



Infraestrutura Nacional de Computação Distribuída

Estratégia de Desenvolvimento

Novos serviços, Renovação de equipamentos e Gestão de recursos humanos

Estado:	FINAL
Disseminação:	CONFIDENCIAL
Autores:	Jorge Gomes (LIP/INCD), Anabela Oliveira (LNEC), Catarina Ortigão (INCD), João Pina (LIP)
URL:	http://www.incd.pt

Data	Autor	Comentários
18-05-2014	Jorge Gomes	Estruturação baseada na proposta e plano de desenvolvimento
20-03-2015	Anabela Oliveira	Atualização
10-10-2019	Catarina Ortigão João Pina Jorge Gomes	Atualização



1. Estratégia global de desenvolvimento

A INCD é uma infraestrutura digital de suporte à investigação. Disponibiliza serviços de computação e dados à comunidade científica e académica nacional, em todos os domínios. A infraestrutura está especialmente vocacionada para prestar serviços de computação e tratamento de dados suportando os investigadores e a sua participação em projectos nacionais e internacionais, assim como a formação avançada através da disponibilização de meios para realização de teses de mestrado e doutoramento.

A INCD é uma evolução da infraestrutura da Iniciativa Nacional Grid (INGRID) criada em 2008/2009 no contexto da Iniciativa Nacional Grid. **A estratégia da INCD assenta no desenvolvimento contínuo do leque de serviços e da capacidade da infraestrutura**, de forma a suportar os requisitos actuais e futuros da comunidade científica .

Pela sua natureza as infraestruturas de investigação digitais requerem uma aposta constante na inovação e desenvolvimento. Só assim é possível acompanhar a constante evolução das fronteiras do conhecimento e dos seus novos requisitos computacionais. Os serviços oferecidos por este tipo de infraestruturas carecem de uma constante evolução e adaptação às necessidades da comunidade científica, em função do estado da arte. A computação científica requer serviços e conhecimentos bastante específicos, que importa desenvolver e suportar de forma profissional e sustentada em articulação com os utilizadores.

As infraestruturas digitais genéricas, como a INCD, devem poder afirmar-se como elementos de suporte à comunidade científica e académica. Só assim se poderá evitar a dispersão de esforços e investimentos, criar soluções comuns, e apoiar de forma eficiente o desenvolvimento destas infraestruturas. **É portanto fundamental dotar a INCD dos recursos materiais, operacionais e humanos** que permitam:

- a) **Desenvolver e implementar** novos serviços e realizar a sua contínua melhoria;
- b) **Renovar e aumentar** a capacidade de processamento e armazenamento da infraestrutura, tanto dos serviços existentes como para novos serviços, de forma a poder responder aos desafios (ex. apoio ao roteiro e projetos financiados pela FCT, projectos estratégicos e/ou de interesse nacional como a participação no CERN, formação avançada, investigação em todos os domínios, etc);
- c) **Trabalhar em conjunto** com todas as infraestruturas, projectos estratégicos e investigadores apoiando o **desenvolvimento e/ou adaptação** das suas ferramentas e plataformas digitais, e sua operacionalização na INCD;
- d) **Apoiar** a participação em **projetos** nacionais e internacionais **de relevância estratégica**.
- e) Prestar **suporte**, consultoria e formação **aos utilizadores**;
- f) Divulgar e **promover o uso** da infraestrutura e seus serviços.



2. Dimensão Europeia e Internacional

Tanto a nível nacional como internacional ainda subsiste um ecossistema fragmentado de serviços de tecnologias de informação e comunicação, com inerentes problemas de acesso e interoperabilidade. É necessário harmonizar este ecossistema para que seja possível explorar todo o potencial dos meios digitais e dados científicos.

O **European Grid Infrastructure (EGI)** e seus parceiros propõem uma visão integrada deste espaço que inclui dados, instrumentos, serviços e conhecimento. Pretende-se assim reduzir as barreiras de acesso e partilha da informação. Esta visão encontra-se alinhada com a iniciativa da comissão Europeia para a criação do **European Open Science Cloud (EOSC)**, caminhando no sentido de uma maior integração entre serviços e dados, e permitindo juntar as ferramentas e a capacidade necessárias à efetiva exploração dos dados científicos.

A INCD como infraestrutura parceira do EGI partilha esta visão com a qual alinou o seu desenvolvimento. A INCD participa no EOSC através do EGI, enquanto nó da federação de computação, e também diretamente enquanto participante em projetos Europeus e outras atividades no âmbito do EOSC. A participação no EGI e no EOSC é realizada em estreita parceria com Espanha através da **infraestrutura Ibérica IBERGRID**, na qual assenta a colaboração Ibérica na computação e processamento de dados distribuídos.

A INCD encontra-se em posição estratégica para apoiar na federação de serviços e dados a nível nacional e internacional promovendo uma utilização mais eficiente dos meios existentes, e a partilha dos recursos e do conhecimento.

A estratégia define os seguintes pilares:

- **Recursos partilhados:** os recursos, dados, instrumentos, serviços digitais, software, publicações, conteúdos educativos, conhecimento e experiência, devem ser partilhados.
- **Acesso aberto:** deve estar bem definido e ser não discriminatório.
- **Políticas:** devem existir políticas de acesso harmonizadas e incentivos ao uso responsável.
- **Gestão:** os serviços devem ser geridos de forma profissional e transparente. Os serviços devem ser estabelecidos e operados numa perspectiva de longo termo.
- **Governança:** os modelos de governo devem incluir os múltiplos intervenientes, incluindo a participação das infraestruturas digitais e comunidades de utilizadores.
- **Cultivo:** apoio persistente e de longo termo por parte das agências de financiamento, para que as infraestruturas possam estabelecer estratégias com vista a um futuro Europeu comum. Estabelecer políticas que permitam apoiar o crescimento e desenvolvimento contínuo das infraestruturas e suas capacidades. Esta



3. Desenvolvimento dos serviços de computação HPC e HTC

A INCD oferece os seguintes serviços de computação avançada do tipo HPC e HTC:

- **High Throughput Computing (HTC):** para processamento single-threaded ou multithreaded.
- **High Performance Computing (HPC):** para processamento paralelo usando Infiniband para baixa latência ou Ethernet.

Ambos os serviços podem ser acedidos através:

- **Acesso Local:** os utilizadores acedem a um cluster interactivo via SSH no qual podem realizar o desenvolvimento das aplicações, e submeter trabalhos a um sistema batch.
- **Acesso federado:** para processamento HTC e HPC sequencial, multithreaded e paralelo, acoplado a gestão de dados em larga escala. Os utilizadores acedem via middleware ao serviço HTP e HTC da INCD, ou a outro sistema compatível disponível nas infraestruturas IBERGRID e/ou EGI.

Estes serviços de processamento são complementados por:

- **Áreas de software NFS:** permitem a disponibilização de software e dados aos processos em execução através do acesso NFS a sistemas de ficheiros centrais.
- **Áreas de software CVMFS:** permitem disponibilizar software aos processos em execução através de um sistema de *cache* geograficamente distribuído.
- **Armazenamento HPC:** permite suportar I/O eficiente por parte de trabalhos paralelos. Atualmente é implementado através de um sistema de armazenamento Lustre otimizado para acesso paralelo.
- **Armazenamento HTC:** para suportar aplicações HTC ou HPC sem requisitos de I/O paralelo.
- **Gateways de armazenamento:** permitem o acesso remoto aos sistemas de armazenamento com diversos protocolos para acesso e transferência de dados.

Importa aumentar a capacidade de processamento e armazenamento associada a estes serviços. Durante 2018/2019 foi efetuado um primeiro aumento da capacidade associada aos serviços de HPC/HTC instalados no centro operacional de Lisboa, com o objetivo de renovar o equipamento mais antigo.

4. Serviços de computação em nuvem

Os investigadores requerem soluções cada vez mais sofisticadas e adaptadas às suas necessidades. Usam software cada vez mais complexo, que requer ambientes de execução específicos e/ou adaptados aos seus requisitos. Cada vez mais são usados métodos de processamento diferentes do tradicional processamento em *batch* (ex. Linux Containers, Kubernetes, Mesos, Hadoop, etc). O acesso aos dados é cada vez mais efectuado através de novas soluções de armazenamento (ex. bases de dados *NoSQL*, armazenamento de objetos, etc). Diversas aplicações de uso comum são disponibilizadas sob a forma de serviços permanentes que requerem elasticidade.



Infraestrutura Nacional de **Computação Distribuída**

Para ir de encontro às novas necessidades é necessário implementar serviços genéricos e flexíveis capazes de suportar um maior leque de requisitos. A computação em nuvem (*cloud computing*) especialmente na sua vertente de *Infrastructure as a Service* (IaaS) permite disponibilizar aos utilizadores acesso *on-demand* a máquinas virtuais ou físicas cuja configuração e gestão pode ser delegada nos próprios utilizadores, com maior flexibilidade e adaptabilidade às suas necessidades específicas.

A INCD implementou um serviço de computação em nuvem do tipo IaaS que já está a ser explorado por diversas comunidades de utilizadores. O serviço encontra-se integrado na cloud federada do European Grid Initiative (EGI) e na IBERGRID. O serviço de computação em nuvem da INCD, oferece uma **base adaptável para a construção e disponibilização de serviços para múltiplas comunidades e projectos, nacionais e internacionais, incluindo o suporte aos ESFRIs e infraestruturas do roteiro.**

Em paralelo foram implementados **serviços de armazenamento** do tipo **block storage** e **object storage** baseados em Ceph. Estes serviços constituem a base para o armazenamento resiliente em larga escala, servindo o serviço de computação em nuvem, e podendo também ser explorados separadamente por outros serviços.

Os serviços de computação em nuvem e seu armazenamento, oferecem uma base flexível para suporte dos requisitos da comunidade científica e académica, assim como para apoio às restantes infraestruturas do roteiro. A existência de meios de computação em nuvem disponíveis à comunidade científica nacional, operados sem fins comerciais, e orientados à investigação deve constituir uma prioridade. É portanto fundamental apoiar o seu desenvolvimento e adoção.

A INCD pretende também suportar computação em nuvem híbrida, permitindo a utilização combinada dos recursos de computação em nuvem da INCD, maximizando e/ou racionalizando a utilização dos meios à disposição dos utilizadores.

Este tipo de utilização pode incluir a combinação de meios da INCD com meios próprios dos utilizadores, ou meios disponibilizados por outros fornecedores de serviço académicos e/ou comerciais. A INCD aderiu aos acordos de computação em nuvem no âmbito do protocolo GÉANT, e avaliou a contração de serviços através destes acordos. A INCD pretende oferecer serviços que facilitem a computação em nuvem híbrida. Assumem particular relevo a autenticação federada, a interoperabilidade e a orquestração.

5. Serviços avançados de computação

Neste contexto inserem-se tecnologias emergentes e/ou serviços de suporte a necessidades específicas que complementam e/ou expandem a oferta de serviços de computação (HPC, HTC, computação em nuvem). A INCD identificou três serviços de interesse imediato.

A INCD pretende expandir o suporte às tecnologias de virtualização leve, especialmente *Linux containers*, que possuem uma cada vez maior adoção por parte da comunidade científica no encapsulamento e execução de serviços e aplicações. Esta tecnologia facilita uma execução eficiente e transparente sobre diversas plataformas minimizando a utilização da virtualização convencional e



Infraestrutura Nacional de **Computação Distribuída**

permitindo ganhos em desempenho e poupança na utilização dos recursos. A INCD desenvolve em parceria com o LIP uma ferramenta flexível para execução de containers em sistemas HPC e HTC. A INCD suporta a execução de Linux containers nos serviços HPC, HTC e computação em nuvem. Pretende-se expandir o suporte à orquestração através da disponibilização de serviços e ferramentas adequadas. Exemplo: kubernetes, Mesos, OpenShift, etc.

Disponibilização de acesso a servidores com elevada capacidade de memória RAM e I/O para a execução de aplicações com requisitos elevados. Enquadram-se nesta tipologia alguns tipos de processamento massivo de dados. Exemplo: sequenciação de genoma, análise massiva de dados.

Disponibilização de acesso a servidores equipados com aceleradores (ex. GPGPUs) para computação acelerada, integrados nos serviços HPC, HTC e de computação em nuvem. Exemplo: inteligência artificial, processamento de imagens, *machine learning*.

6. Integração e alojamento

A INCD está a desenvolver a componente de alojamento e integração de serviços e equipamentos de outras entidades na sua infraestrutura.

- a) Integração de serviços que são alojados na infraestrutura INCD, explorando as diversas capacidades (computação em nuvem, HPC, HTC e armazenamento), esta integração pode envolver a utilização de meios e internacionais através da IBERGRID, EGI, EOSC, EUDAT etc. Exemplo: a INCD colabora com o LNEC na integração e alojamento de diversos serviços de simulação aplicados à engenharia civil e costeira, que usam meios disponibilizados através da IBERGRID e EGI apoiados no âmbito do EOSC pelos projetos EOSC-hub e EOSC-Synergy.
- b) Integração de equipamentos em localizações remotas permitindo um mais fácil acesso, utilização integrada e optimização desde recursos. A integração de sistemas HPC e Cloud em estabelecimentos de ensino superior e centros de investigação é particularmente importante. Exemplo: integração do cluster HPC do ISEC em Coimbra no portfólio de serviços da INCD.
- c) Alojamento de equipamentos pertencentes a outras instituições em condições adequadas em datacenters geridos pela RCTS, integrando estes equipamentos nos serviços centrais da INCD e partilhando a sua capacidade com a comunidade. Exemplo: integração e partilha dos equipamentos de armazenamento e processamento do nó nacional do GBIF e do Tier-2 do WLCG.

7. Armazenamento de longo termo e repositórios de dados

A INCD pretende ainda desenvolver serviços de armazenamento de longo termo e proteção de dados para suportar as crescentes necessidades nesta área. É cada vez mais importante salvaguardar os dados resultantes dos projectos de investigação. Esta necessidade encontra-se patente nas políticas de acesso aberto da FCT. É igualmente importante que estes dados possam ficar armazenados de forma integrada com os serviços de computação em nuvem, HPC e HTC. Esta



Infraestrutura Nacional de **Computação Distribuída**

integração é essencial à exploração dos dados em condições adequadas. A INCD pretende desenvolver um piloto de serviço de arquivo de dados em colaboração com o EUDAT no âmbito do EOSC-hub, e com o CSIC e DANS no âmbito do projeto EOSC-Synergy.

8. Renovação de equipamentos

A infraestrutura da INCD está a ser explorada por diversas comunidades. Em 2019 foi efetuada uma renovação de equipamentos instalados no centro operacional de Lisboa, que permitiu aumentar a capacidade de processamento e armazenamento dos serviços HPC, HTC e cloud. Foi também iniciado em 2019 um segundo centro operacional no norte do país. No entanto, estes investimentos são insuficientes para dar resposta às necessidades crescentes.

Alguns exemplos:

- A capacidade fornecida às experiências do CERN/LHC através do Tier-2 operado pela INCD ao abrigo do MoU de computação do LHC (MoU for Collaboration in the Deployment and Exploitation of the Worldwide LHC Computing Grid) deve ser revista e aumentada anualmente. Apesar do aumento de capacidade realizado em 2019, a capacidade do Tier-2 nacional no WLCG continua quatro vezes abaixo da capacidade nominal necessária, comprometendo a análise de dados e a participação nas experiências do LHC. É importante realçar que o aumento constante da luminosidade do colisionador LHC implica um ajuste constante da capacidade de processamento e armazenamento.
- O LNEC tem requisitos crescentes sobretudo na simulação de dados e no processamento de imagens de satélites de observação da terra decorrentes da investigação em engenharia civil, engenharia costeira, hidráulica, eficiência energética, entre outras.
- A modelação das alterações climáticas exige cada vez maiores capacidades de processamento para a simulação detalhada dos fenómenos associados, a para as simulações que constituem a contribuição nacional para os relatórios do IPCC.
- A utilização de simulação e análise de dados em ciências da vida é cada vez maior. O estudo de proteínas, genoma, e física médica (na terapia e diagnóstico) são algumas das áreas cujo crescimento na INCD é cada vez maior, e exige meios com mais capacidade.

A INCD requer um reforço da capacidade inicialmente prevista em 2013 para:

- a) Apoiar um número cada vez maior de investigadores de múltiplas áreas.
- b) Satisfazer as crescentes necessidades dos investigadores.
- c) Satisfazer as necessidades decorrentes da participação em projetos de investigação nacionais e internacionais.
- d) Respeitar os compromissos relacionados com a participação no CERN e outras organizações.
- e) Participar efetivamente nas actividades de I&D e construção das infraestruturas do roteiro e ESFRIs.
- f) Acompanhar a evolução das tecnologias de informação.
- g) Disponibilizar novos serviços tal como o arquivo de dados e processamento associado.



9. Gestão de recursos humanos

Os recursos humanos necessários ao funcionamento da infraestrutura têm sido suportados quer pela Associação INCD, quer pelos associados através de financiamento próprio e da participação em projectos nacionais e internacionais.

Para disponibilizar serviços com qualidade e suportar um maior leque de utilizadores, é essencial garantir a existência de recursos humanos com elevada competência, tanto para o desenvolvimento como para a operação dos serviços. A estratégia de gestão de recursos humanos assenta nos seguintes bases que importa garantir:

- a) Manter através de contratos permanentes um núcleo de investigadores e pessoal técnico chave, que possui conhecimentos e experiência únicos, e que é essencial para o cumprimento da missão da INCD.
- b) Complemento dos investigadores e técnicos chave através de contratos a termo certo, para as actividades de desenvolvimento.
- c) Evitar a saída de pessoal chave através de condições de trabalho competitivas.
- d) Evitar a perda de conhecimento através de uma redundância adequada nas posições chave.
- e) Garantir os contratos através do Roteiro e/ou de medidas inseridas na estratégia nacional de computação avançada.
- f) Complementar a capacidade de inovação através de contratos de investigação suportados por projectos nacionais e Europeus.

10. Centros e Parcerias

A estratégia de desenvolvimento da INCD passa pela colaboração com os centros de investigação, laboratórios, instituições de ensino superior e comunidades de utilizadores, partilhando e explorando conhecimentos e capacidades. A INCD está a implementar centros operacionais e de suporte, que contribuem para a infraestrutura em áreas como o treino, formação avançada, I&D, suporte, operação, etc. Os centros podem ser do tipo interno quando operados pelos associados da INCD, ou do tipo externo quando operados por outras entidades.

- Os Centros Operacionais operam equipamentos, disponibilizam serviços, prestam apoio aos utilizadores, e trabalham em estreita colaboração com os centros de suporte na customização e melhoramento dos serviços.
- Os Centros de Suporte prestam apoio aos utilizadores desde a triagem dos pedidos de acesso, à utilização dos serviços, passando pela identificação de necessidades, customização de serviços em colaboração com os centros operacionais, adaptação de aplicações, formação e consultadoria.

Centros de ambos os tipos, consoante as suas valências, colaboram no desenvolvimento de novos serviços e no seu melhoramento contínuo.

A infraestrutura possui Centros Operacionais em:



Infraestrutura Nacional de **Computação Distribuída**

Centro Operacional	Descrição
INCD-Lisboa	Instalado no campus do LNEC em Lisboa, datacenter sala-grid. Disponibiliza serviços computação avançada e suporte aos utilizadores que incluem: HPC, HTC, cloud, grid e sistemas de armazenamento de alto desempenho.
INCD-Minho	Instalado no Riba-de-Ave, no datacenter da REN. Disponibiliza serviços de computação avançada: HPC, HTC e armazenamento. Os serviços estão a ser expandidos para incluir cloud e grid.
U.Porto	Centro Externo na Universidade do Porto. Disponibiliza um serviço de HTC integrado na infraestrutura de computação INCD, IBERGRID e EGI, focado em aplicações de ciências da vida.
ISEC	Centro Externo no Instituto Superior de Engenharia de Coimbra. Disponibiliza um serviço de HPC integrado através de uma SLA no portfólio da INCD.

A infraestrutura possui Centros de Suporte em:

Centro de Suporte	Descrição
LIP-Lisboa	No LIP em Lisboa na U.Lisboa. Coordenação técnica. Ponte com as infraestruturas: WLCG, EGI, EOSC, IBERGRID. Suporte a utilizadores. Suporte à Física de Altas Energias e Astropartículas.
LIP-Minho	No LIP no Minho na U.Minho. Disponibiliza competências e suporte para análise de dados, técnicas de Inteligência Artificial e Machine Learning.
LNEC	No LNEC em Lisboa. Coordenação de pilotos de utilização. Suporte a aplicações de Engenharia Civil. Desenvolvimento de novos serviços e ligação às comunidades de Engenharia Civil.
RCTS	Em Lisboa. Ligação com outras infraestruturas. Gestão do alojamento (Centros de Processamento de Dados). Redes e serviços avançados de comunicação de dados.

A INCD pretende ainda implementar um centro operacional apoiado por um centro de suporte na região centro em Coimbra. Outros centros operacionais e de suporte poderão ser estabelecidos em parceria com outras entidades.

11. Orçamento

O orçamento inicialmente estimado em 2013, previa um investimento na INCD de aproximadamente 5 milhões de Euro ao longo de seis anos entre 2014 e 2019. **Infelizmente o início do roteiro sofreu um atraso substancial. O primeiro financiamento foi recebido pela INCD apenas em 2017, para um período de apenas três anos (em vez de seis), e com um montante total muito inferior ao inicialmente previsto (apenas 2.7 milhões).**



Infraestrutura Nacional de **Computação Distribuída**

É essencial um reforço do financiamento que permita a execução do orçamento inicialmente estimado para o período de 2014 a 2019 (5 milhões de Euro), e também o seu reforço em função da evolução das necessidades ocorrida desde então.

11.1 Equipamento

O financiamento incluído na proposta da INCD contempla apenas capacidade para suportar as necessidades identificadas em 2013, assim como alguma capacidade para partilha. **Note-se que o orçamento tinha sido calculado em função das necessidades conhecidas à data. Desde então estas necessidades cresceram.**

11.2 Recursos humanos

Existe necessidade de recursos humanos, focada no desenvolvimento e suporte aos utilizadores. A evolução da ciência e da tecnologia conduz a um aumento das exigências dos investigadores no que toca às TIC. Esta evolução exige um esforço constante de desenvolvimento e atualização por parte da INCD, o que exige uma equipa dedicada, tecnicamente competente, apoiada permanente atualização de conhecimentos.