

1 Nome

INCT-MCS

Instituto Nacional de Ciência e Tecnologia para uma Sociedade Massivamente Conectada:
Uma Abordagem Ciberfísica e Social

2 Coordenação

Coordenador:

Prof. Virgílio A. F. Almeida
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

Subcoordenador:

Prof. Wagner Meira Jr.
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

3 Comitê Gestor

O comitê gestor do INCT-MCS é composto pelos seguintes pesquisadores:

Nome	CPF	Título	Cargo	Instituição
Alberto H. F. Laender	144.029.126-87	PhD	Professor Titular	UFMG
Altigran Soares da Silva	243.039.252-68	PhD	Professor Associado	UFAM
Antonio A. F. Loureiro	391.085.386-20	PhD	Professor Titular	UFMG
Geraldo Robson Mateus	300.670.496-72	PhD	Professor Titular	UFMG
Humberto T. Marques Neto	704.984.406-30	PhD	Professor Adjunto	PUC Minas
José Marcos S. Nogueira	144.132.646-49	PhD	Professor Titular	UFMG
Mario F. M. Campos	244.927.286-00	PhD	Professor Titular	UFMG
Nivio Ziviani	072.302.576-20	PhD	Professor Emérito	UFMG
Virgílio A. F. Almeida	130.465.196-72	PhD	Professor Titular	UFMG
Wagner Meira Jr.	509.960.376-91	PhD	Professor Titular	UFMG

4 Descrição

A missão do Instituto Nacional de Ciência e Tecnologia para uma Sociedade Massivamente Conectada: Uma Abordagem Ciberfísica e Social (INCT-MCS) é entender e desenvolver princípios científicos e tecnologias necessárias para as interações futuras entre seres humanos, computadores e o mundo físico. O INCT-MCS tem seus alicerces no Instituto Nacional de Ciência e Tecnologia para a Web (InWeb, 2008-2014), cujo foco primário é melhorar a integração entre a Web e a sociedade, a fim de tornar os serviços relacionados à informação mais acessíveis e efetivos. O INCT-MCS vai expandir esse foco para integrar os mundos físico e cibernético, considerando-se não somente a Web, mas todo tipo de rede formada por dispositivos conectados.

Os princípios que norteiam o INCT-MCS são: (1) desenvolver sistemas e tecnologias necessárias para construir sistemas ciberfísicos complexos que as pessoas possam usar com grande confiança; (2) acelerar a implantação de sistemas ciberfísicos e sociais em uma sociedade massivamente conectada através do desenvolvimento de métodos, ferramentas e componentes de software e hardware baseados em princípios transversais, validados por meio protótipos e cenários de teste; (3) transferir os resultados científicos e tecnológicos alcançados para a indústria.

Nosso propósito é suportar uma sociedade massivamente conectada por meio do avanço do estado-da-arte em termos de sensoriamento,¹ integração, modelagem e predição, bem como de mecanismos de decisão e atuação em múltiplos campos da ciência da computação e áreas afins. Como cenário particular de estudo, o instituto buscará projetar novos modelos, algoritmos, métodos e tecnologias para conduzir e suportar o planejamento e controle mais efetivo de cidades. Serão abordados temas como a mobilidade urbana, a segurança pública e a proteção do

¹Por “sensoriamento”, referimo-nos a qualquer forma de aquisição de dados, incluindo o monitoramento ambiental e da Web.

ambiente. Tais temas são, naturalmente, chave para promover crescimento econômico e para melhorar a qualidade de vida das pessoas.

As principais questões de pesquisa que pretendemos abordar são: (1) como efetivamente coletar dados do ambiente físico e virtual sem comprometer a sua segurança e a sua privacidade? (2) como integrar e gerenciar *terabytes* de dados ruidosos, incompletos e incertos provenientes de fontes heterogêneas, de modo a transformá-los em fontes valiosas de informação e conhecimento? (3) como transformar a informação em modelos práticos e robustos que permitam o entendimento de sistemas de grande complexidade e a predição de seu comportamento? (4) como tirar vantagem do melhor entendimento das interações entre as pessoas e dispositivos, a fim de que sejam criados novos mecanismos para tomada de decisões e atuação que venham a ter valor para a sociedade? (5) como fechar esse ciclo, isto é, gerar retro-alimentação, que possibilite analisar o impacto dos mecanismos de atuação, a fim de refinar os modelos existentes e permitir o projeto de novos modelos?

Para abordar tais questões, o INCT-MCS é composto por uma equipe de pesquisadores de três instituições de renome do Brasil, além de contar com membros de outras instituições nacionais e internacionais. O campo de atuação dos participantes cobre vários aspectos fundamentais da ciência da computação, como a modelagem e análise de comportamento de usuários, ciência da Web, sistemas paralelos e distribuídos, recuperação de informação, bancos de dados, mineração de dados, aprendizado de máquina, visão computacional e robótica, interações homem-máquina, otimização combinatória, privacidade e segurança.

A equipe proposta expande substancialmente o grupo de pesquisadores que desenvolveu o InWeb, agregando competências complementares para permitir abordar problemas muito mais difíceis e ambiciosos. Tais problemas vão além do mundo virtual, e incluem interações complexas e difusas entre esse mundo e o mundo físico.

5 Equipe

A equipe do INCT-MCS inclui pesquisadores de três instituições: a Universidade Federal de Minas Gerais (UFMG), a Universidade Federal do Amazonas (UFAM), e a Pontifícia Universidade Católica de Minas Gerais (PUC Minas). O ponto de partida dessa equipe foi o InWeb – o Instituto Nacional de Ciência e Tecnologia da Web, que começou em 2008 com um foco principal em serviços informacionais na Web. O InWeb alcançou excelentes resultados, não somente em termos de pesquisa, publicações e formação de recursos humanos, mas também em termos de transferência de tecnologia e relevância social. A transferência de tecnologia materializou-se em duas *startups*, Zunnit² e Neemu³, e no CTWeb – Centro de Tecnologia da Web⁴, uma iniciativa pioneira para auxílio ao aperfeiçoamento tecnológico e para servir de ponte entre a academia e a indústria. A relevância social dos resultados produzidos pelo InWeb pode ser ilustrada pelo Observatório da Dengue⁵, uma iniciativa que foi adotada pelo Ministério da Saúde, e pelo Portal CiênciaBrasil⁶, o qual vem sendo usado por agências de P&D.

Enfatiza-se que 27 pesquisadores da UFMG e da UFAM que participaram do InWeb são hoje parte da equipe do INCT-MCS, que também agrega cientistas de áreas como redes, robótica e outras disciplinas que lidam com aspectos ciberfísicos de ciência da computação. Salienta-se também que a equipe possui grande experiência com projetos de pesquisa (dos quais uma lista parcial aparece na Seção 20.2), e já publicou um número significativo de trabalhos em conferências e periódicos de prestígio relacionados às áreas abordadas pelo INCT-MCS. Por exemplo, 17 pesquisadores têm mais de 60 publicações indexadas na DBLP, uma das mais importantes e completas bibliotecas digitais da ciência da computação. As três instituições oferecem programas de pós-graduação em ciência da computação e muitos dos pesquisadores da UFAM e da PUC Minas obtiveram seus títulos de doutor pela UFMG. Assim, a UFMG vem, de certa forma, norteadando as atividades de pesquisa passada e presente do grupo, por pelo menos quarenta anos. A equipe também possui um impressionante histórico de interações com a indústria, o qual emerge via a criação de *startups* ou projetos de cooperação de pesquisa e desenvolvimento.

²<http://www.zunnit.com>

³<http://neemu.com>

⁴<http://www.ctweb.com.br>

⁵<http://www.observatorio.inweb.org.br/dengue>

⁶<http://www.cienciabrasil.org.br>

Em resumo, o grupo de pesquisa conta com 55 pesquisadores das três instituições participantes. Todos possuem título de doutor e 38 são bolsistas de produtividade do CNPq, sendo quatro de nível 1A, três nível 1B, um nível 1C, dez nível 1D e 20 nível 2. Além disso, sete pesquisadores são membros da Academia Brasileira de Ciências, dos quais três são membros permanentes e quatro são membros afiliados. Finalmente, a equipe conta com oito pesquisadores associados (totalizando 63 pesquisadores) de outras instituições brasileiras que têm integrado ou complementam o time de pesquisadores. Três desses pesquisadores são bolsistas de produtividade do CNPq no nível 1D. A Tabela 1 lista todos os participantes agrupados por instituição, incluindo seu nível de bolsa de produtividade em pesquisa, quando houver. A coluna mais à direita (*Tarefas*) inclui as tarefas atribuídas a cada pesquisador, conforme discutidas na Seção 13.2.

É importante ressaltar também que o INCT-MCS conta com a participação de vários pesquisadores associados internacionais, que mantêm colaborações ativas com membros do INCT-MCS, tendo provido cartas declarando sua intenção de participar ativamente no Instituto. Esses pesquisadores estão listados na Seção 19.1.

Tabela 1: Pesquisadores Participantes (pesquisadores marcados com ★ são membros da Academia Brasileira de Ciências).

ID	Pesquisador	Cargo	Bolsa CNPq	Tarefas
UFMG				
1	Alberto H. F. Laender★	Titular	PQ 1A	S.1 I.2 I.3 I.4 I.6 R.5 A.4
2	Geraldo Robson Mateus	Titular	PQ 1A	R.6 F.1 F.3 F.4
3	Nivio Ziviani★	Emérito	PQ 1A	S.1 I.4 I.5 R.1 R.2 R.5 A.4
4	Virgílio A. F. Almeida★	Titular	PQ 1A	S.1 I.1 R.5 R.7 R.9 R.10 R.11 A.2
5	Antonio A. F. Loureiro	Titular	PQ 1B	S.1 S.3 I.1 I.2 I.4 I.7 A.1–3 F.1–4
6	Gilberto M. Ribeiro	Titular	PQ 1B	S.2 S.5 I.7 R.10 F.6
7	Mario F. M. Campos	Titular	PQ 1B	S.4 I.1 I.7 R.2 A.2 A.3 A.5
8	Wagner Meira Júnior	Titular	PQ 1C	S.3 I.5 R.1 R.2 R.4 R.10 A.2 F.5 F.8 F.9
9	Arnaldo de A. Araújo	Associado	PQ 1D	I.1 I.5 R.2 R.8 R.9
10	José Marcos S. Nogueira	Titular	PQ 1D	S.3 I.7 A.1 A.2 F.1 F.2 F.3 F.6
11	Jussara M. Almeida★	Associado	PQ 1D	S.2 I.2 I.4 I.6 R.7 R.9 R.10 A.4
12	Marcos A. Gonçalves★	Associado	PQ 1D	I.2–4 I.6 R.2 R.4 R.5 R.7 R.8 R.10 A.4 A.6
13	Renato A. C. Ferreira	Associado	PQ 1D	S.3 S.4 S.5 I.5 R.4 F.4 F.5 F.9
14	Adriano A. Veloso★	Adjunto	PQ 2	S.1 S.2 S.3 S.4 I.1 R.3 A.1 A.6
15	Alexandre Cunha	Adjunto	PQ 2	R.6 F.4 F.8
16	Ana Paula Couto da Silva	Adjunto	PQ 2	S.2 R.7 R.9 R.10 A.4 A.6 F.8
17	Clodoveu A. Davis	Adjunto	PQ 2	S.2 I.3 I.4 I.5 I.6 R.9 R.10 A.2 A.6
18	Daniel F. Macedo	Adjunto	PQ 2	A.1 A.2 F.1 F.2 F.6
19	Dorgival O. Guedes Neto	Associado	PQ 2	S.3 S.4 R.4 F.1 F.2 F.3 F.6 F.9
20	Fabrizio B. Souza★	Adjunto	PQ 2	S.2 R.3 R.7 R.10 A.1
21	Fernando M. Q. Pereira	Adjunto	PQ 2	R.4 A.3 F.5 F.7
22	Gisele L. Pappa	Adjunto	PQ 2	S.2 I.4 I.6 R.1 R.7 R.8 R.9 F.8
23	Luis Chaimowicz	Associado	PQ 2	I.1 I.7 A.1 A.2 A.3 A.5
24	Luiz Filipe M. Vieira	Adjunto	PQ 2	A.1 A.2 F.1 F.2 F.3 F.5 F.6
25	Marcos A. M. Vieira	Adjunto	PQ 2	A.1 A.2 A.3 F.1 F.2 F.3 F.5 F.6
26	Mirella M. Moro	Adjunto	PQ 2	I.3 I.4 I.5 I.6
27	Raquel C. Melo-Minardi	Adjunto	PQ 2	R.3 R.5 R.7 R.8
28	Sebastián A. Urrutia	Adjunto	PQ 2	R.6 R.10 A.4 F.1
29	Thiago F. Noronha	Adjunto	PQ 2	R.6 R.10 F.1 F.2
30	William R. Schwartz	Adjunto	PQ 2	S.4 I.1 R.2 R.8 R.9
31	Adriano C. M. Pereira	Adjunto		S.1 S.2 S.4 I.4 I.6 R.1 R.8 R.9
32	Berthier Ribeiro-Neto	Associado		I.1 I.5 R.1 R.2 R.5 R.9 R.11
33	Douglas G. Macharet	Adjunto		I.1 R.3 R.9 A.1 A.3 A.5
34	Erickson R. do Nascimento	Adjunto		S.4 I.1 I.2 R.2 R.3 R.8 A.5 F.8
35	Ítalo F. S. Cunha	Adjunto		S.3 I.7 R.11 F.2 F.3 F.6
36	Jefersson A. dos Santos	Adjunto		S.4 I.1 I.5 R.1 R.4 R.5 R.8 R.9
37	Jeroen van de Graaf	Adjunto		S.1 S.5 I.3 F.7

38	Leonardo B. Oliveira	Adjunto		S.5 I.2 I.3 A.2 F.3 F.6 F.7
39	Loïc P. G. Cerf	Adjunto		R.1 R.2 R.4 R.5
40	Olga N. Goussevskaia	Adjunto		R.3 F.1 F.2 F.3
41	Pedro O. S. V. de Melo	Adjunto		S.5 I.4 R.7 R.9 R.10 A.1 F.8
42	Raquel O. Prates	Associado		S.5 R.3 R.7 R.10 A.1 A.2 A.6
43	Rodrygo L. T. Santos	Adjunto		S.1 I.2 I.3 R.5 R.8 A.4 F.8
UFAM				
44	Altigran S. da Silva	Associado	PQ 1D	S.1 I.1 I.2 I.3 I.4 I.6 R.9
45	Edleno S. de Moura	Associado	PQ 1D	S.1 I.5 R.5 A.4
46	Horácio A. B. F. Oliveira	Adjunto	PQ 2	A.3 A.4 F.3 F.4
47	Marco A. P. Cristo	Adjunto	PQ 2	R.1 R.5 R.8 R.9
48	André L. C. Carvalho	Adjunto		I.5 R.4 A.4 F.5
49	David B. F. de Oliveira	Adjunto		S.1 I.5 R.3 R.5
50	João M. B. Cavalcanti	Associado		I.5 I.6 R.5
51	Moisés G. de Carvalho	Adjunto		I.2 I.3 I.4 I.6 R.4 R.8
PUC Minas				
52	Raquel A. F. Mini	Adjunto	PQ 2	R.11 A.2 F.1 F.3 F.4
53	Fátima L. P. D. Figueiredo	Adjunto		R.11 A.2 F.1 F.2 F.3
54	Humberto T. Marques Neto	Adjunto		S.2 I.2 R.3 R.5 R.7 R.9 R.10 A.4
55	Wladimir C. Brandão	Adjunto		S.1 I.2 I.3 R.5 A.4
PESQUISADORES ASSOCIADOS				
56	Djamel Sadok (UFPE)	Associado	PQ 1D	S.1 S.3 I.1 I.2
57	Francisco Brasileiro (UFCE)	Associado	PQ 1D	S.3 S.4 R.4 F.1 A.1 A.2
58	Judith Kelner (UFPE)	Associado	PQ 1D	I.4 I.7 A.1–3 F.1–4
59	Alvaro R. Pereira Junior (UFOP)	Associado		S.1 I.4 I.5
60	Anderson A. Ferreira (UFOP)	Associado		I.2–4 I.6 R.2
61	Carlos Figueiredo (UEA, FUCAPI)	Adjunto		S.1 S.3 I.1
62	Fernando Parreiras (FACE/FUMEC)	Adjunto		I.2 I.3 I.4 R.5
63	Leonardo Rocha (UFSJ)	Adjunto		S.5 I.5 R.4

A equipe de pesquisadores do INCT-MCS consolida uma história de colaborações de sucesso entre seus membros. Essas colaborações podem ser visualizadas por meio da rede de coautoria dos membros do grupo pertencentes a cada uma das instituições participantes no Instituto, conforme mostrada na Figura 1, onde arestas representam pelo menos um artigo, nos últimos cinco anos, entre os pesquisadores conectados. Esse grafo mostra que os vários pesquisadores vêm desenvolvendo cooperações de sucesso, compartilhando e reusando conhecimento. Assim, o time já representa uma unidade coesa, com competências complementares, e potencial para continuar a produzir avanços científicos e tecnológicos significantes. A presente proposta formaliza essa coesão. Além disso, os pesquisadores associados (nacionais e internacionais) adicionam competências complementares ao grupo, ao mesmo tempo em que contribuem para aumentar sua visibilidade internacional.

6 Estrutura do Instituto

O INCT-MCS será coordenado pelo Departamento de Ciência da Computação da Universidade Federal de Minas Gerais. Sua estrutura organizacional é composta por um coordenador e um subcoordenador, um conselho gestor, e vários coordenadores de linhas de pesquisa.

O *Coordenador Geral* deverá organizar e coordenar todos os trabalhos relacionados ao Instituto. As atividades e tarefas sob sua coordenação são basicamente de três tipos: a gestão financeira do Instituto, a realização da pesquisa proposta e a gestão de relações do INCT-MCS com a comunidade. Além disso, o Coordenador Geral será o representante do INCT-MCS perante órgãos e entidades externas. O Coordenador Geral será diretamente auxiliado pelo *Subcoordenador*, que poderá representá-lo e tomar decisões em seu nome sempre que necessário.

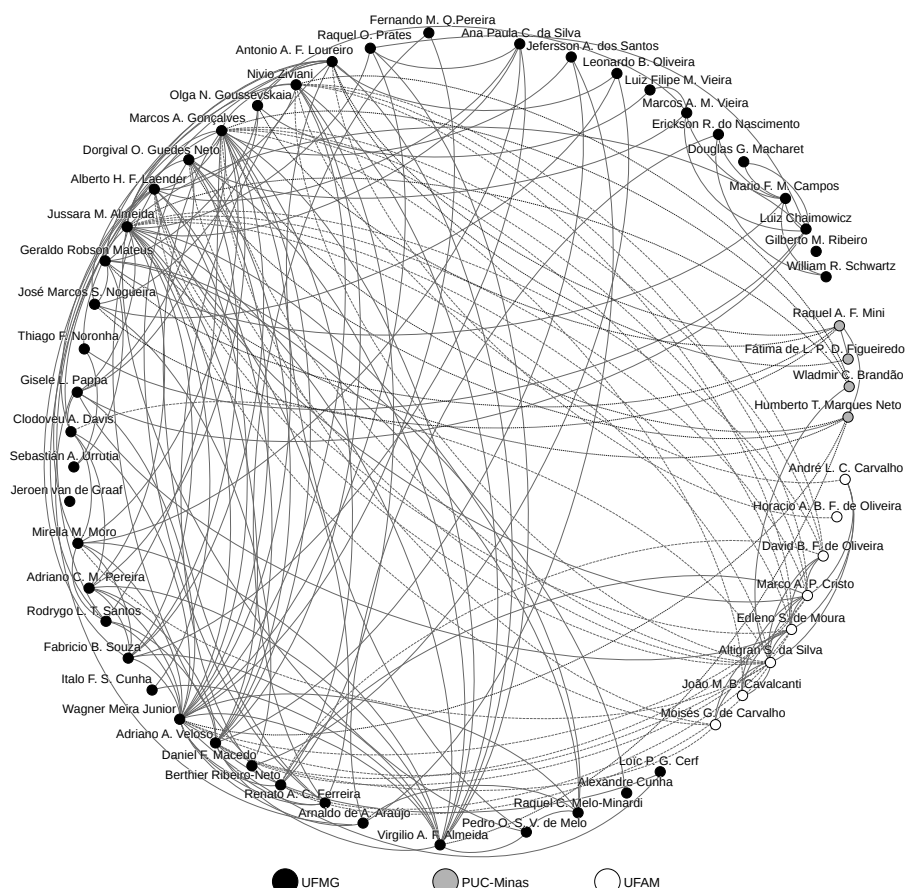


Figura 1: Rede de Colaboração dos Pesquisadores do INCT-MSC.

O *Conselho Gestor* auxiliará o Coordenador Geral na gestão administrativa, financeira e patrimonial do INCT-MCS. O Conselho Gestor será responsável pela supervisão e coordenação administrativa do Instituto, pelo planejamento e execução do orçamento, e pela implementação dos procedimentos financeiros, contábeis e patrimoniais. Será sua responsabilidade esclarecer e simplificar as ações de gestão do Instituto. Além disso, será de responsabilidade do Conselho Gestor a elaboração dos relatórios solicitados pelos órgãos de fomento e apoio externos ao Instituto, tais como Universidades, CNPq e Ministérios. O Conselho Gestor será composto por ao menos um representante de cada instituição participante, conforme listados na Seção 3, e será auxiliado por secretariado.

Os *Coordenadores de Linhas de Pesquisa*, apresentadas na Seção 10, coordenarão diretamente as atividades de pesquisa em cada linha proposta. Serão também responsáveis por decisões do dia-a-dia da pesquisa, como a seleção de bolsistas e a requisição de recursos computacionais, bibliográficos e de custeio para os laboratórios. Os coordenadores das linhas de pesquisa também responderão pelos recursos alocados às suas equipes e pela prospecção, identificação e seleção dos meios de divulgação científica. Os coordenadores reportarão seus resultados ao Coordenador Geral.

7 Instituições Participantes

As instituições que participam do INCT-MCS são UFMG, UFAM, e PUC Minas. Os pesquisadores dessas instituições irão realizar tarefas de pesquisa e desenvolvimento no contexto do projeto, tal como indicado nas cartas de apoio apresentadas. A integração entre as instituições será realizada através da participação de pesquisadores de diferentes instituições em linhas de pesquisa comuns, e de pesquisadores de uma mesma instituição em diferentes linhas. Além disso, propõe-se empregar vários mecanismos de interação, conforme descritos na Seção 8.

8 Mecanismos de Interação

Em projetos de longo prazo, como o proposto aqui, e especialmente quando os membros incluem instituições localizadas em diferentes estados e regiões do país (no nosso caso, Amazonas e Minas Gerais), é importante definir com cuidado especial a metodologia de implementação das atividades planejadas, de forma a assegurar o sucesso do projeto. Os membros do time atual já possuem um histórico de trabalharem juntos em diversos projetos anteriores e têm uma experiência positiva no que se refere à realização de projetos conjuntos. Essas conquistas passadas demonstram que a distância geográfica não representa um obstáculo à cooperação e realização deste projeto. Além disso, tecnologias correntes, tais como sistemas de e-mail e vídeo conferências como Skype, Google Hangouts, WebEx, e Adobe Connect, permitem encontros e discussões entre os parceiros mesmo estando fisicamente em lugares diferentes.

Os mecanismos de interação propostos incluem as seguintes atividades:

- Encontros técnicos. Reuniões de natureza essencialmente técnica, semanais ou quinzenais, com objetivo de avaliar o desenvolvimento das tarefas e discutir o progresso da pesquisa. Esses encontros irão ocorrer separadamente para cada linha de pesquisa, envolvendo os coordenadores de cada linha, assim como todos os pesquisadores e alunos participantes. Os encontros também poderão contar com a presença de membros de outras linhas de pesquisa, quando isso for necessário. Essas reuniões abordarão aspectos da pesquisa específicos de cada problema. Sempre que for possível e considerado necessário, participantes localizados em diferentes locais geográficos irão participar remotamente via vídeo conferência.
- Encontros de progresso. Reuniões dos coordenadores de linhas de pesquisa, que irão ocorrer mensalmente, e cujo objetivo será avaliar o desenvolvimento geral do projeto, corrigir quaisquer distorções, discutir ações para manter as tarefas no cronograma e estabelecer conexões entre as diferentes linhas. Quando necessário, vídeo conferências serão utilizadas.
- Encontros de revisão. Reuniões realizadas anualmente (em frequência maior, se for o caso), que contarão com a presença física de todos os membros do Instituto, pesquisadores e estudantes, com o objetivo de apresentar e discutir os resultados da pesquisa conduzida, avaliar o desenvolvimento do projeto e fazer ajustes quando necessário.
- Workshops e seminários. Eventos abertos à comunidade, com o objetivo de apresentar resultados parciais do projeto. Esses eventos servirão também como mecanismos de interação com empresas interessadas em parcerias com o Instituto com o intuito de transformar os resultados de pesquisa em tecnologias e produtos.

Este projeto de pesquisa é dividido entre várias linhas de pesquisa, cada uma composta por vários desafios e tarefas correspondentes, como descrito detalhadamente nas Seções 10, 12 and 13. Cada linha possui um coordenador, cujas responsabilidades são descritas na Seção 6, assim como um time de pesquisadores e alunos. As tarefas serão divididas por linha de pesquisa e não por instituição, o que irá contribuir para a integração de pesquisadores de diferentes instituições. Além disso, um mesmo pesquisador poderá participar de várias linhas de pesquisa, o que facilitará a troca de informações entre os participantes do projeto.

Nós também planejamos estimular visitas técnicas envolvendo membros do Instituto e outros pesquisadores internacionais com quem nosso grupo já colabora (veja a lista na Seção 19). O objetivo dessas visitas será promover trabalhos de pesquisa conjuntos e troca de experiências com grupos internacionais trabalhando em áreas relacionadas às pesquisadas no Instituto. Publicações em conferências de alta qualidade serão encorajadas, em particular quando associadas a visitas técnicas a esses grupos internacionais. Dado que conferências são o principal veículo de publicação em Ciência da Computação, a participação nesses eventos é muito importante para continuar expandindo a visibilidade internacional do grupo.

9 Programa de Educação

As instituições participando deste projeto possuem programas de pós-graduação credenciados pela CAPES nos níveis 7 (UFMG), 5 (UFAM) e 4 (PUC Minas). Portanto, pode-se esperar que um número significativo de alunos de doutorado, mestrado e graduação seja formado como resultado da criação deste Instituto. Ao entrarem no mercado de trabalho, esses estudantes vão disseminar o conhecimento adquirido neste projeto e desenvolver serviços e aplicações que beneficiarão diretamente a sociedade, gerando resultados interessantes em muitos setores da economia. Na verdade, a formação desses profissionais e pesquisadores pode ser considerado um dos principais resultados deste Instituto.

O INCT-MCS também agregará jovens pesquisadores recém-doutores que atualmente recebem bolsas de pós-doutorado. Esta participação é uma excelente forma de integrá-los ao ambiente cooperativo de instituições de pesquisa, facilitando sua interação com novos grupos, oferecendo-lhes recursos e oportunidade de orientar alunos em tópicos avançados, assim alavancando sua carreiras e estimulando a sua permanência nessas instituições.

Os resultados científicos produzidos no contexto do INCT-MCS poderão ser incorporados aos currículos de graduação e pós-graduação dos cursos oferecidos pelas instituições participantes. Além disso, a experiência adquirida durante o desenvolvimento deste projeto poderá beneficiar disciplinas já ministradas, em áreas tais como Redes Complexas, Recuperação de Informação, Gerenciamento de Dados da Web, Mineração de Dados, Sistemas Distribuídos, Visão Computacional, Robótica, entre outros.

Finalmente, o INCT-MCS também planeja desenvolver programas de residências, com tópicos adaptados para necessidades do mercado. Nesses programas, profissionais de computação irão receber treinamento em áreas avançadas de competência deste Instituto, com oportunidades de participação em atividades de pesquisa e desenvolvimento. Esses profissionais poderão aplicar o conhecimento adquirido e se preparar para satisfazer a demanda de empresas de tecnologia da computação.

10 Linhas de Pesquisa

A fim de apoiar uma sociedade massivamente conectada, a agenda de pesquisa do INCT-MCS irá lidar com problemas específicos relacionados com os seguintes objetivos gerais: sensoriamento, integração, análise com base em modelagem e predição de sistemas e comportamentos complexos, bem como atuação. Estes problemas, que serão mais detalhados na Seção 12, exigem expertises complementares, sendo agrupados em oito linhas específicas de pesquisa, como resumido na Figura 2. Cada arco colorido representa uma linha de pesquisa, com o seu comprimento abrangendo os objetivos que a linha vai cobrir. Note que todas as faixas de pesquisa estão associadas a pelo menos dois dos quatro objetivos supracitados, ao passo que cada objetivo é coberto por pelo menos quatro linhas de pesquisa. Uma alta integração entre os objetivos e as linhas de pesquisa é fundamental para o sucesso deste projeto. Além disso, todo o trabalho desenvolvido em cada linha será guiado pelos seis aspectos descritos no centro da Figura 2, ou seja, adaptabilidade, performance, privacidade, confiabilidade, escalabilidade e segurança, que se relacionam com as propriedades desejadas bem como com as preocupações envolvidas na concepção de soluções para uma sociedade massivamente conectada.

10.1 Comunicação Ubíqua

Numa sociedade massivamente conectada existe um grande fluxo de informação entre agentes sociais e os mundos físicos e virtuais. Os dados originados do mundo físico são coletados do ambiente por dispositivos integrados na rede. Estes dados, associados com dados gerados pelos agentes sociais conectados no mundo virtual, geram ações nos ambientes físicos e virtuais. Neste sentido, esta linha de pesquisa tem como objetivo investigar protocolos de comunicação, algoritmos, modelos e dispositivos para o desenvolvimento e integração de redes heterogêneas de grande escala, composto por objetos/coisas/sensores/dispositivos que coletam dados e agem sobre o ambiente físico. Entre os desafios de pesquisa, existem problemas de escalabilidade para coletar grandes quantidades de dados geograficamente distribuídos; uso eficiente do meio físico de comunicação; tomada de decisão e atuação no

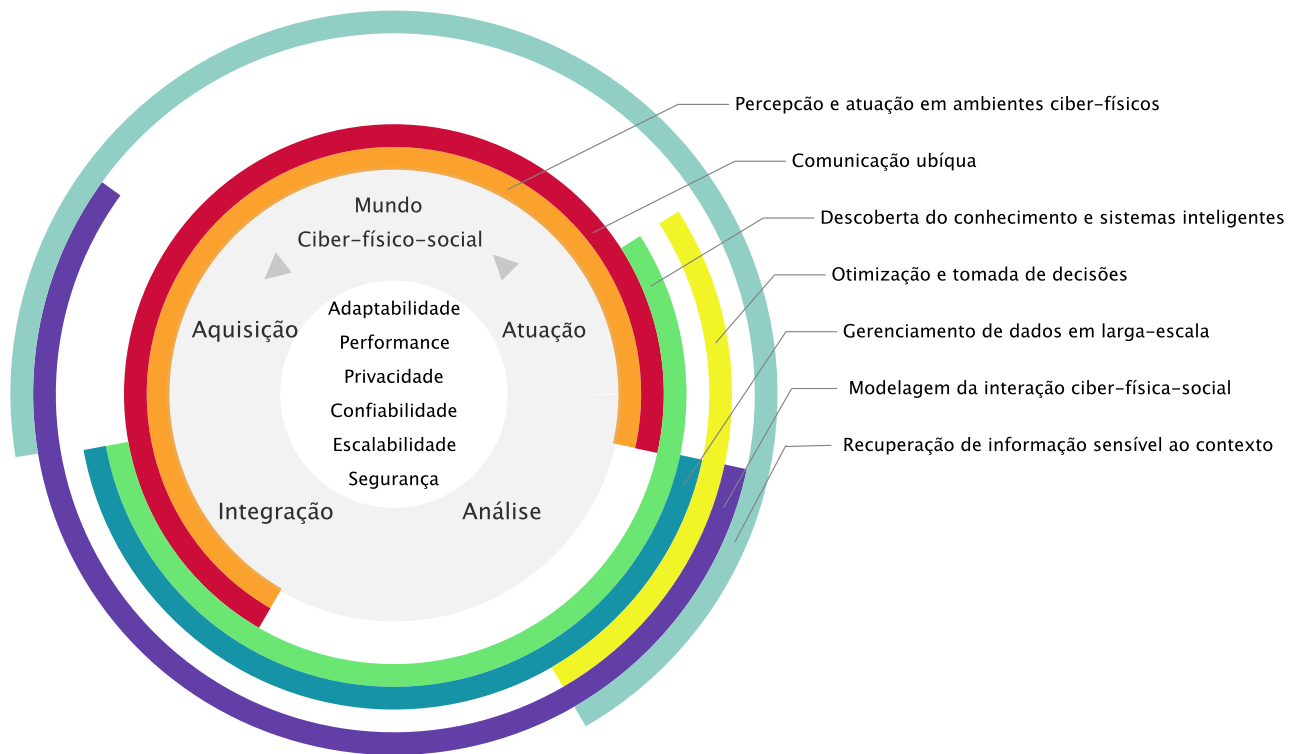


Figura 2: Linhas de Pesquisa do INCT-MCS.

ambiente; identificação, segurança e privacidade; operação e gerenciamento eficiente e inteligente de grandes redes heterogêneas dinâmicas e sujeitas a falhas frequentes.

10.2 Gerenciamento de Dados em Larga Escala

Em um mundo massivamente conectado, a quantidade e variedade de fontes de dados vem continuamente crescendo de forma impressionante. A Internet vem servindo como um repositório para a maioria destes dados, que por sua vez podem ser totalmente estruturados ou ser completamente desestruturados. A utilidade de grandes volumes de dados heterogêneos não está limitada a exploração de cada fonte individualmente. Há muito a ser ganho através da integração de diversas grandes fontes de dados na busca por informação. Todos os tipos de dados coletados do meio ambiente através de sensores, câmeras e dispositivos de computação móvel, ou fornecidos pelos cidadãos, em iniciativas de *crowdsourcing* ou e-ciência, devem ser integrados com dados e informações que podem ser obtidas na Internet a partir de fontes como conjuntos de dados estruturados, documentos e redes sociais. A pesquisa no gerenciamento de dados altamente conectados em larga em escala e multi-fonte busca a resposta de inúmeros desafios que são relacionados à coleta, sanitização, armazenamento, integração e análise de dados que se originam de várias fontes ciber-físicas e sociais.

10.3 Modelagem da Interação Ciberfísica-social

Outra característica de um mundo massivamente conectado está relacionada com as várias interações entre um grande número de entidades heterogêneas, como usuários, aplicações e elementos físicos. Isso leva ao surgimento de várias redes sociais, que se comportam de acordo com padrões que variam ao longo de várias escalas de tempo e através de vários tipos de relacionamentos. O entendimento e a modelagem de tais padrões leva ao conhecimento prático e teórico sobre o comportamento e a evolução de grupos sociais e das suas interações com o mundo

cibernético. Em termos mais gerais, a análise e a modelagem de como as pessoas interagem com aplicativos virtuais e umas com as outras através de tais aplicativos e de elementos físicos podem levar a conhecimentos fundamentais que direcionem o projeto de aplicações e serviços que melhor atendem às necessidades e expectativas dos usuários. A elaboração de modelos parcimoniosos para descrever e prever o comportamento individual e coletivo (ou social) das pessoas ao interagir com aplicações virtuais e com elementos físicos (por exemplo, sensores e robôs) exige novas técnicas de análise e simulação que lidam com desafios relacionados com a amostragem e representatividade de dados, integração de múltiplas fontes de evidência do comportamento humano, níveis heterogêneos de abstração e de qualidade, bem como a reação rápida às mudanças do mundo real.

10.4 Recuperação de Informação Sensível ao Contexto

Em um mundo massivamente conectado espera-se um aumento sem precedentes na produção e consumo de informação através de múltiplos canais. Enquanto sistemas modernos de recuperação de informação têm lidado razoavelmente com a escala da informação produzida, muito trabalho ainda tem que ser feito para satisfazer plenamente as necessidades dos usuários de informação relevante em um mundo cada vez mais inundado de dados. Em particular, os usuários contemporâneos têm necessidades complexas de informação, que são transmitidas não apenas por uma consulta submetida a uma máquina de busca ou um perfil de preferências registrado em um sistema de recomendação, mas também pelo contexto em que essas necessidades são manifestadas. Por exemplo, os sinais contextuais podem incluir dados socioeconômicos dos usuários, localização, condições físicas e mentais, e seus círculos sociais. Por sua vez, esses sinais contextuais podem ser derivados de uma infinidade de diferentes fontes, incluindo a Web, redes sociais online, sensores ambientais e sensores acoplados a dispositivos móveis. A pesquisa em recuperação de informação sensível ao contexto visa explorar sinais contextuais arbitrários e inter-relacionados para a produção de máquinas de busca e de recomendação mais eficazes.

10.5 Descoberta de Conhecimento e Sistemas Inteligentes

Um mundo massivamente conectado produz fluxos de dados enormes, incompletos, ruidosos, heterogêneos e assíncronos, tanto de fontes cibernéticas como de fontes físicas. Extrair conhecimento e fazer previsões a partir desses dados exige novos modelos, algoritmos e plataformas computacionais que devem se adaptar às diferentes demandas e trabalhar de forma integrada. Os modelos a serem projetados apoiarão as tarefas de análise e previsão em uma sociedade maciçamente conectada, enquanto os algoritmos tornarão tais modelos aplicáveis na prática. As plataformas computacionais que serão construídas para executar os algoritmos propostos avançarão o estado-da-arte em termos de abstrações, modelos de programação e linguagens, suporte em tempo de execução, usando hardware de baixo custo.

10.6 Otimização e Tomada de Decisão

Problemas de logística envolvem o gerenciamento do fluxo de pessoas e mercadorias entre centros de oferta e demanda, limitado por vários tipos de restrições. Naturalmente, em um mundo massivamente conectado, tais problemas se apresentam em inúmeras situações, uma vez que a mobilidade é uma característica fundamental da nossa sociedade. Assim, o objetivo desta linha de pesquisa é modelar problemas de logística no âmbito das tecnologias que levam em conta a forma como as pessoas e os objetos interligados em rede interagem uns com os outros nos mundos físico e virtual. O desafio implica em alcançar soluções otimizadas através do uso de algoritmos clássicos e extensões para incorporar características operacionais complexas, com especial atenção para aspectos como escalabilidade, dinamicidade, estocasticidade, adaptabilidade e desempenho.

10.7 Percepção e Atuação em Ambientes Ciberfísicos

Como mencionado anteriormente, um mundo massivamente conectado gera uma quantidade enorme de dados incertos, ruidosos, incompletos e frequentemente pouco confiáveis a partir de sensores físicos e virtuais. Uma atuação e interação segura e confiável tanto com os seres humanos quanto com o ambiente exige a integração

e filtragem eficientes desses dados. A busca de soluções para um desafio tão fará avançar os limites do estado-da-arte em Visão Computacional e Robótica, para apoiar as tarefas a serem realizadas pela parte de atuação do sistema pelos (possivelmente heterogêneos) agentes autônomos (humanos, robôs e outros atuadores) em ambientes dinâmicos e complexos.

11 Qualificação e Justificativa

A adoção explosiva de dispositivos físicos e aplicações Web, incluindo redes sociais online, junto com o acesso a recursos computacionais na nuvem estão transformando o jeito que as pessoas trabalham, aprendem, se divertem e se comunicam. Essa transformação é dirigida pela quantidade e pela qualidade das conexões de rede hoje disponíveis, que atualmente permitem que bilhões de pessoas interajam entre si e com dispositivos eletrônicos. Nessa sociedade massivamente conectada, o desenvolvimento de sistemas ciberfísicos e sociais pode diminuir o custo de vida das pessoas e aumentar a sustentabilidade das formas de geração e consumo de recursos.

Sistemas ciberfísicos e sociais são sistemas projetados que dependem da integração perfeita de algoritmos computacionais e componentes físicos que, por outro lado, são também usados de forma integrada. Progressos em tais sistemas permitirão avanços na capacidade, adaptabilidade, escalabilidade, resiliência, segurança e usabilidade dos mesmos, que excederão em muito as mesmas características que existem nos sistemas atuais do mundo físico e cibernético. Novas tecnologias ciberfísicas transformarão o jeito que pessoas interagem com sistemas projetados – da mesma forma que a Internet transformou o jeito que pessoas interagem com a informação. Essa transformação trará inovações e fomentará a competição saudável nos mais variados setores da sociedade produtiva, como aqueles relacionados à mobilidade urbana, à segurança pública, à agricultura, à geração de energia, à manufatura de produtos de alta tecnologia e à saúde.

No entanto, a construção de uma arquitetura que integre sistemas tão avançados é um desafio por muitas razões. Por exemplo, tal arquitetura deve integrar vários componentes heterogêneos. Um componente chave é a Web, atualmente ubíqua na vida das pessoas. Frequentemente as pessoas usam as aplicações Web para realizar diversas atividades, de compras online e transações bancárias a interação e socialização com amigos e familiares, compartilhamento de experiências, troca de opiniões e de conhecimento. Outros componentes importantes dessa arquitetura são os vários tipos de sensores ambientais e humanos (sensores embutidos em vestimentas, carros, câmeras, etc), atuadores (robôs) e outros dispositivos móveis conectados (smartphones e tablets) que estão cada vez mais presentes em nossa sociedade.

A integração de componentes heterogêneos em sistemas robustos que melhorem a qualidade de vida das pessoas envolve operações com elementos complexos, os quais são conectados em diferentes escalas e suportados por tecnologias diversas (redes cabeadas ou sem-fio e redes ad-hoc). Tais elementos, por seu vez, geram massivas quantidades de dados em diferentes níveis de abstração e qualidade, em múltiplas escalas temporais e espaciais, e o fazem em tempo real. A tarefa de lidar com tais características, sem comprometer a privacidade dos usuários, impõe desafios éticos, políticos, sociais, científicos, tecnológicos e até de engenharia. As oportunidades para a construção de soluções científicas e tecnológicas inovadoras para enfrentar esses desafios nos motivam a criar o INCT-MCS.

12 Objetivos, Metas e Indicadores

O objetivo deste projeto é prover suporte para uma sociedade massivamente conectada. Para tanto, o projeto buscará avançar o estado da arte em técnicas de sensoriamento, integração, análise e atuação. Como mencionado, a pesquisa irá abordar o problema de como coletar dados eficazmente de ambientes físicos e virtuais sem comprometer a segurança e privacidade. O objetivo é integrar uma grande quantidade de dados ruidosos de diversas fontes em informações valiosas, e sintetizar as informações em modelos práticos que permitam compreender sistemas e comportamentos complexos. Essa melhor compreensão permitirá desenvolver mecanismos de atuação para prestar serviços de alto valor para a sociedade. Fecha-se o ciclo a partir da coleta de feedbacks para refinar o sensoriamento, a integração, análise e os mecanismos de atuação. A Figura 3 resume como os quatro objetivos principais abor-

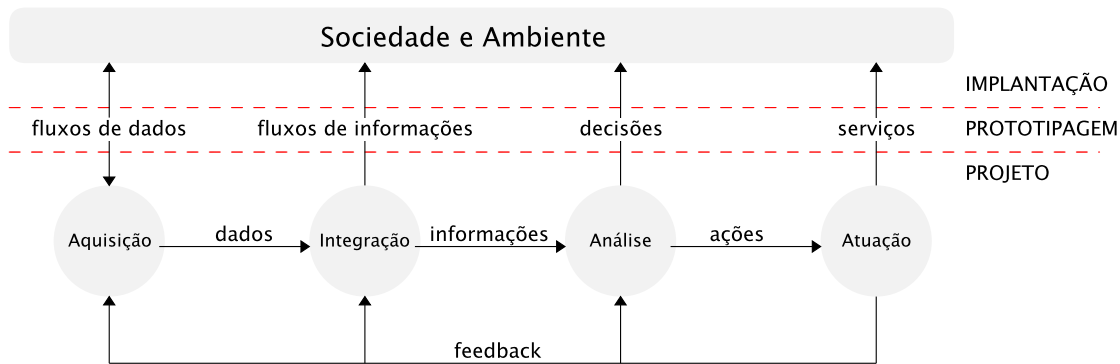


Figura 3: Visão Geral dos Principais Objetivos do INCT-MCS.

dados neste projeto estão interligados. A seguir, cada objetivo é descrito, juntamente com os desafios específicos relacionados ao mesmo.

12.1 Sensoriamento

O sensoriamento (ou coleta de dados) tem por objetivo obter, de forma simples e contínua, amostras de múltiplas fontes de informação que vão desde sistemas naturais, tais como a Floresta Amazônica, até sistemas ciberfísicos, como as grandes cidades. Amostras de dados podem ser obtidas de fontes dinâmicas e heterogêneas, incluindo sensores físicos (câmeras, GPS, sensores de distância e movimento, etc), sensores virtuais (coletores da Web, bancos de dados públicos, monitores de dados online, etc) e sensores humanos (dados voluntários, ciência do cidadão, iniciativas de participação pública, etc). Além do contínuo crescimento da Web e da explosão de redes sociais online, o barateamento e a modernização do hardware de sensoriamento tem levado a um crescimento sem precedentes na quantidade de fluxos de dados disponíveis em tempo real [1]. Nesse cenário, a monitoração eficiente de volumes tão grandes de informação é um problema aberto.

A meta final desta fase é desenvolver mecanismos eficientes para a observação do mundo físico como um repositório de informações sujeito a mudanças contínuas. Grandes desafios estão atrelados a essa meta, tais como:

- Como projetar sistemas de coleta de dados que lidem, de forma eficiente, com os compromissos entre a representatividade da informação obtida e o custo, em termos de energia, espaço, latência e financeiro dos mecanismos usados para obtê-la?
- Que mecanismos devem ser usados ou desenvolvidos para coletar informação proveniente de fluxos de dados muito grandes, ruidosos e sujeitos a erros, levando em consideração, restrições de segurança e privacidade?
- Como permitir e encorajar usuários a compartilhar informação e assegurar a sua privacidade, de modo que dados representativos e não tendenciosos possam ser obtidos?

12.2 Integração

Na fase de integração de dados, o objetivo principal é projetar algoritmos e estruturas de dados que processem e combinem diferentes tipos de dados—temperatura, localização, umidade, velocidade, opiniões—a diferentes níveis de abstração—texto, imagens, vídeos, gestos e ações—para extrair informação útil. Para tal objetivo seja alcançado, fazem-se necessários algoritmos que lidem com fluxos de dados massivos e que disponham de operações como agregação, filtragem e indexação em tempo (quase) real. Integração é, portanto, uma fase crítica do projeto proposto, pois a informação constitui os alicerces sobre os quais modelos e mecanismos de atuação serão construídos.

As tarefas e problemas discutidos nesta seção suscitam vários desafios de pesquisa, alguns dos quais são listados a seguir:

- (a) Como integrar múltiplas fontes de dados, heterogêneas e complexas, em diferentes níveis de abstração?
- (b) Como projetar algoritmos que sejam capazes de armazenar, agregar, filtrar e indexar os dados coletados de forma eficiente?
- (c) Como avaliar a qualidade da informação derivada dos dados agregados?
- (d) Como alcançar os três objetivos anteriores preservando a privacidade dos indivíduos?

12.3 Análise

Outro objetivo fundamental é projetar novos modelos para descrição e predição de comportamento individual e coletivo em cenários que podem ser dinâmicos, complexos e heterogêneos. Esses modelos devem ser capazes de receber grandes quantidades de informação como entrada, informação essa proveniente de múltiplas fontes e sujeita a vários níveis de incerteza e redundância. É intenção desta proposta redefinir o estado-da-arte nos campos de computação natural, otimização, aprendizado de máquina e recuperação de informação, a fim de que possa ser plenamente explorada e aproveitada a informação proveniente de múltiplas fontes, disponível em diferentes formatos e existente em diferentes níveis de qualidade. A ideia é que esses modelos possam direcionar tomadas de decisão e atuação eficientes. Nesse contexto, os maiores desafios são:

- (a) Como desenvolver modelos parcimoniosos, porém gerais e precisos, considerando-se a grande quantidade de características humanas e ambientais que precisam ser modeladas?
- (b) Como construir modelos eficientes que forneçam respostas em tempo (quase) real? Como adaptar esses modelos para lidar com mudanças contextuais em tempo (quase) real?
- (c) Como validar os modelos propostos, uma vez que eles podem ser alimentados com dados incompletos e muitas vezes ruidosos?

12.4 Atuação

O último objetivo é o projeto e implantação de mecanismos de atuação capazes de interagir com a sociedade e com o ambiente. O principal desafio desta fase é o risco de desastres. Uma decisão errada tomada por um dispositivo que controla veículos autônomos ou robôs, que aloca policiais, ambulâncias ou bombeiros, que controla o trânsito, ou guia decisões governamentais, pode comprometer seriamente o bem-estar social, o ambiente e a segurança pública. Mecanismos de atuação precisam ser confiáveis, pois seus serviços podem ser requisitados a qualquer momento sob condições adversas. Esses mecanismos devem também se integrar com o ambiente e com a sociedade de forma suave, bem como serem capazes de aprender e se adaptar continuamente.

A fim de desenvolver mecanismos de atuação que sejam confiáveis, resistentes e não intrusivos, pretende-se responder às seguintes questões:

- (a) Como construir uma vasta infra-estrutura de custo efetivo que seja capaz de prover coleta de dados, computação, comunicação e armazenamento em situações adversas, preservando a segurança e privacidade?
- (b) É possível o desenvolvimento de arcabouços de teste que permitam a quantificação das limitações dos mecanismos de atuação, a fim de que sejam identificados cenários e ambientes em que tais mecanismos sejam confiáveis?
- (c) Como isolar o impacto de diferentes mecanismos de atuação, coletar *feedback*, e rapidamente adaptar tais mecanismos para lidar com mudanças no comportamento de seus usuários ou do ambiente em que são utilizados?

13 Metodologia

Este projeto será desenvolvido em três etapas, correspondentes às fases de projeto, protótipo e implementação, como ilustrado na Figura 3. Na primeira etapa, grupos trabalhando com os ambientes físicos e virtuais vão trocar ideias, preparar os seus dados e infraestrutura, bem como desenvolver modelos e aplicações que permitem a integração desses dois mundos. Na etapa seguinte, as iniciativas ligadas aos mundos físico e virtual se tornarão fortemente conectadas, com as atividades de pesquisa focando em protótipos de tecnologias baseados em modelos e aplicações desenvolvidas na fase anterior, identificando também novos desafios. A última fase envolve a identificação de cenários de aplicação interessantes em que essa integração possa ser benéfica e demonstre a eficácia das nossas soluções através de sua implantação em uma configuração de maior escala.

A metodologia proposta para o desenvolvimento deste projeto, que pode ser revista em cada uma das três fases acima referidas, consiste de todos (ou um subconjunto de) os seguintes passos:

1. Revisão da literatura: levantar a bibliografia sobre métodos e modelos de ponta para resolver ou modelar o problema sob análise.
2. Definição de fontes de dados: identificar as fontes de dados de interesse e definir quais dados devem ser coletados.
3. Aquisição de dados: construir métodos para coletar, armazenar e indexar fluxos de dados dinâmicos e heterogêneos, provenientes de várias fontes em diferentes cenários.
4. Caracterização de dados: entender os dados coletados, sua representatividade, incerteza intrínseca e sua qualidade.
5. Integração de dados: integrar dados do mundo físico ou virtual, preservando suas características individuais, como qualidade e privacidade.
6. Desenvolvimento de novos modelos descritivos e/ou preditivos de comportamento individual ou coletivo, no mundo físico e/ou virtual.
7. Avaliação de modelos: entender e propor métodos apropriados para avaliar modelos que lidam com dados incompletos ou ruidosos.
8. Desenvolvimento de infraestrutura: desenvolver infraestrutura de comunicação de custo eficiente e protocolos que integrem redes de comunicação heterogêneas.
9. Atuação: desenvolver algoritmos e métodos que possibilitem que agentes físicos e virtuais coordenem e atuem sobre o ambiente ciberfísico.
10. Desenvolvimento de protótipos: desenvolver um protótipo que demonstre à sociedade os benefícios de integrar a informação proveniente de ambientes virtuais e físicos.

Esses 10 passos são ainda divididos em tarefas, as quais são descritas na tabela 2

13.1 Tarefas

A Tabela 2 apresenta as principais tarefas que serão realizadas para alcançar os objetivos apresentados na Seção 12. Agrupamos as tarefas seguindo os principais objetivos. Estas tarefas podem estar relacionadas a várias linhas de pesquisa (em diferentes domínios). As tarefas do grupo de infraestrutura irão dar suporte a outras tarefas, por exemplo, provendo armazenamento compartilhado, computação escalável, e comunicação confiável. A coluna mais à esquerda na tabela apresenta os indicadores das tarefas que serão utilizadas no restante do documento para referenciar as tarefas.

Tabela 2: Descrição das Tarefas a Serem Realizadas no INCT-MCS.

ID	DESCRIÇÃO
SENSORIAMENTO	
S.1	Desenvolver algoritmos robustos e eficientes para coletar dados da Web e de sensores acoplados a seres humanos ou instalados no ambiente. Esses algoritmos devem levar em conta o consumo de energia, a taxa de amostragem, a temporalidade e a privacidade dos dados coletados.
S.2	Desenvolver técnicas efetivas de amostragem para a coleta de conjuntos de dados representativos.
S.3	Desenvolver arquiteturas distribuídas para a coleta de dados em larga escala e em tempo (quase) real.
S.4	Analisar a disponibilidade temporal e espacial de diferentes fluxos de dados.
S.5	Desenvolver mecanismos e técnicas para encorajar usuários a compartilhar dados sem colocar em risco a sua segurança e privacidade.
INTEGRAÇÃO	
I.1	Projetar métodos para integrar fontes de dados que produzem informação em diferentes escalas de tempo e níveis de abstração. Esses dados vão de informações simples, como temperatura e umidade, até informações complexas, como vídeos e fotos.
I.2	Desenvolver mecanismos para emparelhar registros de entidades individuais como usuários, carros, telefones, e rastreá-los ao longo de diferentes domínios, preservando a sua privacidade.
I.3	Desenvolver mecanismos para alinhar esquemas complexos de uma mesma entidade.
I.4	Desenvolver mecanismos para produzir visões agregadas de múltiplas entidades que sejam produzidas por filtros sensíveis a informações contextuais, tais como tempo, localização e aspectos culturais.
I.5	Desenvolver mecanismos escaláveis para armazenar, indexar e recuperar dados de forma eficiente.
I.6	Projetar novas métricas e metodologias para medir o nível de incerteza dos dados coletados e a qualidade da informação derivada a partir deles.
I.7	Projetar protocolos de coordenação para melhorar a qualidade dos dados e evitar medições redundantes.
ANÁLISE	
R.1	Projetar novos algoritmos para descobrir de forma eficiente, em conjuntos de dados complexos, as regiões que contêm conhecimento específico, como os padrões que caracterizam o comportamento de subcomunidades físicas ou virtuais.
R.2	Projetar novos modelos e algoritmos que descubram ou expliquem com eficiência a organização global de dados complexos, tais como classes de comportamento, e que possam classificar essas entidades.
R.3	Projetar técnicas de visualização expressivas e robustas para dados e modelos que suportem tarefas tais como análise, predição e atuação física ou virtual.
R.4	Investigar a paralelização, distribuição e integração dos algoritmos usados na mineração de dados, na computação natural e no aprendizado de máquina.
R.5	Investigar modelos e mecanismos que possam inferir quão relevantes são partes da informação para usuários, dadas restrições de contexto.
R.6	Desenvolver novas otimizações e modelos de simulação que melhorem o processo de tomada de decisão, levando-se em conta aspectos relacionados à heterogeneidade, dinamicidade e escalabilidade.
R.7	Propor metodologias de classificação que possibilitem o entendimento de comportamentos individuais e coletivos, a partir de dados ruidosos, heterogêneos e incompletos.
R.8	Projetar técnicas de seleção de características que levem em conta como variáveis diferentes impactam a acurácia dos modelos gerados.
R.9	Desenvolver modelos precisos e parcimoniosos que descrevam e possam prever o comportamento de indivíduos e grupos, tais como mobilidade e interesses de pessoas. Esses modelos devem aprender com e adaptar-se a respostas provenientes de usuários e do ambiente.
R.10	Realizar simulações e experimentos no mundo físico para validar os modelos de comportamento propostos. Verificar a sensibilidade desses modelos à parametrização.
R.11	Construir um arcabouço que permita entender o impacto de informação incompleta e obsoleta sobre a modelagem de sistemas complexos como tráfego veicular e mobilidade de usuários.
ATUAÇÃO	
A.1	Construir e instalar um estudo de caso de tamanho médio para prover uma validação inicial das soluções propostas. Um estudo de caso razoável pode ser levado adiante no campus de uma universidade.
A.2	Construir e instalar um estudo de caso de grande escala para testar exaustivamente as soluções propostas no mundo concreto, dado um grande volume de entidades heterogêneas interconectadas, tais como usuários, sensores e agentes eletrônicos. Considera-se aqui um estudo de caso em uma cidade.

- A.3 Investigar e desenvolver mecanismos de coordenação escaláveis, descentralizados e robustos, que suportem grande quantidade de agentes como sensores e robôs cooperando entre si e realizando tarefas complexas, ainda que somente informação local esteja disponível.
- A.4 Desenvolver aplicações robustas e cientes de privacidade, capazes de recomendar ações aos seus usuários, a fim de melhorar a qualidade de suas rotinas em termos de mobilidade, segurança pública e preservação do ambiente.
- A.5 Desenvolver aplicações que usem interações inovadoras para melhorar a acessibilidade e imersão em ambientes virtuais tais como os encontrados nas interações entre robôs e humanos, ou em salas de cirurgia.
- A.6 Investigar mecanismos não intrusivos para encorajar, coletar e filtrar respostas de usuários.

INFRAESTRUTURA

- F.1 Desenvolver mecanismos de alocação de recursos, políticas de enfileiramento e algoritmos de controle de tráfego que possam se comunicar de forma efetiva sobre redes de diferentes tecnologias.
- F.2 Desenvolver protocolos de comunicação capazes de conectar redes com suporte parcial a IP e dispositivos sem acesso a uma pilha completa de rede, como câmeras e sensores.
- F.3 Instalar uma infraestrutura de comunicação eficiente para integrar redes heterogêneas, que incluam redes locais, redes globais e redes formadas por aparelhos móveis, tais como redes de sensores e redes veiculares.
- F.4 Propor mecanismos flexíveis e extensíveis capazes de desonerar dispositivos de baixo poder computacional, tais como sensores de baixo custo e telefones de pouca autonomia energética, a fim de suportar aplicações complexas que usem tais dispositivos.
- F.5 Projetar e implementar arcabouços de computação elásticos o suficiente para suportar arquiteturas heterogêneas que incluem máquinas de múltiplos núcleos, placas gráficas e processadores de sinais.
- F.6 Projetar métodos e ferramentas para monitorar e controlar a comunicação, o armazenamento da infraestrutura de serviços desenvolvida durante esse projeto.
- F.7 Projetar técnicas de análise estática e dinâmica e técnicas criptográficas para assegurar a privacidade e a integridade de dados transmitidos por redes heterogêneas e seguras.
- F.8 Construir e manter repositórios atualizados para todos os dados coletados e para todos os componentes de software produzidos durante esse projeto.
- F.9 Projetar uma arquitetura flexível e modular para integrar o software produzido, a fim de facilitar a implementação e instalação de serviços e mecanismos de atuação.

13.2 Cronograma

A Tabela 3 mostra uma estimativa do tempo que será dedicado a cada tarefa. Tons mais escuros indicam maior esforço despendido pela equipe de pesquisadores. Quando apropriado, rótulos são usados para descrever a evolução de tarefas que demoram mais tempo. Nota-se que as tarefas A.1 e A.2 requerem a integração de resultados de diversas outras atividades, e serão completadas no quarto e no sexto ano de pesquisa, respectivamente. A tabela também apresenta os pesquisadores associados com cada tarefa (conforme números vistos na tabela 1).

Tabela 3: Cronograma de Tarefas e Pesquisadores Participantes.

Tarefa ID	Semestre												Participantes
	1	2	3	4	5	6	7	8	9	10	11	12	
S.1													1 3 4 5 14 30 36 42 44 45 49 55
S.2													6 11 14 16 17 20 22 30 54
S.3													5 8 10 13 14 19 34
S.4													7 13 14 19 29 30 33 35
S.5													6 13 36 37 40 41
I.1													4 5 7 9 14 23 29 31 32 33 35 44
I.2													1 5 11 12 33 37 42 44 51 54 55
I.3													1 12 17 26 36 37 42 44 51 55
I.4													1 3 5 11 12 17 22 26 30 40 44 51
I.5													3 8 9 13 17 26 31 35 45 48 49 50
I.6													1 11 12 17 22 26 30 44 50 51

14 Cenário Atual e Resultados Esperados

14.1 Cenário Atual

14.1.1 Sensoriamento

No campo da coleta de dados online, vários problemas abertos ainda existem, a despeito da maturidade dos sistemas de coleta de informação na Web [5]. Em particular, a coleta de dados na Web oculta (ou *deep Web*) ainda mostra-se incipiente, tanto mais quando se leva em consideração a necessidade de obtenção de dados a partir de sistemas fechados, como as redes sociais online. De fato, a obtenção eficiente de amostras não tendenciosas, considerando-se o acesso limitado a essas redes, é um problema em aberto [12]. Um problema igualmente desafiador é a alocação de recursos de coleta, mantendo-se um compromisso entre cobertura e atualidade para suportar análises em tempo real [26]. Do ponto de vista do sensoriamento ambiental, escalabilidade e resistência a falhas são desafios chave. Considerando-se o problema de escala, a sempre crescente quantidade de dados coletada por sensores demandará novas soluções para o processamento de grandes fluxos de dados [2]. Da mesma forma, um melhor entendimento

de quais dados irão impactar os modelos obtidos subsequentemente é chave para uma amostragem eficaz [22]. Considerando-se a resistência a falhas desses sistemas, existe uma grande necessidade de desenvolvimento de mecanismos de coordenação mais avançados, que sejam capazes de lidar com a bateria limitada dos sensores e com as fontes de transmissão de dados muitas vezes heterogêneas [1]. Finalmente, os avanços na tecnologia de sensores permitem que hoje seja coletada uma infinidade de dados acerca das atividades diárias de uma pessoa [14]. Neste sentido, ainda é um desafio o desenvolvimento de técnicas que permitam o compartilhamento de dados, ao mesmo tempo que assegurem a privacidade do indivíduo [11].

14.1.2 Integração

O problema da integração de dados é atualmente estudado por diversos pesquisadores que trabalham nos mais variados campos da ciência da computação. Pesquisadores que estudam grandes volumes de dados (“Big Data”) lidam com desafios relacionados a fontes de informação que são numerosas, heterogêneas e autônomas. Essas fontes de dados interagem de forma complexa e crescem continuamente [31]. Devido a esse padrão de comportamento, massas de dados frequentemente exibem comportamento de cauda pesada e inter-dependências não triviais [17]. Assim, um dos desafios que serão atacados neste projeto é gerar os metadados corretos para descrever a informação armazenada [16]. Outro desafio é realizar a agregação de dados de forma online e escalável, ao invés de utilizar o pós-processamento [25]. Ao mesmo tempo, é vital que seja avaliada a qualidade dos dados coletados, a fim de que conjuntos de informação cada vez mais limpos possam ser construídos [24].

De maneira geral, a alta disponibilidade de dados originou vários problemas de pesquisa relacionados à integração da informação. Por exemplo, considere-se o problema de identificar entidades (usuários, lugares, etc) entre múltiplos domínios [15, 20] Essa identificação torna possível um entendimento profundo de traços psicológicos, que não somente são essenciais para várias disciplinas científicas, mas também são fundamentais em sistemas de recomendação online [33]. Em várias aplicações da vida real, tais como aquelas voltadas para *crowd-sourcing* ou computação móvel, indivíduos compartilham seus dados a fim de receber serviços específicos de todo o sistema. Nesse contexto, pesquisadores estão hoje preocupados com o compromisso entre a disponibilidade de dados e a sua privacidade [17]. Dados pessoais integrados podem ser usados para melhor entender a sociedade [28] e mesmo para salvar vidas [10]. Entretanto, esses dados pessoais podem revelar informações privadas de forma indesejada, ou mesmo ilegal [8]. Sabe-se que a integração de dados com manutenção de privacidade é possível em teoria [7], mas a carga computacional das técnicas existentes hoje é muito alta, exigindo a proposta de novas soluções mais eficientes.

14.1.3 Análise

A crescente disponibilidade de poder computacional e o surgimento de ferramentas cada vez mais expressivas possibilita hoje a coleta de dados que quantificam de forma confiável a complexidade de sistemas cibernéticos. Modelos computacionais dirigidos a dados emergiram como ferramentas apropriadas para estudar diversos fenômenos, que vão desde o aparecimento de epidemias [13] até a propagação de informação em redes sociais [21]. Esses modelos têm por meta prover os fundamentos sobre os quais podem ser entendidas as tendências emergentes e as propriedades não-lineares que frequentemente caracterizam os sistemas cibernéticos [30]. Em um sentido amplo, podemos categorizar os modelos dirigidos a dados como descritivos e preditivos [32]. Enquanto os primeiros permitem descrever dados de uma forma que torne aparente as tendências sociais [29], padrões [4, 27], e situações excepcionais, os últimos permitem prever o comportamento tanto de indivíduos quanto de sociedades e tomar decisões [6]. Modelagem descritiva e preditiva são frequentemente usadas em conjunto. Um modelo descritivo lida, de maneira quantitativa, com os principais características dos dados, enquanto um modelo preditivo é, em sua essência, uma relação matemática entre essas mesmas características e uma certa variável alvo. Este projeto pretende desenvolver modelos que combinam dados coletados de múltiplas fontes heterogêneas, que frequentemente possuem diferentes níveis de abstração. Neste cenário, identificar o menor grupo de variáveis de entrada que são suficientes para gerar modelos precisos é uma tarefa particularmente desafiadora, dado que neste contexto é comum lidar com um grande conjunto de variáveis, que frequentemente não são correlacionadas e podem ter ruído, inconsistência e prover somente uma visão incompleta do sistema em estudo.

14.1.4 Atuação

O desenvolvimento de metodologias de coleta e atuação é um dos problemas mais fundamentais que os agentes precisam superar a fim de que possam executar, com sucesso, tarefas de forma cooperativa. Em cenários do mundo real, agentes mecânicos, físicos e biológicos devem lidar com a incerteza e o ruído presente em dados adquiridos por diferentes tipos de sensores. Interações intra-agentes podem ocorrer somente entre dispositivos autônomos de natureza eletromecânica (como robôs) [19], exigindo estratégias de coordenação seguras e confiáveis. Essas interações também podem ocorrer entre agentes heterogêneos, tais como entre seres humanos, robôs e o ambiente [23]. Nesse caso, tem-se problemas ainda mais desafiadores, pois a dinâmica de estruturas orgânicas e ambientais é difícil de modelar e prever. Uma questão importante relacionada à coordenação de problemas nesta linha de pesquisa é a correta extração e representação de informações ambientais [9], bem como a comunicação entre vários dispositivos e agentes que executam aplicações em diferentes contextos [3], que requerem novas infraestruturas de comunicação [18].

14.2 Resultados Qualitativos

Os esforços de pesquisa neste projeto serão direcionados para módulos independentes e complementares dentro de uma arquitetura de comunicação. A coordenação prevista pela arquitetura permite que pesquisa básica, projeto, testes, protótipos, integração e implantação possam ser continuamente realizados pela equipe e que possam gradualmente ser integrados à medida que o projeto avançar.

Nos dois primeiros anos, componentes serão desenvolvidos e testados em um ambiente de integração flexível, onde a experimentação científica com as idéias e conceitos elaborados por cada um possam ser validados. A integração então prosseguirá durante toda a duração do projeto e de forma gradual levará a protótipos que serão implantados e testados em pequena escala. À medida que o projeto avançar, esperamos que protótipos evoluam para produtos integrados capazes de serem implantados em larga escala e que sejam capazes de abordar de forma eficaz problemas atuais da sociedade.

A integração de componentes permitirá que o projeto atinja um grande número de domínios de aplicação. Por exemplo, em relação à mobilidade urbana, módulos de coleta de dados podem reunir dados brutos a cerca do tráfego a partir de sensores embarcados em smartphones e/ou veículos. Além disso, tais dados podem ser integrados e armazenados para fins de análises estatísticas e espaço-temporais, dando suporte ao estudo de problemas de mobilidade. Assim, uma ferramenta de mineração de dados será capaz de buscar padrões e eventos, anomalias, distúrbios, avarias e incidentes correlacionados com eventos monitorados por outros sensores. Em outras palavras, será possível gerar retorno por meios físicos e virtuais à sociedade, tais como alertas, previsões, alarmes, mensagens aos usuários de redes sociais, mensagens em locais públicos, reprogramação de semáforos ou comandos enviados a unidades de atendimento de emergências ou a unidades policiais. Os mesmos componentes poderão ser integrados para outras aplicações em particular.

14.3 Resultados Quantitativos

O projeto em questão resultará em uma longa lista de resultados científicos e tecnológicos. Esses resultados foram quantificados e são exibidos na Tabela 4.

14.4 Produtos

A seguir, são apresentados os produtos a serem entregues por biênio.

Anos 1 e 2:

1. Adquirir e implantar elementos de infraestrutura de computação, sensores e equipamentos de rede;
2. Contratar serviços de computação em nuvem;
3. Criar um Web site para o Instituto;

Tabela 4: Resultados Quantitativos do INCT-MCS.

Descrição	Quantidade
Arquitetura geral	1
Biblioteca de componentes e módulos	1
Know-how ou patentes	10
Portal web	1
Artigos em conferências	95 por ano = 570
Artigos em periódicos	31 por ano = 186
Cursos/Seminários	1 por ano = 6
Dissertações de mestrado	48 por ano = 288
Escolas de Verão	1 por ano = 6
Teses de doutorado	12 por ano = 72
Visitas técnicas	5 por ano = 30
Residentes (após ano 1)	6 por ano = 30

4. Desenvolver a primeira versão do componente geral e arquitetura de comunicações;
5. Estabelecer as primeiras iniciativas de coleta e armazenamento da Web e de plataformas de sensores de dados;
6. Desenvolver a primeira versão da biblioteca de componentes de software e módulos, incluindo componentes para a coleta de dados, armazenamento, indexação, processamento de fluxo, mineração de dados, aprendizado de máquina, e outros;
7. Escrever artigos científicos para publicação de resultados em eventos internacionais de qualidade;
8. Organizar dois workshops anuais, reunindo todos os membros da equipe, incluindo estudantes, para discutir, avaliar e integrar os principais resultados dos esforços de pesquisa e decidir sobre eventuais adaptações e novas direções;
9. Propiciar dez visitas técnicas de pesquisadores pertencentes ao Instituto a pesquisadores colaboradores internacionais, visando ampliar a visibilidade dos resultados alcançados pelo projeto, bem como a geração de novos resultados correlacionados ao projeto;
10. Promover dois cursos/seminários a serem lecionados pelos colaboradores internacionais aos alunos relacionados as atividades do Instituto. Esta é uma oportunidade para o intercâmbio científico com os pesquisadores do Instituto;
11. Organizar um programa de residência em temas da indústria relacionados com as principais áreas do INCT-MCS, envolvendo cerca de 6 residentes, com duração de um mês, começando no segundo ano do Instituto.

Anos 3 e 4:

1. Adquirir os elementos complementares de infraestrutura de hardware;
2. Contratar serviços de implantação e manutenção do sensores;
3. Estender o desenvolvimento de módulos de software e componentes, com ênfase na integração e geração de protótipos funcionais;
4. Desenvolver a primeira versão da arquitetura geral de componentes e comunicação;
5. Escrever artigos científicos para publicação de resultados em eventos internacionais de qualidade;
6. Enviar pedidos de patentes com base em resultados de pesquisa;

7. Organizar dois workshops anuais, reunindo todos os membros da equipe, incluindo estudantes, para discutir, avaliar e integrar os principais resultados do esforço de pesquisa e decidir sobre eventuais adaptações e novas direções;
8. Organizar workshops envolvendo autoridades do governo e representantes da indústria para comunicar resultados e discutir a cooperação para transferência de tecnologia e implantação em cenários de grande escala;
9. Propiciar dez visitas técnicas de pesquisadores pertencentes ao Instituto aos pesquisadores colaboradores internacionais, visando ampliar a visibilidade dos resultados alcançados pelo projeto, bem como a geração de novos resultados correlacionados ao projeto;
10. Promover dois cursos/seminários a serem lecionados pelos colaboradores internacionais aos alunos relacionados às atividades do Instituto. Esta é uma oportunidade para o intercâmbio científico com os pesquisadores do Instituto.
11. Organizar um programa de residência em temas da indústria relacionados com as principais áreas do INCT-MCS, envolvendo cerca de 6 residentes, com a duração de um mês.

Anos 5 e 6:

1. Consolidar componentes, módulos e protótipos em produtos, voltados para cenários de aplicação previamente definidos;
2. Implantar soluções e produtos em grande escala, monitorando seu uso e impacto social;
3. Manter a biblioteca de componentes de software;
4. Divulgar resultados, protótipos e produtos, a fim de obter repercussão social dos resultados do projeto;
5. Escrever artigos científicos para publicação de resultados em eventos internacionais de renomados;
6. Enviar pedidos de patentes com base em resultados de pesquisa;
7. Organizar dois workshops anuais, reunindo todos os membros da equipe, incluindo estudantes, para discutir, avaliar e integrar os principais resultados do esforço de pesquisa e decidir sobre eventuais adaptações e novas direções.
8. Propiciar dez visitas técnicas de pesquisadores pertencentes ao Instituto aos pesquisadores colaboradores internacionais, visando ampliar a visibilidade dos resultados alcançados pelo projeto, bem como a realização de novos resultados correlacionados ao projeto.
9. Promover dois cursos/seminários a serem lecionados pelos colaboradores internacionais aos alunos relacionados às atividades do Instituto. Esta é uma oportunidade para o intercâmbio científico com os pesquisadores do Instituto.
10. Organizar um programa de residência em temas da indústria relacionados com as principais áreas do INCT-MCS, envolvendo cerca de 6 residentes, com duração de um mês.

15 Potencial de Transferência Tecnológica

Este projeto tem um grande potencial para produção de patentes, *know-hows*, protótipos e produtos tecnológicos, considerando tanto as conquistas do InWeb e as experiências prévias dos pesquisadores participantes do projeto, como o inegável apelo econômico a empresas do setor de TI ou empresas que incorporam hardware e software relacionados a TI aos seus produtos. Em particular, devemos mencionar que o escritório de P&D do Google na América Latina foi originalmente baseado na aquisição, em 2005, da Akwan, uma startup criada na UFMG em 1999. O escritório do Google, localizado em Belo Horizonte, no momento, emprega mais de 150 engenheiros, e mais da metade dos mesmos são formados na UFMG. Além disso, o InWeb resultou em duas startups operacionais: Zunnit e Neemu. Zunnit representou um modelo pioneiro de transferência no Brasil, onde uma universidade pública federal, a UFMG, recebeu participação acionária como compensação pela transferência de tecnologia. Finalmente,

o grupo inovou recentemente mais uma vez ao criar o CTWeb – Centro de Tecnologia para a Web, cujo objetivo é melhorar, gerar e transferir tecnologia, assim como desenvolver novos produtos a partir dos resultados obtidos no InWeb. O CTWeb materializa uma nova sistematização do processo de inovação baseado em resultados de pesquisa, valorizando tanto o papel da academia quanto o da indústria ao longo do processo.

Os objetivos de pesquisa e tecnologia do INCT-MCS já atraíram o interesse de várias companhias e organizações. As atividades de caracterização e modelagem do comportamento e interação de usuários têm um enorme potencial de agregar valor a uma ampla gama de serviços sendo desenvolvidos para a Web. Por exemplo, as principais fontes de renda das maiores máquinas de busca, como Google e Yahoo, são serviços de anúncios, recomendação e moderação, motivando o seu interesse formal nos resultados do INCT-MCS. Provedores de Internet, conteúdo e serviços, como o Universo Online (UOL), também têm enorme interesse em incorporar conteúdo de anúncios provido pelos mesmos de forma a aumentar suas receitas. UFMG e UFAL têm um relacionamento duradouro de pesquisa e tecnologia com UOL Inc., como manifestado pelo seu interesse formal nos resultados do INCT-MCS. Também vale mencionar que Google, Yahoo e UOL já apoiaram a submissão do InWeb em 2008, e desde então os laços têm se fortalecido em várias dimensões, incluindo transferência de tecnologia.

Além disso, o grupo tem procurado parcerias para transferência de tecnologia em outras vertentes do INCT-MCS. CEITEC, um grande centro brasileiro de projeto de circuitos integrados, manifestou seu interesse em redes, sensores e aplicativos a serem desenvolvidos. Fiat, a montadora de automóveis, quer acompanhar de perto os resultados relacionados a redes veiculares. Outras duas companhias que provêm serviços de TI em cenários ciberfísicos e sociais também manifestaram o seu interesse formalmente: Way Taxi e Módulo GRC. Way Taxi explora o novo mercado de transporte cliente-para-provedor, enquanto Módulo é um provedor global líder de GRC (Governança, Risco e Cumprimento) e Soluções Inteligentes de Governo. Módulo já estabeleceu uma parceria com o CTWeb para transferência de tecnologia e tem a intenção de fortalecer e ampliar este processo. Finalmente, há diversas pequenas empresas de TI, representadas pela Fumsoft, que também manifestaram o seu interesse nas atividades do INCT-MCS.

Como mencionado anteriormente, o grupo possui mecanismos pioneiros e diversificados para transferir a tecnologia desenvolvida, contando com o apoio da CTIT – Coordenadoria para Transferência e Inovação Tecnológica da UFMG, que ajuda a transferir para a sociedade tecnologias inovadoras criadas dentro da universidade. CTIT provê apoio legal e administrativo e define políticas de propriedade intelectual. Essas políticas garantem que a propriedade intelectual seja assegurada nos interesses do participante do projeto que a tenha criado.

Como estratégia, nós pretendemos seguir uma política agressiva de código aberto, para a maioria mas não todos os resultados, em que a chave do sucesso para um impacto duradouro é uma rápida e pervasiva disseminação de protótipos para uma ampla comunidade de pesquisadores.

Neste contexto, nós vemos como potenciais parceiros não apenas companhias estabelecidas que podem se beneficiar diretamente dos resultados de pesquisa, mas também novos investidores que possam criar novas empresas baseadas nos resultados de pesquisa do Instituto, assim como já ocorreu anteriormente,

É importante ainda ressaltar que o domínio das tecnologias derivadas deste projeto é de extrema importância para o desenvolvimento tecnológico e social do País, tendo em vista o papel fundamental que sistemas computacionais conectados à Internet desempenham hoje no mundo. A parceria entre o InWeb, o INCT da Dengue e o Ministério da Saúde em torno do Observatório da Dengue inovou no monitoramento epidemiológico e ajudou milhões de brasileiros.

Finalmente, o domínio dessa tecnologia coloca o nosso País com uma vantagem competitiva aos nossos vizinhos sul-americanos e a outros países em desenvolvimento, como a China e a Índia, que hoje despontam como centros de desenvolvimento de software e serviços de tecnologia da informação.

16 Transferência de Conhecimento

Este projeto possui diversas iniciativas que visam o alcance da sociedade como um todo. Entre elas, podemos citar as Escolas de Verão e os cursos de treinamento para escolas técnicas e de segundo grau. Nas Escolas de Verão, os estudantes serão introduzidos, ao longo de algumas semanas do ano, à Ciência da Computação como um todo, e aos

tópicos de pesquisa descritos nesta proposta. O objetivo é despertar interesse na área entre os jovens, assim como identificar e captar talentos que não seriam descobertos de outra forma. Com o mesmo intuito, o Instituto poderá eventualmente patrocinar a organização de Feiras de Computação, no mesmo estilo das já reconhecidas Feiras de Ciências.

Outra iniciativa de transferência de conhecimento é baseada em protótipos, como já implementado pelos portais Observatório da Web e CiênciaBrasil no contexto do InWeb. Esses protótipos, além de demonstrar a efetividade e inovação das pesquisas do Instituto, provêm para a população como um todo informações de valor agregado e serviços relevantes usualmente não disponíveis.

17 Orçamento

Esta seção apresenta o orçamento do projeto (Tabela 6), o qual está dividido em despesas operacionais, capital fixo e bolsas. Cada tipo de despesa é detalhado nas próximas seções.

17.1 Despesas Operacionais

As despesas operacionais compreendem custos com viagens (diárias e passagens), material de reposição e serviços de terceiros, totalizando R\$ 4.925.999,90.

Diárias: R\$ 1.476.000,00

Descrição: As diárias requisitadas são para estadias nacionais e internacionais. As diárias internacionais servirão principalmente para cobrir visitas técnicas de membros associados ao INCT-MCS, para cobrir visitas técnicas de convidados do exterior no Brasil, e para custear a participação de membros do INCT em conferências de alto impacto. As diárias nacionais irão suportar interações entre os grupos que constituem o INCT, e permitirão que seus membros participem de simpósios nacionais.

Justificativa: Visitas técnicas e apresentação de artigos em congressos técnicos são mecanismos chave para o aprendizado e para a divulgação de resultados.

Passagens aéreas: R\$ 1.140.000,00

Descrição: As passagens aéreas solicitadas são tanto para viagens nacionais quanto internacionais. As viagens internacionais permitirão que os pesquisadores do INCT possam participar de visitas técnicas, que pesquisadores estrangeiros possam vir para o Brasil, e que resultados de pesquisa possam ser apresentados em congressos. As passagens nacionais irão suportar interações entre os grupos que constituem o INCT, e permitirão que seus membros participem de simpósios nacionais.

Justificativa: Visitas técnicas e apresentação de artigos em congressos técnicos são mecanismos chave para o aprendizado e para a divulgação de resultados.

Material de reposição nacional: R\$ 109.000,00

Descrição: Além de material para escritório (papel, tóner, canetas, etc), distinguem-se dois tipos de material que serão adquiridos: material de reposição e material relacionado à Internet das Coisas (ou *Internet of Things*). Material de reposição inclui discos, memória, teclados, mouses, etc. Material relacionado à Internet das Coisas inclui sensores, RFIDs, adaptadores, baterias e interface de comunicação, entre outras partes necessárias aos várias experimentos e implantações planejados.

Justificativa: Os materiais de escritório são necessários para as atividades diárias do projeto, o que inclui a escrita e revisão de artigos. As partes de reposição são chave para a manutenção dos equipamentos de pesquisa, e o material de sensores é necessário para a realização de experimentos preditos nos estudos de caso. Esse material é também necessário para o desenvolvimento de software e para a manutenção dos dispositivos embarcados usados no projeto.

Material de reposição importado: R\$ 190.677,97

Descrição: Todo material de reposição importado é relacionado à Internet das Coisas, o que inclui sensores, RFIDs, adaptadores, baterias, antenas, etc. Esse material é necessário para a realização dos experimentos que o

projeto prevê.

Justificativa: Esse material vai suportar os experimentos que usam dados coletados de usuários reais que carregam telefones celulares, por exemplo. O material é também importante para o desenvolvimento de software e para a manutenção de dispositivos embarcados. Finalmente, o material é útil para a construção de redes de acesso experimentais.

Serviços de Terceiros: R\$ 2.010.321,93

Descrição: Serviços de terceiros irão apoiar o pessoal de comunicação e de gestão de mídia, a execução de experimentos, a organização e participação de eventos, serviços na nuvem, a melhoria da infraestrutura, e o desenvolvimento de hardware e software, além de taxas operacionais e adicionais do projeto.

Justificativa: Os serviços de terceiros proporcionarão a força de trabalho especializada que falta à equipe do projeto, permitirão a execução profissional de atividades de interação como eventos organizados pelo INCT-MCS ou não, e também apoiarão atividades de pesquisa (por exemplo, serviços de nuvem). Despesas operacionais serão pagas à FUNDEP, fundação da UFMG, pela gestão do projeto (5 % do orçamento) e as taxas adicionais são o custo de importação de equipamentos, peças e suprimentos (18 %).

17.2 Bens de Capital

Os bens de capital compreendem equipamentos, livros, e a licença de softwares, e resulta em R\$ 1.418.990,00.

Equipamento Nacional: R\$ 953.434,44

Descrição: Os equipamentos nacionais a serem adquiridos podem ser divididos em três categorias: de uso geral, Internet das Coisas, e robótica. A seguir serão descritas brevemente cada uma das categorias. Os equipamentos de uso geral são compostos de servidores, armazenadores, laptops, desktops, dispositivos móveis, impressoras e aparelhos de ar condicionado. Uma variedade de equipamentos relacionados com a Internet das Coisas serão adquiridos, em especial os dispositivos de aquisição e transmissão de dados, leitores de RFID, de infraestrutura para redes veiculares, equipamentos para laboratório de apoio e rede de acesso experimental. Diferentes tipos de plataformas robóticas serão adquiridas, tais como um manipulador holonômico móvel (Kuka youbot), humanóides de pequena escala (NAO), dispositivos de interação humano-robô (por exemplo, Jibo, Duplo Telepresença, Google Glass) e um enxame de robôs móveis simples.

Justificativa: Esses equipamentos serão utilizados para criar os vários laboratórios do INCT-MCS. Os equipamentos relacionados com a Internet das Coisas vão apoiar experimentos usando dados coletados de usuários reais que usam celulares e etiquetas RFID, bem como redes veiculares. Esses equipamentos também são necessários para o desenvolvimento de software e manutenção de dispositivos embarcados, bem como a construção de uma rede de acesso para os experimentos. Dispositivos robóticos serão integrados com diferentes sensores, a fim de permitir a percepção e atuação no mundo real, tanto com os seres humanos como com outros sistemas ciberfísicos. Os experimentos serão realizados tanto no ambiente de laboratório como em condições reais. Portanto, o laboratório deve ser equipado com infraestrutura de captura de movimento (por exemplo, rastreador Objeto/Quadrotor e rastreador Face/expressão) para possibilitar a construção de ambientes de teste realistas.

Equipamento Importado: R\$ 355.555,56

Descrição: Os equipamentos importados a serem adquiridos podem ser divididos em duas categorias: Internet das Coisas e robótica. Vários equipamentos relacionados com a Internet das Coisas serão adquiridos, em especial os dispositivos de aquisição e transmissão de dados, leitores de RFID, de infraestrutura para redes veiculares, equipamentos para dar suporte a laboratórios e redes experimentais. Diferentes tipos de plataformas robóticas serão adquiridas, tais como um manipulador holonômico móvel (Kuka youbot), humanóides de pequena escala (NAO), dispositivos de interação humano-robô (por exemplo, Jibo, Duplo Telepresença, Google Glass) e um enxame de robôs móveis simples. Serão adquiridos também sensores de laser (por exemplo, Hokuyo), câmeras e outros sensores visuais que irão equipar os robôs, equiparão os seres humanos e o meio ambiente. Novos sensores e atuadores serão desenvolvidas com o auxílio de um dispositivo de impressão 3D.

Justificativa: O equipamento para Internet das Coisas será necessário para a realização dos experimentos usando os dados coletados a partir de usuários reais utilizando celulares e etiquetas RFID bem como redes veiculares.

Tabela 5: Bolsas Solicitadas.

Tipo	Código	#	Valor	Valor	Total
			Mensal	Anual	Total
Graduação	IC	20	400,00	96.000,00	576.000,00
Mestrado	GM	12	1.500,00	216.000,00	1.296.000,00
Doutorado	GD	6	2.594,00	186.768,00	1.120.608,00
TecDev Doc	DTI A	1	4.000,00	48.000,00	288.000,00
TecDev Mas	DTI B	1	3.000,00	36.000,00	216.000,00
TechDev Bac	DTI C	2	1.100,00	26.400,00	158.400,00
Total		44		609.168,00	3.655.008,00

Esses equipamentos também são necessários para o desenvolvimento de software, manutenção de dispositivos embarcados e construção das redes que serão usadas em experimentos. Dispositivos robóticos serão integrados com diferentes sensores, a fim de habilitar percepção e atuação no mundo real, seja com sistemas de sensores usados por seres humanos, seja com outros sistemas ciberfísicos. Experimentos serão realizados em ambiente de laboratório e em condições reais. Assim, os laboratórios de pesquisa deverão estar equipados com infraestrutura para captura de movimentos que venha a prover confiança na construção de conjuntos de dados.

Livros: R\$ 10.000,00

Descrição: livros relacionados aos tópicos do projeto.

Justificativa: livros de referência que suportem e fundamentem o desenvolvimento de pesquisa e tecnologia.

Licenças de Software: R\$ 100.000,00

Descrição: Serão adquiridas licenças de software para o apoio do desenvolvimento de hardware (Altium e IAR) e funcionamento do instituto em termos de gerência, comunicação, visualização, e publicação.

Justificativa: As licenças para desenvolvimento de hardware são necessários para a construção do hardware que serão utilizados nos experimentos. As outras licenças são necessárias para a produtividade da equipe e gestão profissional e comunicação.

17.3 Bolsas

Requisita-se as bolsas discriminadas na tabela 5 para suportar as atividades de pesquisa e desenvolvimento tecnológico. As bolsas de pesquisa irão manter alunos de doutorado, mestrado e iniciação científica. Esses alunos estarão matriculados nos programas de graduação e pós-graduação das instituições que integram o INCT-MCS. As bolsas de desenvolvimento tecnológico (DTI A, DTI B e DTI C) financiarão a equipe responsável por desenvolver e implantar os protótipos.

18 Bolsas Requisitadas a Agências Estaduais de Fomento

Nenhuma bolsa será solicitada à Fapemig.

19 Parcerias Institucionais

Esta seção apresenta as colaborações acadêmicas com pesquisadores de instituições internacionais, bem como as interações com outros INCTs.

Tabela 6: Resumo do Orçamento.

Tipo	Item	Quantidade (R\$)
Despesas operacionais	Diárias	1.476.000,00
	Passagens	1.140.000,00
	Material de reposição nacional	109.000,00
	Material de reposição importado	190.677,97
	Serviços de terceiros	2.010.321,93
Capital	Equipamento nacional	953.434,44
	Equipamento importado	355.555,56
	Livros	10.000,00
	Licenças de software	100.000,00
Bolsas		3.655.008,00
Total		9.999.997,90

19.1 Pesquisadores Associados Internacionais

A equipe mantém parcerias acadêmicas com pesquisadores de diversas instituições internacionais que trabalham em áreas relacionadas com os nossos objetivos e metas. A Tabela 7 apresenta algumas das cooperações em andamento. Nomes marcados com um asterisco identificam os pesquisadores que já têm projetos formais de cooperação internacional em curso, como também bolsas de pesquisador visitante com a equipe deste projeto. Todos esses pesquisadores assinaram cartas de intenção para participar do INCT-MCS como pesquisadores associados, as quais estão em poder do Coordenador do Instituto.

Tabela 7: Pesquisadores Internacionais Associados.

Nome	Instituição
Alberto Elfes	CSIRO, Austrália
Aline Carneiro Viana*	INRIA Saclay, França
Amedeo Napoli*	Loria, França
Azzedine Boukerche	University of Ottawa, Canadá
Benjamin Bustos*	Universidad de Chile, Chile
Bernard Gendron	Université de Montreal, Canadá
Carlos Castillo	QCRI, Catar
Chedy Raïssi *	INRIA, França
Christophe Diot	Technicolor, França
Christos Faloutsos	Carnegie-Mellon University, EUA
Craig Macdonald	University of Glasgow, Reino Unido
Daniel D. Lee	University of Pennsylvania, EUA
Daniel Riesco	Universidad Nacional de San Luis, Argentina*
Daniele Quercia	Yahoo! Research em Barcelona, Espanha
Darryl Veitch	University of Melbourne, Austrália
David R. Choffnes	Northeastern University, EUA
Edward Fox	Virginia Tech, EUA
Ethan Katz-Bassett	University of Southern California, EUA
Fabrice Rastello*	ENS, Lyon, França
Frederico Fonseca*	Pennsylvania State University, EUA
Germán Montejano	Universidad Nacional de San Luis, Argentina*
Gilles Bertrand*	ESIEE Paris, Université Paris-Est, França

Gonzalo Navarro	Universidad de Chile, Chile
Hao Chi Wong	Intel Labs, EUA
Iadh Ounis	University of Glasgow, Reino Unido
Jean-François Boulicaut	INSA-Lyon, França
Johannes Buchmann	Centre of Advanced Security Darmstadt, Alemanha
Jörn Müller-Quade	Karlsruhe Institute of Technology, Alemanha
Krishna P. Gummadi	Max Planck Institute for Software Systems, Alemanha
Laure Gonnord*	ENS Lyon, França
Laurent Najman*	ESIEE Paris, Université Paris-Est, França
Manfred Hauswirth	Digital Enterprise Research Institute, Irlanda
Marco Ajmone Marsan	Politecnico di Torino, Itália
Marco Fiore*	CNR, Itália
Mário J. Silva	Instituto Superior Técnico de Lisboa, Portugal
Matthieu Cord*	LIP6, UPMC Paris 6, Sorbonnes Universités, França
Meeyoung Cha	Korea Advanced Institute of Science and Technology, Coreia do Sul
Mehdi Kaytoue*	INSA de Lyon, França
Michela Meo	Politecnico di Torino, Itália
Mohammed J. Zaki*	Rensselaer Polytechnique Inst., EUA
Peter Y. A. Ryan	Université de Luxembourg
Philip R. O. Payne	Ohio State University, EUA
Philippe Gosselin*	ENSEA, Université Cergy-Pontoise, França
Philippe Mahey	Université Blaise-Pascal, Clermont-Ferrand, França
Renata Teixeira	INRIA, França
Ricardo Baeza-Yates	Yahoo! Labs Barcelona, Espanha
Roberto Uzal	Universidad Nacional de San Luis, Argentina*
Ruzena Bajcsy	CITRIS, Berkeley University, EUA
Srinivasam Parthasarathy	Ohio State University, EUA
Sylvain Collange*	INRIA, França
Ümit V. Çatalyürek	Ohio State University, EUA
Vijay Kumar	University of Pennsylvania, EUA
Wolfgang Nejdl	L3S Research Center, Alemanha
Yacine Ghamri-Doudane*	Univ. de La Rochelle, França

19.2 Interações com Outros INCTs

Considerando o escopo do INCT-MCS, muitas interações com outros INCTs poderão ser criadas. Entre os INCTs que podem ser parceiros, citamos: INCT em Dengue, INCT de Medicina Molecular (INCTMM), INCT de Análise Socioeconômica e Demográfica (INTASED) e INCT de Nano-dispositivos Semi-condutores (DISSE), entre outros. Os objetivos de todos esses institutos envolvem aquisição, integração e/ou análise de dados, e eles se beneficiarão de uma metodologia para tal que integre os mundos físicos e virtual.

É importante destacar que o Inweb, precursor do INCT-MCS, já possui uma parceria com o INCT em Dengue. Juntos, os dois institutos desenvolveram o Observatório da Dengue, uma plataforma de monitoramento de dengue em redes sociais. O Observatório da Dengue é utilizado atualmente pelo Ministério da Saúde para monitorar epidemias de dengue no Brasil. Essa parceria será mantida com a criação do INCT-MCS.

20 Infraestrutura Disponível

Esta seção apresenta a infraestrutura computacional disponível nas três instituições participantes do INCT-MCS, bem como a relação de alguns dos principais projectos recentemente desenvolvidos pelos seus pesquisadores.

20.1 Infraestrutura Computacional

As três instituições participantes deste projeto oferecem, como contrapartida, recursos computacionais no valor de aproximadamente R\$ 2.650.000,00 (dois milhões, seiscentos e cinquenta mil reais), além do capital humano, conforme descrito a seguir.

UFMG. O Departamento de Ciência da Computação (DCC) da UFMG possui 43 pesquisadores que estarão diretamente ligados ao projeto, além de alunos de graduação, mestrado e doutorado. Além disso, a UFMG dispõe de sistema informatizado de acesso a periódicos e provê uma equipe de apoio técnico, gerencial e jurídico para proteção intelectual por meio da Cooedenação de Transferência e Inovação Tecnológica (CTIT). Os pesquisadores do DCC estão alocados em nove laboratórios: Laboratório para Tratamento da Informação (LATIN) – coordenado pelo Prof. Nivio Ziviani, Laboratório de Bancos de Dados (LBD) – coordenado pelo Prof. Alberto Laender, Laboratório de Análise e Modelagem de Desempenho (CAMPS) – coordenado pelo Prof. Virgílio Almeida, Laboratório de Visualização de Alto Desempenho (LVAD) – coordenado pelo Prof. Wagner Meira Jr., Laboratório de Vídeo sob Demanda (VOD) - coordenado pela Prof. Jussara Almeida, Núcleo de Processamento Digital de Imagens (NPDI) - coordenado pelo Prof. Arnaldo de Albuquerque Araújo, Laboratório de Visão e Robótico (VerLaB) – coordenado pelos Profs. Mário Campos e Luiz Chaimowicz, Laboratório de Redes de Computadores – coordenado pelo Prof. José Marcos S. Nogueira, Laboratório de Pesquisa Operacional (LAPO) – coordenado pelo Prof. Geraldo Robson Mateus, e *Wireless, Informational, Sensing, Embedded Systems & Models, Algorithms and Protocols Lab* - coordenado pelo Prof. Antonio A.F. Loureiro. Esses laboratórios possuem uma excelente infraestrutura computacional que, juntamente com os equipamentos solicitados nesta proposta, apoiarão a execução do projeto. Ao todo, os laboratórios possuem cerca de 60 servidores de alto desempenho organizados em dois *clusters*, mais de 80 estações de trabalho para tarefas de desenvolvimento e visualização gráfica, e capacidade de armazenamento de mais de 40 terabytes. Os laboratórios estão conectados por meio de uma rede alta velocidade (1Gbps) e utiliza o *backbone* da Rede Nacional de Ensino e Pesquisa (RNP). Esses recursos correspondem a uma contrapartida de aproximadamente R\$ 2.200.000,00.

UFAM. O Instituto de Computação da UFAM conta com oito pesquisadores que, juntamente com seus alunos de doutorado, mestrado e graduação, participarão do projeto. Em termos de infraestrutura, o grupo conta com um bem equipado laboratório que será usado no projeto. Esse laboratório conta com sete servidores de alto desempenho para experimentos, e está conectado à rede da UFAM e à Rede Nacional de Ensino e Pesquisa (RNP). Esses recursos correspondem a uma contrapartida de aproximadamente R\$ 350.000,00.

PUC Minas. A PUC Minas conta com quatro pesquisadores no projeto, que coordenam laboratórios que são usados pelos seus alunos de mestrado e iniciação científica. Esses laboratórios serão unificados para apoiar as atividades do INCT-MCS na instituição e contarão com oito estações de trabalho conectadas à Internet e à RNP, além de um servidor de alto desempenho para execução de experimentos. Esses recursos equivalem a uma contrapartida de aproximadamente R\$ 100.000,00.

20.2 Projetos Anteriores

Esta seção apresenta uma seleção dos principais projetos do grupo de pesquisadores (UFMG, UFAM e PUC Minas) nos últimos 5 anos, selecionados pela relevância para o contexto do INCT-MCS. No total, os pesquisadores do Instituto proposto coordenaram ou participaram em mais de 172 projetos de pesquisa, que somados superam o montante de R\$ 22.000.000,00 (vinte e dois milhões de reais, – aproximadamente 9.8 milhões de dólares americanos), conforme sumarizado na Tabela 8.

A seguir apresentamos uma breve descrição dos dez projetos mais relevantes, incluindo seus títulos, valores, vigências, agências financiadoras e participantes.

1. **Título:** InWeb – Instituto Nacional de Ciência e Tecnologia para a Web (CNPq/Fapemig, 2008–2015)

Valor: R\$4.097.439,11

Participantes: Virgílio A. F. Almeida (coordenador), Nivio Ziviani (subcoordenador), Alberto H. F. Laender, Arnaldo A. Araújo, Berthier Ribeiro-Neto, Wagner Meira Jr, Renato A. C. Ferreira, Clodoveu A. Davis, Dorgival Guedes Neto, Jussara M. Almeida, Marcos A. Gonçalves, Mirella M. Moro, Raquel O. Prates,

Tabela 8: Resumo de Projetos Anteriores.

Instituição	No. de Projetos Financiados	Agência de Fomento	Verba (R\$)	Total (R\$)
UFMG	57	CNPq	R\$6.968.158,47	R\$20.658.186,16
	79	Fapemig	R\$5.383.071,50	
	7	CAPES	R\$556.209,16	
	12	Empresas	R\$7.750.747,03	
UFAM	4	CNPq	R\$376.671,64	R\$1.340.001,45
	3	Fapeam	R\$603.329,81	
	2	Universo Online	R\$360.000,00	
PUC Minas	3	CNPq	R\$48.678,30	R\$239.778,30
	3	Fapemig	R\$144.000,00	
	2	Empresas	R\$47.100,00	

Adriano C. M. Pereira, Gisele L. Pappa, Altigran S. da Silva, Edleno S. De Moura, João M. B. Cavalcanti, José Palazzo M. de Oliveira, Carlos A. Heuser, Leandro Krug Wives, Renata M. Galante, Viviane M. Orengo, Cristina Murta, Evandrino G. Barros, Fabiano Botelho.

- Título:** Monitoramento Aéreo (Petrobrás, 2013 – 2016)
Valor: R\$ 3.424.520,74
Participantes: Mario F. M. Campos (coordenador), Erickson Rangel do Nascimento, William Rosbon Schwartz.
- Título:** Morelit – Monitoramento Remoto de Linhas de Transmissão (Cemig/Aneel, 2012–2015)
Valor: R\$ 2.037.907,20
Participantes: José Marcos S. Nogueira (coordenador), Daniel F. Macedo, Luiz. H. A. Correia, Luiz F. M. Vieira, Mario F. M. Campos.
- Título:** Veículos Autônomos (Instituto de Tecnologia da Vale, 2013–2016)
Valor: R\$ 931.739,00
Participantes: Mario F. M. Campos (coordenador), Luiz Chaimowicz, Erickson Rangel do Nascimento, Douglas Guimarães Macharet.
- Título:** C5 – Centro de Competência em Comunidades de Cooperação e Conhecimento (Fapemig, 2009–2012)
Valor: R\$ 795.960,00
Participantes: Wagner M. Júnior (coordenador).
- Título:** eCoSoC - Energy-Efficient Instrumentation to Secure Systems-on-a-Chip Devices (Intel Labs, 2012 – 2015)
Valor: US\$ 210.000,00
Participantes: Leonardo B. Oliveira (coordenador), Fernando Pereira, David Ott, Hao-chi Wong, Fernando Teixeira, Pablo Marcondes, Gustavo Vieira, Vitor Paisante, Luiz Felipe Saggioro, Antonio Maia, Bruno Silva e Henrique Santos.
- Título:** ALOA – Algoritmos, Otimização e Aplicações (Fapemig, 2009–2014)
Valor: R\$ 503.952,54
Participantes: Geraldo R. Mateus (coordenador), Antônio A. F. Loureiro, Clarindo I. P. da Silva e Pádua, José M. S. Nogueira, Rodolfo S. F. de Resende, Luiz Chaimowicz, Alexandre S. da Cunha, Sebastián A. Urrutia, Ricardo H. C. Takahashi, Maurício C. de Souza, Marcone J. F. Souza, André G. dos Santos, José E. C. Arroyo, Raquel A. F. Mini, Sérgio R. de Souza, Ricardo M. A. Silva.
- Título:** CAMPS – Centro de Análise e Modelagem de Performance de Sistemas (Hewlett Packard, 2009)

Valor: R\$ 345.238,54

Participantes: Virgílio A. F. Almeida (coordenador), Wagner M. Júnior, Jussara M. Almeida, Dorgival O. G. Neto.

9. **Título:** Técnicas para Tratamento de Documentos Semi-estruturados na Web (Fapeam, 2012 – 2015)

Valor: R\$ 277.789,81

Participantes: Altigran Soares da Silva (coordenador), Edleno Silva de Moura, João Marcos B. Cavalcanti, Marco Antônio P. Cristo, Daniel R. Fernandes.

10. **Título:** eVOX Pesquisa (Fapeam, 2014–2016)

Valor: R\$ 246.808,00

Participantes: Edleno Silva de Moura (coordenador), Pavel Calado, Altigran Soares da Silva, David Fernandes, João M. B. Cavalcanti, André Carvalho, Adriano Veloso, Maria da Graça C. Pimentel, Moisés Carvalho.

Referências

- [1] Charu C. Aggarwal. *Managing and Mining Sensor Data*. Springer Publishing Company, Incorporated, 2013.
- [2] H. Andrade, B. Gedik, and D. Turaga. *Fundamentals of Stream Processing: Application Design, Systems, and Analytics*. Cambridge University Press, 2014.
- [3] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805, 2010.
- [4] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Web Science (WebSci'12)*, pages 33–42, 2012.
- [5] Ricardo Baeza-Yates and B. Barla Cambazoglu. Scalability and efficiency challenges in large-scale web search engines. In *Companion Publication of the 23rd International Conference on World Wide Web Companion (WWW Companion'14)*, pages 185–186, 2014.
- [6] Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, and Christos Faloutsos. CopyCatch: stopping group attacks by spotting lockstep behavior in social networks. In *22nd International Conference on World Wide Web (WWW'13)*, pages 119–130, 2013.
- [7] David Chaum, Ivan Damgård, and Jeroen van de Graaf. Multiparty computations ensuring privacy of each party's input and correctness of the result. In *Advances in Cryptology (CRYPTO'87)*, pages 87–119, 1987.
- [8] Jon P. Daries, Justin Reich, Jim Waldo, Elise M. Young, Jonathan Whittinghill, Andrew Dean Ho, Daniel Thomas Seaton, and Isaac Chuang. Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, 57(9):56–63, September 2014.
- [9] Erickson R. do Nascimento, Gabriel L. Oliveira, Antônio W. Vieira, and Mario F. M. Campos. On the development of a robust, fast and lightweight keypoint descriptor. *Neurocomputing*, 120:141–155, 2013.
- [10] Deborah Estrin. Small data, where n = me. *Communications of the ACM*, 57(4):32–34, April 2014.
- [11] B. Faltings, J. J. Li, and R. Jurca. Incentive mechanisms for community sensing. *IEEE Transactions on Computers*, 63(1):115–128, Jan 2014.
- [12] Maksym Gabiello, Ashwin Rao, and Arnaud Legout. Sampling online social networks: An experimental study of Twitter. In *ACM Conference on SIGCOMM (SIGCOMM'14)*, pages 127–128, 2014.
- [13] Bruno Gonçalves, Duygu Balcan, and Alessandro Vespignani. Human mobility and the worldwide impact of intentional localized highly pathogenic virus release. *Scientific reports*, 3, 2013.
- [14] Cathal Gurrin, Alan F Smeaton, and Aiden R Doherty. Lifelogging: personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.
- [15] Johannes Hoffart, Yasemin Altun, and Gerhard Weikum. Discovering Emerging Entities with Ambiguous Names. In *23rd International Conference on World Wide Web (WWW'14)*, pages 385–396, Republic and Canton of Geneva, Switzerland, 2014.
- [16] H. V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, and Cyrus Shahabi. Big data and its technical challenges. *Communications of the ACM*, 57(7):86–94, July 2014.

- [17] Taeho Jung, XuFei Mao, Xiang-Yang Li, Shao-Jie Tang, Wei Gong, and Lan Zhang. Privacy-preserving data aggregation without secure channel: Multivariate polynomial evaluation. In *Conference of the IEEE Communications Society (INFOCOM'13)*, pages 2634–2642, April 2013.
- [18] Euisin Lee, Eun-Kyu Lee, M. Gerla, and S.Y. Oh. Vehicular cloud networking: architecture and design principles. *IEEE Communications Magazine*, 52(2):148–155, February 2014.
- [19] Quentin Lindsey, Daniel Mellinger, and Vijay Kumar. Construction of cubic structures with quadrotor teams. In *Proceedings of Robotics: Science and Systems (RSS'11)*, Los Angeles, CA, USA, June 2011.
- [20] Siyuan Liu, Shuhui Wang, Feida Zhu, Jinbo Zhang, and Ramayya Krishnan. HYDRA: large-scale social identity linkage via heterogeneous behavior modeling. In *ACM SIGMOD international conference on Management of data (SIGMOD'14)*, pages 51–62, 2014.
- [21] Yasuko Matsubara, Yasushi Sakurai, B Aditya Prakash, Lei Li, and Christos Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 6–14, 2012.
- [22] Massimo Melucci. Contextual search: A computational framework. *Foundations and Trends in Information Retrieval*, 6(4–5):257–405, 2012.
- [23] Alexander Mortl, Martin Lawitzky, Ayse Kucukyilmaz, Metin Sezgin, Cagatay Basdogan, and Sandra Hirche. The role of roles: Physical cooperation between humans and robots. *The International Journal of Robotics Research*, 31(13), November 2012.
- [24] Ravali Pochampally, Anish Das Sarma, Xin Luna Dong, Alexandra Meliou, and Divesh Srivastava. Fusing data with correlations. In *ACM SIGMOD international conference on Management of data (SIGMOD'14)*, pages 433–444, 2014.
- [25] Yingmei Qi, Lei Cao, Medhabi Ray, and Elke A. Rundensteiner. Complex event analytics. In *ACM SIGMOD international conference on Management of data (SIGMOD'14)*, pages 229–240, 2014.
- [26] Denis Shestakov. Current challenges in web crawling. In Florian Daniel, Peter Dolog, and Qing Li, editors, *Web Engineering*, volume 7977 of *Lecture Notes in Computer Science*, pages 518–521. Springer Berlin Heidelberg, 2013.
- [27] Thiago H. Silva, Pedro O. S. Vaz de Melo, Jussara M. Almeida, Mirco Musolesi, and Antonio A. F. Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *8th International Conference on Weblogs and Social Media (ICWSM'14)*, 2014.
- [28] Thiago H. Silva, Pedro O. S. Vaz De Melo, Jussara M. Almeida, and Antonio A. F. Loureiro. Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51, February 2014.
- [29] J. Ugander, L. Backstrom, C. Marlow, and Jon M. Kleinberg. Structural diversity in social contagion. *Proc. National Academy of Sciences*, 109(16):5962–5966, 2012.
- [30] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.
- [31] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1):97–107, January 2014.
- [32] M. J. Zaki and W. Meira Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press, May 2014.
- [33] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, and Xing Xie. Mining Novelty-seeking Trait Across Heterogeneous Domains. In *23rd International Conference on World Wide Web (WWW'14)*, pages 373–384, Republic and Canton of Geneva, Switzerland, 2014.