

## 4.3 Ferramentas de Publicação e Citação de Dados de Pesquisa e Objetos Digitais

HENSE, A.; QUADT, F. Acquiring high quality research data. **D-Lib Magazine**, v. 17, n. 1/2, 2011. doi:10.1045/january2011-hense.  
<<http://www.dlib.org/dlib/january11/hense/01hense.html>>.

# Formato do documento x dados

- Arquivos de textos: formato (HTML, RTF, Word, PDF)

Arquivos de dados: formato (XML, planilhas, formatos binários) - otimizados para determinados domínio.

- Formatos individuais não podem ser abertos com ferramentas padrão, precisam de software de análise e visualização.

# Conteúdo do documento x dados

- Documento transfere conhecimento subjetivo dentro de uma estrutura lógica, didática, escrito e lido por humanos. Interpreta dados, usa tabelas, figuras e imagens para ilustrar.
- Dados são produzidos direta ou indiretamente por computadores ou instrumentos - importados e processados por um computador. Arquivos de dados são objetivos e simples, e nem sempre sujeitos à inspeção direta do humano.

# Publicação do artigo

- O processo de avaliação da qualidade científica (SQA) - peer-review – referer - tem uma longa tradição. O revisor precisa ter uma profunda compreensão do tema. Sendo os *papers* otimizados para o entendimento de humanos, em um volume restrito de páginas, SQA é viável na maioria dos casos.

# Publicação dos Dados

- Grandes formatos de dados não suscetíveis a percepção do humano - Revisor faz uso de softwares  
- validação de um arquivo XML - transferência livre de erros de origem para o destino de localização
- Avaliação dos metadados - tipo de dados e domínio  
- cruciais para navegação e buscas dos dados
- O nível de qualidade dos dados não é "peer-avaliação" mas "aprovado - pelo-autor" – garantia formal de qualidade.

# Pesquisa e navegação -Textos

- Indexáveis, metadados explícitos, classificados, catalogados. Repositórios suportam todos os formatos de arquivos de texto.
- São salvos como arquivo único e armazenados em repositório - arquivo de longo prazo - acesso direto ao documento.
- Textos não cumprem os requisitos para o gerenciamento de arquivos de dados -vinculação a recursos externos - links quebrados e sincronização.

# Pesquisa e navegação - Dados

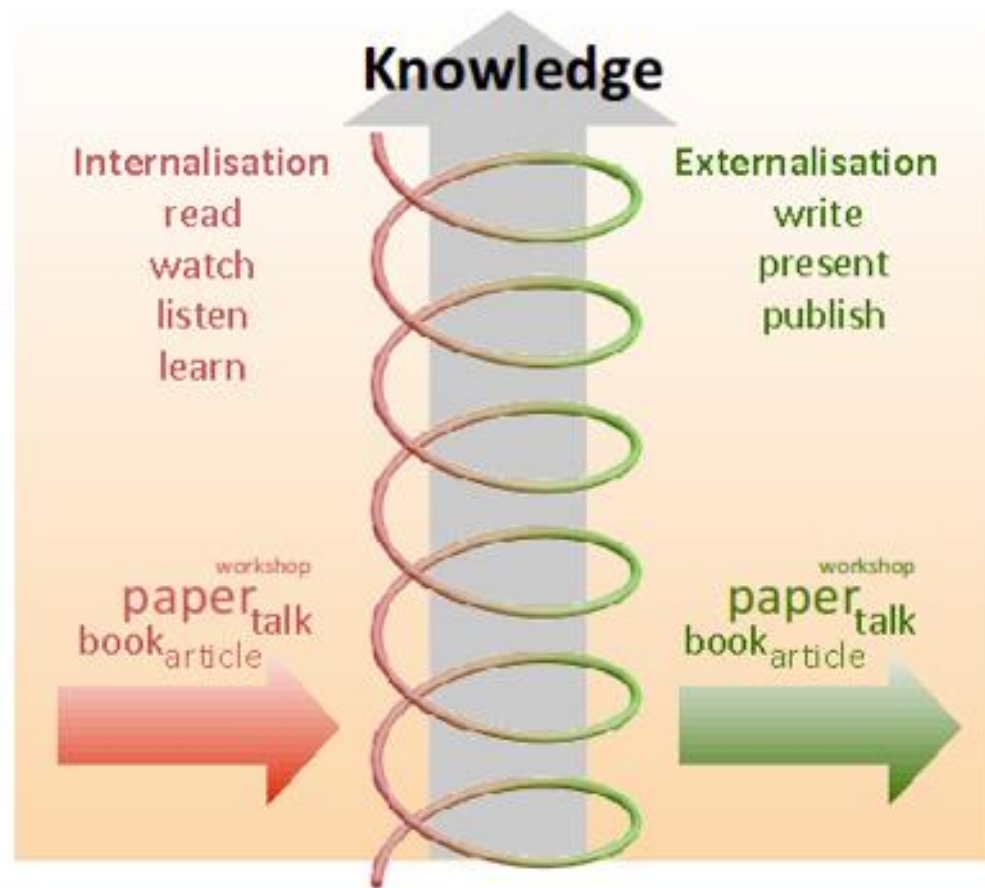
- Para conjuntos de dados em formatos de arquivo não reconhecidos pelos motores de busca a indexação é restrita aos metadados.
- Arquivos de dados - gigabytes e terabytes – problemas de espaço de armazenamento, largura de banda, backup, e custos altos.

# Processo de Internalização e Externalização do conhecimento

- Internalização descreve a transição do conhecimento explícito ao implícito - leitura de documentos ou manuais.
- Externalização é o processo de expressar conhecimento implícito em conceitos explícitos.
- A externalização é a disseminação do conhecimento pessoal - apresentações em conferências e publicação de artigos.



*Figure 1: The knowledge spiral for a researcher. (HENSE; QUADT)*



# Busca dos dados em repositórios

- Nova pesquisa começa com a busca para obter dados relevantes - ferramentas e domínio da pesquisa
- Qualidade e quantidade dos itens de dados dos repositórios
- Quantidade e qualidade da entrada de dados de outros pesquisadores.
- Atratividade do repositório - funcionalidade, design, performance, estabilidade, usabilidade, parâmetros de uso, aceitação e reputação na comunidade.

# Inspeção de dados

- O pesquisador inspeciona, visualiza e processa os dados
- Familiariza-se com os dados e aprende com eles - procedimento essencial na fase de internalização
- Transição direta do conhecimento explícito ao implícito.

# Experimento

O pesquisador aplica o conhecimento adquirido para novos problemas

Realiza novos experimentos ou processa os dados a partir de um ponto de vista diferente

Ganha novos dados que são novas contribuições para a comunidade.

# Interpretar e criar novos dados

Preparação da publicação de dados - primeira fase da externalização.

O pesquisador interpreta e cria novos dados para atrair outros pesquisadores.

Produz um paper que se baseia nos dados destaca as partes interessantes.

Registro em repositórios o arquivo de dados aumentado com novos metadados.

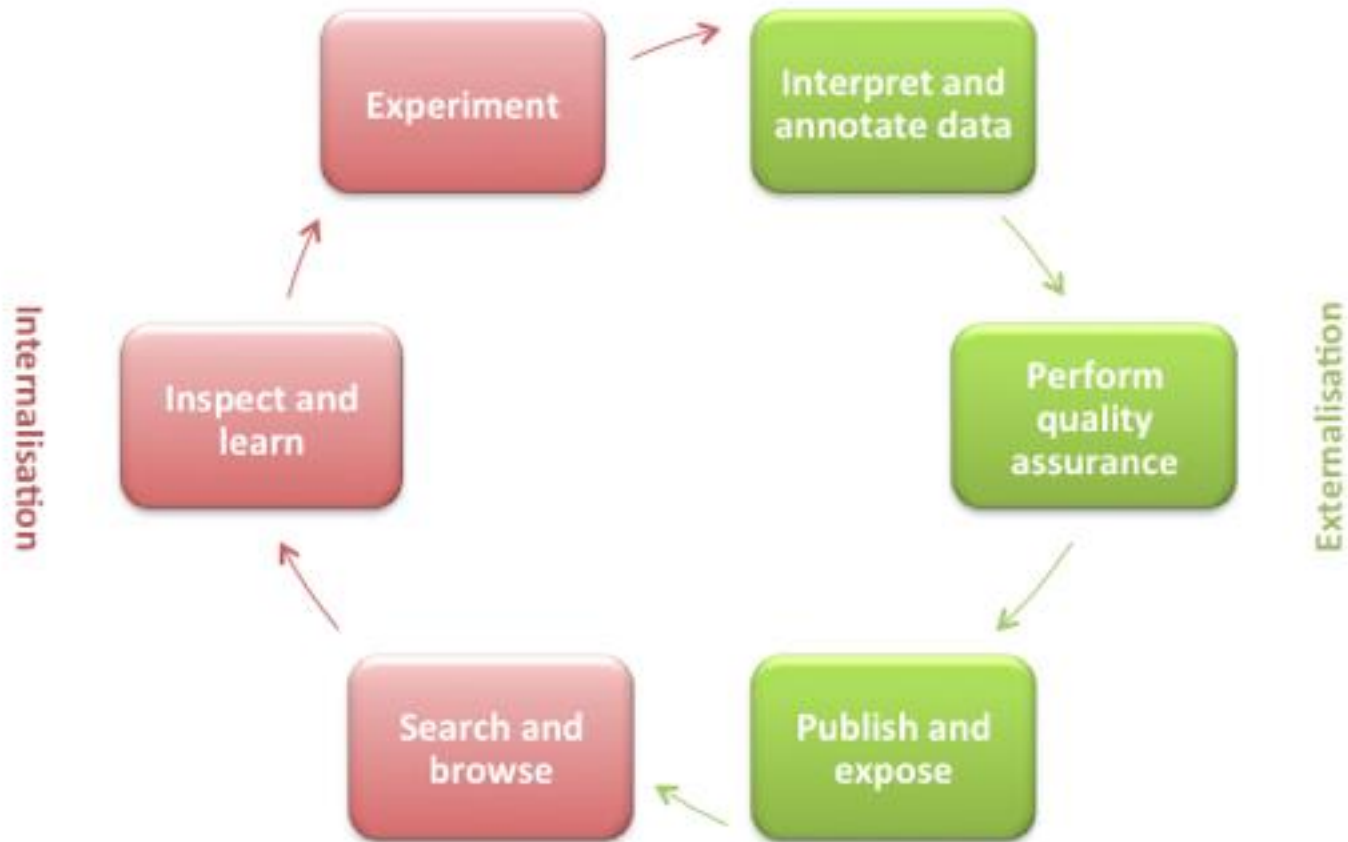
# Garantia da qualidade científica

- Formal e / técnico - garantia da qualidade pelo próprio autor - é difícil encontrar os próprios erros.
- Avaliação interpares - um ou + pesquisadores reveem o paper e os dados e podem recusar a apresentação.
- Metadados - cruciais para registrar os dados em motores de busca e catálogos de repositórios.
- A garantia de qualidade, no que diz respeito à imutabilidade após a publicação.

# Publicação

- “Publicar” - armazenar os arquivos de dados para compartilhamento público em espaço disponível a longo prazo
- Atribuir identificadores persistentes DOI ou outros.
- Expor à matrícula nos motores de busca e catálogos de repositório.
- Alterar os dados após a publicação - bloquear os dados (imutabilidade) e publicar dados corrigidos em novas versões.
- Inicia o ciclo de publicação de outro pesquisador que começa a atividade de 'pesquisar e navegar '.

*Figure 2: The data publication cycle.  
(HENSE, A.; QUADT)*





# Ferramentas de Citação

- Identificadores persistentes garantem colaboração da Ciência em nível internacional.
- Instituições nacionais se uniram em 2010 e fundaram uma iniciativa internacional - DataCite.
- DataCite - agência sem fins lucrativos – ferramenta para registrar conjuntos de dados de pesquisa e atribuir identificadores persistentes.
- Benefícios - reduzir custos de infraestrutura, integração das infraestruturas nacionais, recursos avançados de pesquisa - consistência de conjuntos de dados dos pesquisadores

# Ferramentas para Repositório de Dados

- **Harvard Dataverse Network:** <http://thedata.org/>
- Software de repositórios de dados de pesquisa,
- Para comunidades com práticas arquivísticas profissionais,
- Preservação em longo prazo,
- Controle e compartilhamento dos dados entre pesquisadores,
- Visibilidade na web e citações acadêmicas formais por suas contribuições de dados