisa e ensino do país, na forma de painéis.

- Seminários, Jornadas e Encontros temáticos como: a Semana de Programação Massivamente Paralela, as Jornadas de Ecologia Teórica, Ciência de Dados, etc.
- Minicursos Avulsos.

➤ **PIBIC e PIBITI**

Com a finalidade de despertar vocação científica e incentivar talentos potenciais entre estudantes de graduação universitária, mediante participação em projetos de pesquisa, orientados por pesquisadores qualificados, o LNCC mantém também, desde 1998, o Programa Institucional de Bolsas de Iniciação Científica – PIBIC. Em 2014 a Pós-Graduação foi contemplada com o Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação – PIBITI.

Para avaliar os programas, todos os anos, no mês de agosto, o LNCC realiza a Jornada de Iniciação Científica, onde são apresentados em forma de pôsteres os trabalhos desenvolvidos pelos alunos, os quais são avaliados pela Comissão Interna do PIBIC/PIBITI e por avaliador convidado externo. A partir dessas apresentações, é elaborado o relatório Anais da Jornada, o qual é enviado ao CNPq, que avalia o mecanismo de controle de qualidade dos Programas.

Atualmente, as universidades com bolsistas destes programas são: Universidade Federal de Juiz de Fora (UFJF), Universidades Católica de Petrópolis (UCP), Universidade Estadual do Rio de Janeiro (UERJ), Universidade Estácio de Sá (UNESA), Instituto Superior de Tecnologia em Ciências da Computação de Petrópolis (FAETERJ-RJ), Universidade Federal do Vale do Jequitinhonha do Mucuri (UFVJM), Fundação Centro de Ciências e Educação Superior a Distância do Estado do Rio de Janeiro (CEDERJ/RJ), Centro Universitário Serra dos Órgãos (UNIFESO).

➤ **CBAB/CABBIO: “Ferramentas de Bioinformática Aplicadas às Análises de Sequências Transcritômicas”**

O CBAB/CABBIO é um curso anual de ferramentas de Bioinformática de 80 horas de aula (32 horas teóricas e 48 horas práticas) representa um dos cursos oferecidos pelo Centro Brasileiro-Argentino de Biotecnologia. Este curso vem sendo ministrado no LNCC desde sua criação em 2010, de forma ininterrupta, com enorme demanda durante toda sua vigência. A motivação principal para a criação deste curso foi, fundamentalmente, atender a significativa demanda por parte dos alunos de Pós-graduação que necessitam do processamento e análise de grande volume de dados provenientes das plataformas de sequenciamento de alto desempenho (genoma, metagenoma e transcrito) para seus

projetos de pesquisa. Anualmente, são oferecidas 40 vagas divididas entre alunos brasileiros (incluindo alunos da PPG-LNCC) e estrangeiros vindos da Argentina, Uruguai, Paraguai e Colômbia. Além disso, pode-se destacar que este curso tem contribuído com a formação de mais de 200 alunos de Pós-graduação na área de Bioinformática no Brasil, principalmente dando o suporte teórico e prático para o desenvolvimento de dissertações de mestrado e teses de doutorado de alta qualidade científica, fazendo jus ao papel Nacional do LNCC na área de Bioinformática.

➤ **Rede Avançada em Biologia Computacional (RABICÓ)**

A Rede atua na área de Bioinformática e Biologia Computacional, sendo financiada pela CAPES através do EDITAL BIOLOGIA COMPUTACIONAL – CAPES nº 051/2013. Esse projeto é desenvolvido por uma colaboração entre LNCC, UFRJ (COPPE, Departamento de Genética e Instituto de Bioquímica Médica) e UFRGS (Centro de Biotecnologia – Cbiot).

Objetivando a formação e aperfeiçoamento de recursos humanos, vários cursos específicos foram criados para atender aos integrantes da rede, sendo que esses cursos estão abertos para a comunidade científica do País. Até a presente data, a rede ministrou mais de 43 cursos e capacitou mais de 230 alunos na área de Biologia Computacional, desde o início das atividades em 2015. Este Programa conta com o suporte do Sistema Nacional de Computação de Alto Desempenho (SINAPAD) do MCTI, coordenado pelo LNCC.

➤ **Escola de Modelagem Molecular em Sistemas Biológicos**

A Escola de Modelagem Molecular em Sistemas Biológicos (EMMSB) é bianual e em 2016, completou a sua oitava edição. Conta com a participação de alunos e pesquisadores interessados em áreas relacionadas à modelagem molecular computacional, dinâmica molecular, modelagem molecular aplicada ao desenvolvimento de fármacos, metodologias de *docking* receptor-ligante, métodos computacionais, bioinformática e cálculos estocásticos e quânticos sobre biomoléculas.

A escola vem sendo ministrada há mais de 14 anos, sob a coordenação do LNCC, com a parceria do Instituto de Biofísica Carlos Chagas Filho da UFRJ. De caráter marcadamente multidisciplinar a escola ministra palestras e minicursos teóricos e práticos (básicos e avançados), envolvendo um amplo perfil de participantes tanto no nível da formação (i.e., participação de alunos de graduação, mestrado, doutorado, pós-doc e pesquisadores) quanto na área de formação (e.g., Biologia, Biomedicina, Física Química, Farmácia, Matemática, Engenharia, Computação e áreas afins). O evento tem alta demanda no País e são oferecidas em cada edição 180 vagas.

➤ **Ciclo de Palestras “Fique por Dentro”**

Desde 2008, o LNCC/MCT realiza uma extensa atividade de popularização da ciência voltada à comunidade do município de Petrópolis e região.

O Ciclo de Palestras: Fique por Dentro é anual e a programação inclui palestras sobre assuntos variados, proporcionando à sociedade maior proximidade com a ciência, cultura e educação. As palestras são abertas ao público e acontecem no auditório do LNCC, em Petrópolis.

Em 2016, por exemplo, o Seminário Educação *on-line* Professor Global foi realizado por uma empresa da Incubadora do LNCC e vencedora do prêmio ASSESPRO – RJ em 2012, 2013 e 2015, como a melhor Solução Educacional do Rio de Janeiro.

Os temas das palestras realizadas nos últimos anos evidenciam, esse caráter plural ao abordar diversas áreas de conhecimento: Realidade Virtual, Gerenciamento de Projetos e as Certificações PMP e CAPM do PMI (EUA), Ciência de Dados, Ocupação de Áreas de Risco e Mudanças Climáticas na Região Serrana Sul do Rio de Janeiro.

4.3 INTERCÂMBIOS

O LNCC tem por tradição utilizar a realização de eventos científicos para capacitação interna e como veículo de disseminação do conhecimento científico. As escolas realizadas são mecanismos eficientes que cooperaram para a consolidação das pesquisas desenvolvidas nas cinco grandes áreas de concentração do Programa de Pós-Graduação em Modelagem Computacional do LNCC. Com a realização de eventos promovidos pelo Programa, esperam-se reflexos positivos no desenvolvimento de teses de doutorado e dissertações de mestrado em temas ligados às suas áreas de atuação e o estabelecimento de possíveis colaborações entre grupos do Brasil e do exterior.

Os projetos de pesquisa, convênios e cooperações fomentam intercâmbios e parcerias científicas entre pesquisadores do LNCC e de outras instituições do exterior em diversos países.

Este fato se reflete também no intercâmbio de alunos entre o LNCC e instituições de Pós-Graduação do Brasil e do exterior.

5. PILAR ESTRUTURANTE IV: PROMOÇÃO DA INOVAÇÃO TECNOLÓGICA NAS EMPRESAS

A inovação tecnológica tem como finalidade promover o desenvolvimento de novos produtos, processos e serviços, ou significativa melhoria nos já existentes, buscando aumentar a competitividade. Para tanto, torna-se fundamental o envolvimento e integração dos agentes Estado, Academia e Empresa.

No LNCC, o ecossistema de inovação é composto pelo Núcleo de Inovação Tecnológica (NIT) e Incubadora de Empresas. O NIT tem como função implementar e administrar a política de inovação do LNCC, através da gestão dos processos de propriedade intelectual, registro de *software* e de componentes, *royalties*, etc. Já a Incubadora de Empresas tem como objetivo estimular a criação e o desenvolvimento de empreendimentos baseados em tecnologias inovadoras.

Esse ecossistema tem por objetivo estratégico criar e gerir mecanismos que transformem o conhecimento e as tecnologias geradas no LNCC em novos produtos, processos e serviços de alto valor agregado e inseridos no mercado nacional e internacional, protegendo a propriedade intelectual dessas invenções, produtos e tecnologias.

Objetivos Específicos:

- Através do Núcleo de Inovação Tecnológica – NIT
 - Prospectar e fortalecer parcerias com empresas, instituições governamentais e setores da sociedade civil;
 - Prospectar tecnologias desenvolvidas nos laboratórios de pesquisa do LNCC elegíveis ao processo de inovação;
 - Prospectar o mercado para identificação de demandas elegíveis ao desenvolvimento de projetos em parceria com o LNCC;
 - Criar mecanismos e instrumentos (modelos, padrões, referências) para a prestação de serviços e transferência de tecnologia entre LNCC e parceiros;
 - Manter articulação permanente com órgãos de fomento em busca de apoio financeiro ao desenvolvimento dos projetos do LNCC.
 - Tomar as providências necessárias para a proteção de direitos autorais e propriedade intelectual nos projetos do LNCC;
 - Participar de processos de inovação originados em entidades e empresas, apoiando-os com serviços técnicos, equipamentos e pesquisas.

- Através da Incubadora de Empresas:

- Estimular a criação e o desenvolvimento de empreendimentos baseados em tecnologias inovadoras;
- Criar e gerir mecanismos que transformem o conhecimento e as tecnologias geradas na instituição em novos produtos e processos de alto valor agregado;
- Promover a associação entre pesquisadores e empreendedores;
- Disseminar uma cultura empreendedora no LNCC.
- Apoiar o alcance de viabilidade técnica e econômica das empresas incubadas.

Metas:

- Através do Núcleo de Inovação Tecnológica – NIT
 - Cumprimento das exigências legais de garantia de direitos autorais e propriedade intelectual das novas tecnologias desenvolvidas pelo LNCC;
 - Elaboração e atualização permanente do portfólio das invenções, produtos e tecnologias desenvolvidas pelo LNCC ou com a participação de seus pesquisadores em parceria com outras ICTs;
 - Ampliação e fortalecimento das parcerias entre grupos de pesquisa e empresas do setor corporativo para o desenvolvimento conjunto de projetos de inovação.
 - Ampliação da geração de patentes;
 - Ampliação das parcerias que levem a geração de inovações.

- Através da Incubadora de Empresas
 - Manter edital de incubação para seleção de empresa permitindo a renovação e ampliação do número de Empresas Incubadas e Graduadas na Incubadora do LNCC;
 - Implantação do modelo de governança CERNE na Incubadora do LNCC.
 - Transformação das empresas incubadas em empresas de mercado ou viabilização da comercialização das tecnologias desenvolvidas por essas empresas.

6. ORGANIZAÇÃO ADMINISTRATIVA

Objetivos Específicos:

- Aprimorar as atividades administrativas e os seus mecanismos de gestão, visando atender ao LNCC com eficiência e eficácia.
- Planejar, monitorar e avaliar ações de capacitação e avaliação de desempenho a partir da identificação dos conhecimentos, habilidades e atitudes necessárias ao desempenho das funções dos servidores.
- Estimular a integração, cooperação e comunicação eficaz entre as equipes maximizando seu desempenho por meio de comprometimento, desenvolvimento de competências e espaço para empreender.

Metas:

- Implantação do Sistema Eletrônico de Informações (SEI) adotado pelo MCTIC no âmbito do Processo Eletrônico Nacional (PEN) em toda tramitação documental no LNCC;
- Evolução do SIGTEC (Sistema de Informações Gerenciais e Tecnológicas) permitindo maior interoperabilidade com os sistemas do MCTIC em especial a plataforma Lattes
- Melhoria e otimização nos processos de gestão;
- Revisão da Estrutura Organizacional do LNCC procurando adequá-la as novas demandas das áreas de pesquisa e serviços computacionais e considerando o cenário de restrição orçamentária;
- Implantação de calendário de planejamento e acompanhamento plurianual compatível com os processos de planejamento do MCTIC
- Ampliação do suporte da Coordenação de Administração (CAD) à gestão administrativa de projetos de pesquisa e desenvolvimento, especialmente os transversais;
- Elaboração de Plano Anual de Capacitação com foco em competências.