

# Ciberinfraestrutura para integração, acesso, compartilhamento e reuso de dados de pesquisa da área nuclear

Luana F. Sales<sup>1</sup>, Luís Fernando Sayão<sup>2</sup>

<sup>1</sup>Instituto de Engenharia Nuclear (IEN) – Comissão Nacional de Energia Nuclear (CNEN)  
Rua General Severiano, n.90 – 22294-901 Botafogo - RJ – Brazil

<sup>2</sup>Centro de Informações Nucleares (CIN) - Comissão Nacional de Energia Nuclear (CNEN)

lsales@ien.gov.br, lsayao@cnen.gov.br

**Abstract.** *This article describes the current stage of development of a cyberinfrastructure oriented to integration, curation, access, sharing and reuse of research data and information in the field of nuclear sciences. Conceptually the project is focused on the idea of approaching research in nuclear area of the new scientific paradigm, called e-science. The technological base of cyberinfrastructure is centered on a system of digital repositories that integrates research data and e-prints, and on an information model whose metadata structure provides interoperability and management of information resources. Among the services provided by the system are emphasized the access to the data and information resources, the support to content production, the research data curation and digital preservation, and the generation of a indicators sets of research activities.*

**Resumo.** *O artigo descreve o estágio atual do desenvolvimento de uma ciberinfraestrutura voltada para a integração, curadoria, acesso, compartilhamento e reuso de dados e informações de pesquisa no domínio da área nuclear. Conceitualmente o projeto está centrado na ideia da aproximação da pesquisa na área de ciências nucleares do novo paradigma científico, chamado de e-science. A base tecnológica da ciberinfraestrutura está baseada em um sistema de repositórios digitais que integra dados de pesquisa e e-prints, e em um modelo de informação cuja estrutura de metadados proporciona a interoperabilidade e a gerenciamento dos recursos informacionais. Entre os serviços oferecidos pelo sistema destacam-se o acesso aos dados e informações e a produção de conteúdos, a curadoria e a preservação digital, e ainda geração de conjunto de indicadores de atividades de pesquisa.*

## 1. Introdução

A pesquisa na área nuclear aproxima-se muito do novo paradigma científico chamado e-science (HEY, 2009). Essa nova forma de fazer ciência é compreendida como uma ciência orientada por dados, apoiado pelo uso intensivo de computação, redes e recursos de informação distribuídos. Nesse novo contexto, o desenvolvimento científico em áreas altamente tecnológicas, como a de ciências nucleares, depende mais e mais da colaboração entre pesquisadores, acadêmicos e especialistas em informação e em computação. Para se intensificar essa interlocução, é necessário a criação de uma “ciberinfraestrutura”, que integre várias tecnologias e metodologias, que melhore a qualidade da pesquisa e torne mais veloz a transformação desse conhecimento em benefícios para a sociedade.

O alto grau de interdisciplinaridade e a produção massiva de dados e informações, torna os ciclos de geração de conhecimentos científicos atuais, verdadeiras atividades sociais,

que tem uma necessidade vital de interatividade, compartilhamento de recurso e de um diálogo constante entre pares geograficamente distantes. É necessário, portanto, oferecer as comunidades científicas ferramentas e espaços virtuais que intensifiquem essa interlocução, que encurte o ciclo da comunicação científica e que amplie a visibilidade e o alcance das pesquisas científicas.

Entretanto outros contextos devem ser considerados na pesquisa atual, como ética, transparência e visibilidade. Os mecanismos de transparência que testemunhem as atividades que se desenrolam nos laboratórios se tornam essenciais, na medida em que investimentos volumosos são feitos por parte da sociedade, que agora exige retornos visíveis. Sendo assim, torna-se imprescindível o desenvolvimento de espaços virtuais de memória onde dados e informações digitais possam ser tratados, armazenados e preservados para acesso de longo prazo por parte de todas as comunidades interessadas. A computação e a ciência da informação nos oferecem tecnologias, padrões e metodologias, sintetizadas pelos repositórios digitais, que hoje cumprem a missão de custodiar a memória das instituições de pesquisa e de torná-las interoperáveis e acessíveis. Essa custódia compreende, além da preservação conhecimento, metodologias de curadoria de dados de pesquisa de forma que os *data sets* possam ser reutilizados como insumo para novas pesquisas. Além do mais, os repositórios institucionais disponibilizam estoques de informações que, tratadas e analisadas, são um instrumento valioso para a gestão de CT&I.

Por fim, as instituições de pesquisa precisam transferir os conhecimentos gerados nos seus laboratórios na forma de benefícios perceptíveis pela sociedade, por meio, por exemplo, da inovação tecnológica e pela disseminação adequada de novos saberes. A transferência de conhecimento para segmentos específicos da sociedade torna-se, como no caso da Comissão Nacional de Energia Nuclear<sup>1</sup> (CNEN), de extrema importância. Isso se traduz, na maioria das vezes pela criação, seleção e a formatação de conteúdos, que possam ser disseminados via cursos a distância, de interesse de comunidades envolvidas na utilização da energia nuclear. O apoio da tecnologia para a difusão desses conhecimentos se torna crucial dado à diversidade de público e das dimensões continentais do país.

Esta iniciativa tem por objetivo unificar nacionalmente as atividades de pesquisa desenvolvidas nas diversas unidades da CNEN, sendo uma proposta para compartilhamento e reuso de dados de pesquisa. Para tal, a CNEN se propõe o desenvolvimento de uma infraestrutura que integre tecnologias, recursos informacionais e metodologias que dêem suporte a geração, o armazenamento, a preservação e a transferência do conhecimento nuclear, sendo uma tentativa de responder aos novos desafios no domínio nuclear gerados pelo novo paradigma da eScience.

## **2. Materiais**

A matéria-prima necessária para desenvolvimento dos repositórios digitais é o objeto digital, i.e, e-prints e dados, somados aos metadados que garantem a estrutura e o significado dos objetos digitais e a sua capacidade de serem preservados e reusados.

---

<sup>1</sup> [www.cnem.gov.br/](http://www.cnem.gov.br/)

No contexto da CNEN os objetos digitais que compõem a Ciberinfraestrutura são: produção técnico-científica (artigos de periódicos, livros e capítulos de livros (livres de copyright), teses e dissertações, publicações de eventos, projetos, relatórios técnicos e patentes.); materiais de ensino (apostilas, apresentações, trabalhos de aluno) e dados de pesquisa oriundos dos laboratórios (observacionais, experimentais e computacionais)

Todos esses materiais são acompanhados por um modelo de metadados que garantem que eles sejam preservados, acessados, gerenciados, recuperados e disseminados internamente ou externamente, quando o material não possui restrição de copyright ou algum sigilo. Assim, este modelo considerou em seu desenvolvimento metadados descritivos, técnicos, administrativos e de preservação para a representação descritiva de cada objeto.

### **3. Métodos**

O desenvolvimento da Ciberinfraestrutura está em estágio inicial. As etapas metodológicas se dividem em:

- 1) Levantamento dos tipos de materiais que irão compor a ciberinfraestrutura, conforme relatado no capítulo anterior e identificação dos metadados necessários para cada tipo.
- 2) Desenvolvimento de um repositório modelo em um dos Institutos da CNEN, com a finalidade de ser replicado nas unidades que ainda não possuem nenhuma iniciativa de repositório institucional ou de preservação da memória.
- 3) Identificação dos sistemas de preservação da memória técnico-científica existentes em outras unidades da CNEN, por exemplo Base PTC (produção técnico-científica) do IPEN e Base memória do CIN-CNEN.
- 4) Identificação de um núcleo mínimo comum de metadados que possam servir de padrão para integração do sistemas e intercâmbio de dados.
- 5) Concepção de uma metodologia para curadoria dos objetos digitais depositados no repositório.
- 6) Criação de um Gateway de integração de todos os serviços e produtos com facilidades para personalização e características de ergonomia e acessibilidade. Duas instâncias de acesso: informações com acesso aberto para toda a sociedade e informações restritas aos pesquisadores da comunidade nuclear.

### **4. Resultados Parciais**

Das etapas metodológicas descritas acima, estão concluídas: o levantamento dos tipos de materiais que irão compor a ciberinfraestrutura, a identificação dos metadados necessários para cada tipo de material e a implementação de um repositório modelo no âmbito de uma das unidades da CNEN: O Instituto de Engenharia Nuclear (IEN), identificação de outros sistemas de preservação da memória existentes na CNEN, identificação do núcleo comum de metadados para possibilitar a integração das bases e a interoperabilidade com outros repositórios internacionais.

## 5. Conclusões e Perspectivas

O presente trabalho é fruto de uma experiência em andamento que visa desenvolver e disseminar em âmbito nacional uma infraestrutura que integre tecnologias, recursos informacionais e metodologias que deem suporte a geração, armazenamento, a preservação e a transferência do conhecimento nuclear, a partir da criação de ambientes virtuais interativos que proporcione a troca de informações e dados, de práticas e experiências, bem como a geração compartilhada de conhecimento, o desenvolvimento remoto de projetos e a integração com sistemas internacionais de pesquisa.

Conceber e desenvolver ferramentas voltadas para a gestão de CT&I que inclua a geração de indicadores, métricas para avaliação do retorno para a sociedade, grau de inovação, geração de patentes, etc é outro desejo futuro.

## 6. Referências

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin. Jim Gray on eScience: A Transformed Scientific Method. In: Hey, Tony; Tansley, Stewart; Tolle, Kristin (Orgs.). **The Fourth Paradigm: Data-Intensive Scientific Discovery**, 2009. Disponível em: [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_jim\\_gray\\_transcript.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf). Acesso em: 20 dez. 2011.

GRAY, Jim et al. Scientific data management in the coming decade. **ACM SIGMOD Record**, v. 34, n. 4, p. 34-41, 2005.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure vision for 21st century discovery**. 2007. Disponível em: <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>. Acesso em: 13 mar. 2013.

SOEHNER, Catherine; STEEVES, Catherine; WARD, Jennifer. **E-Science and Data Support Services**. 2010. Disponível em: <http://arl.nonprofitsoapbox.com/storage/documents/publications/escience-report-2010.pdf> Acesso em 13 abr 2013.