

I ENCONTRO DA REDE BRASILEIRA DE REPOSITÓRIOS DIGITAIS

TEMÁTICA: REPOSITÓRIOS DIGITAIS E CIÊNCIA ABERTA

09 a 11/08/2022

Horário: 9h às 13h

unesp

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

GPnti

Grupo de Pesquisa
Novas Tecnologias em Informação

Coleta e reuso de dados para povoamento de repositórios digitais

Silvana Aparecida Borsetti Gregorio Vidotti

silvana.vidotti@unesp.br

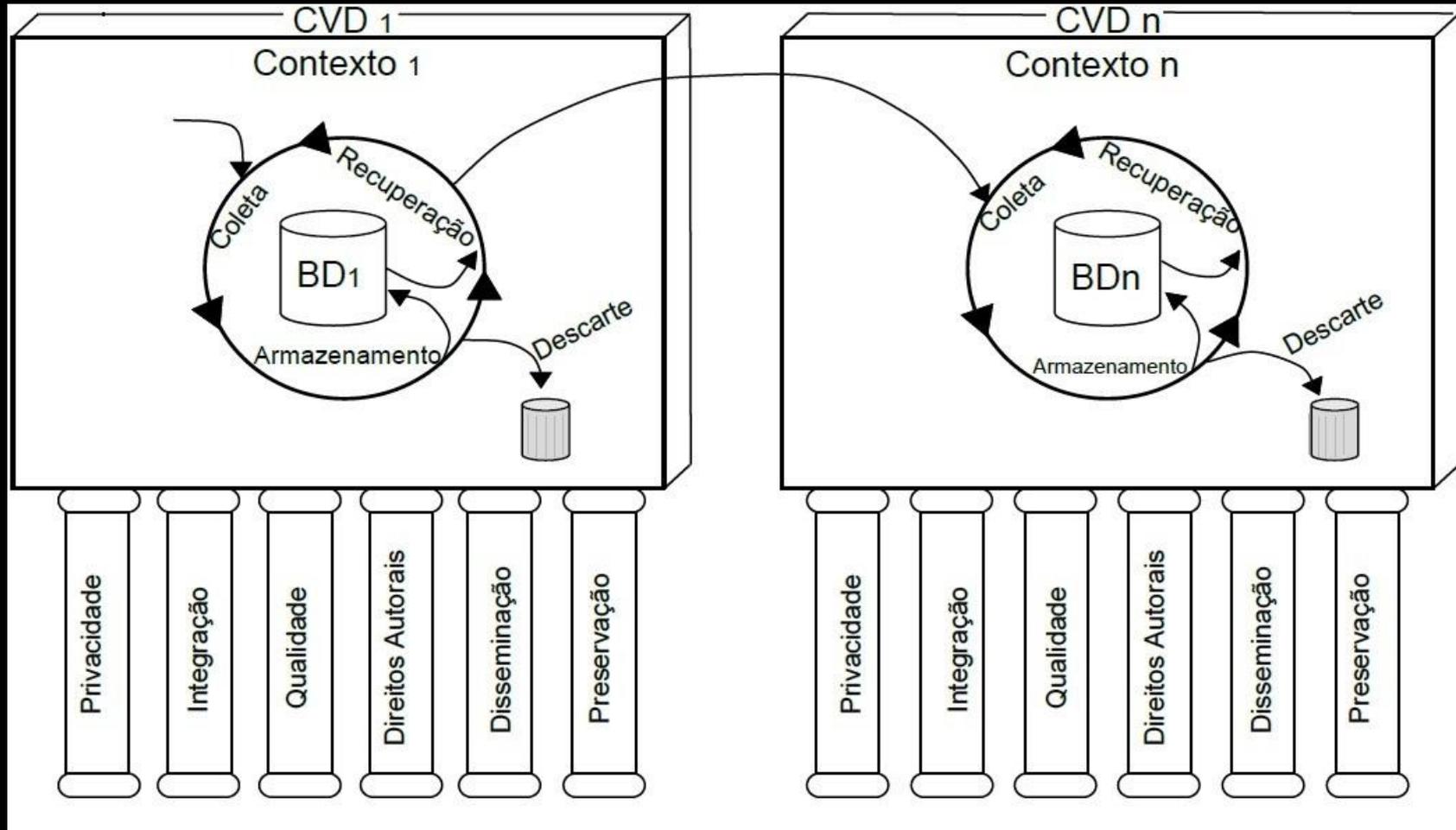


Repositórios Digitais

“Os **Repositórios Digitais** oferecem uma infraestrutura adequada para armazenar, gerenciar, reutilizar e organizar materiais/objetos digitais [...]; isso requer não apenas software e hardware, mas também políticas, processos, serviços e pessoas, além de conteúdo e metadados.”

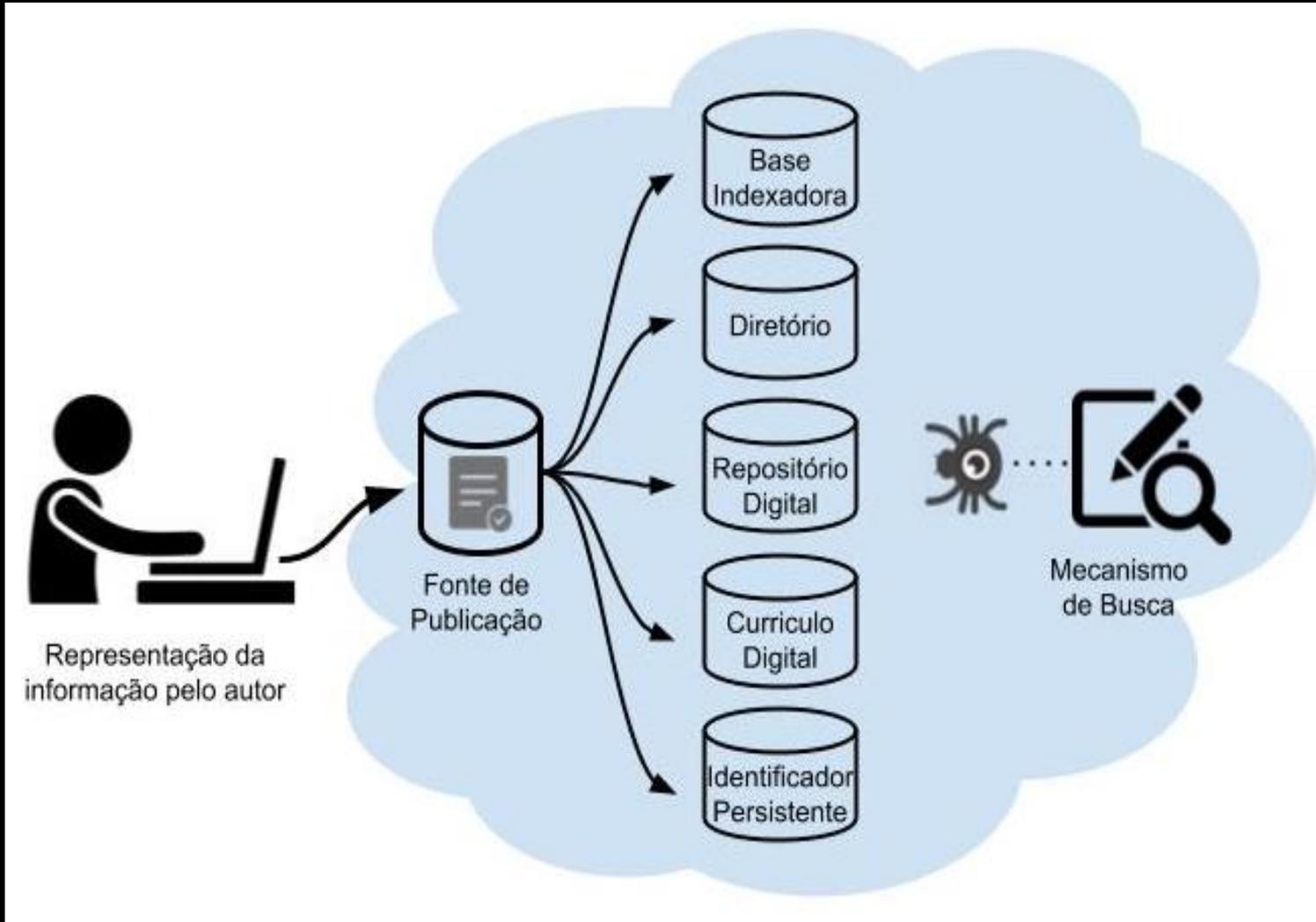
<http://www.dcc.ac.uk/sites/default/files/documents/resource/briefing-papers/digital-repositories.pdf>

Ciclo de Vida dos Dados para a Ciência da Informação (CVD-CI)

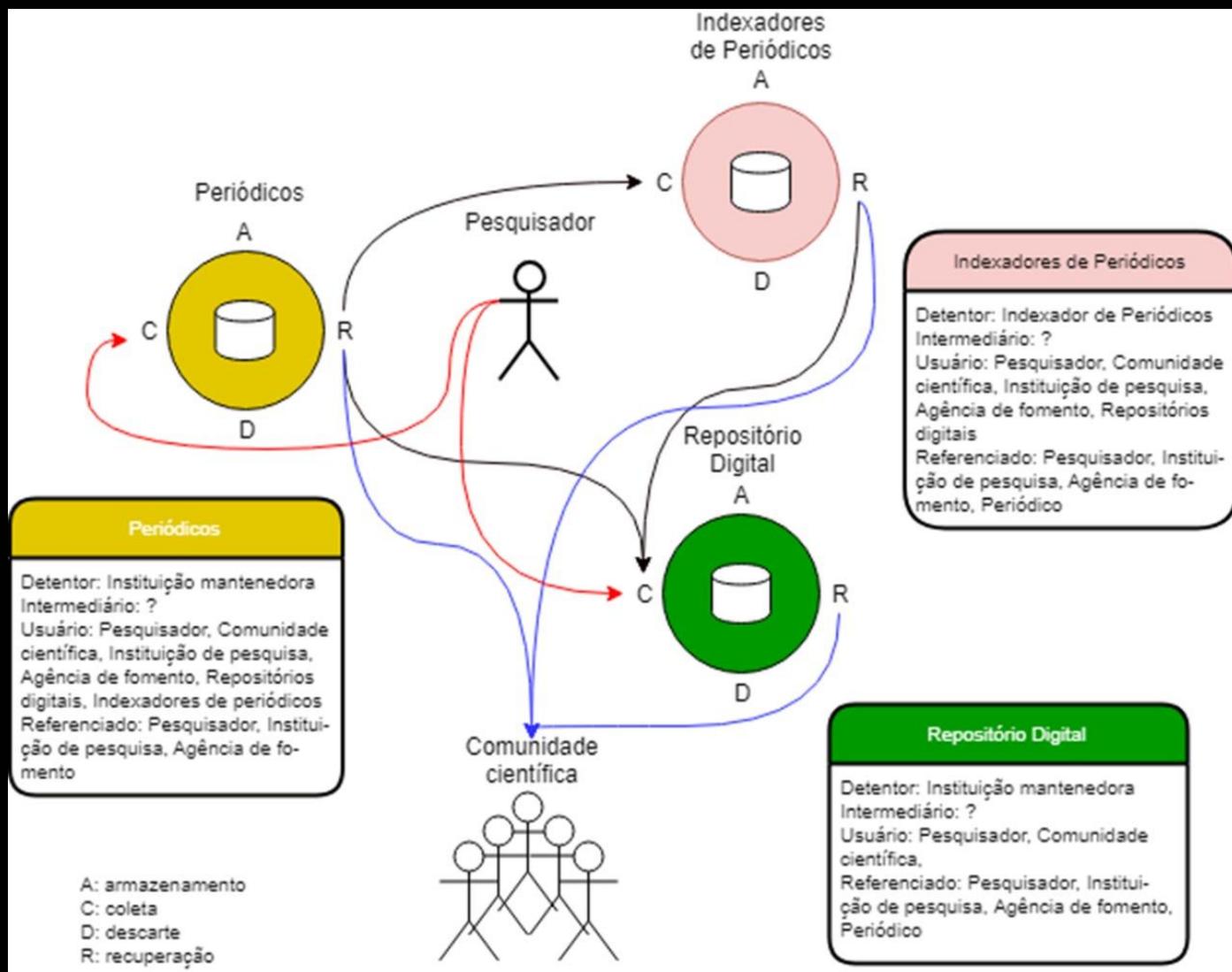


(SANTANA, 2016)

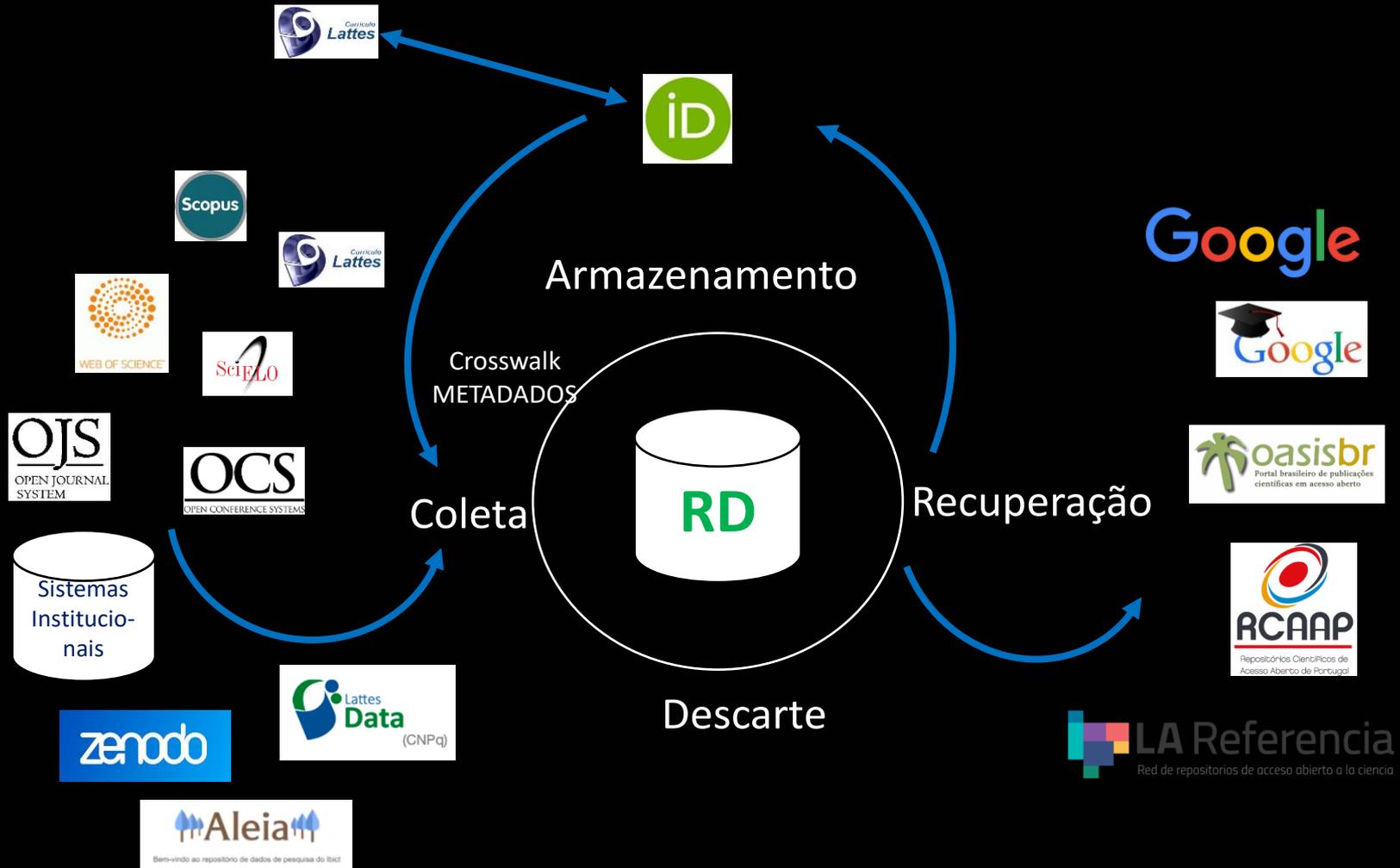
Autor - representação



CVD e o processo de publicação e acesso



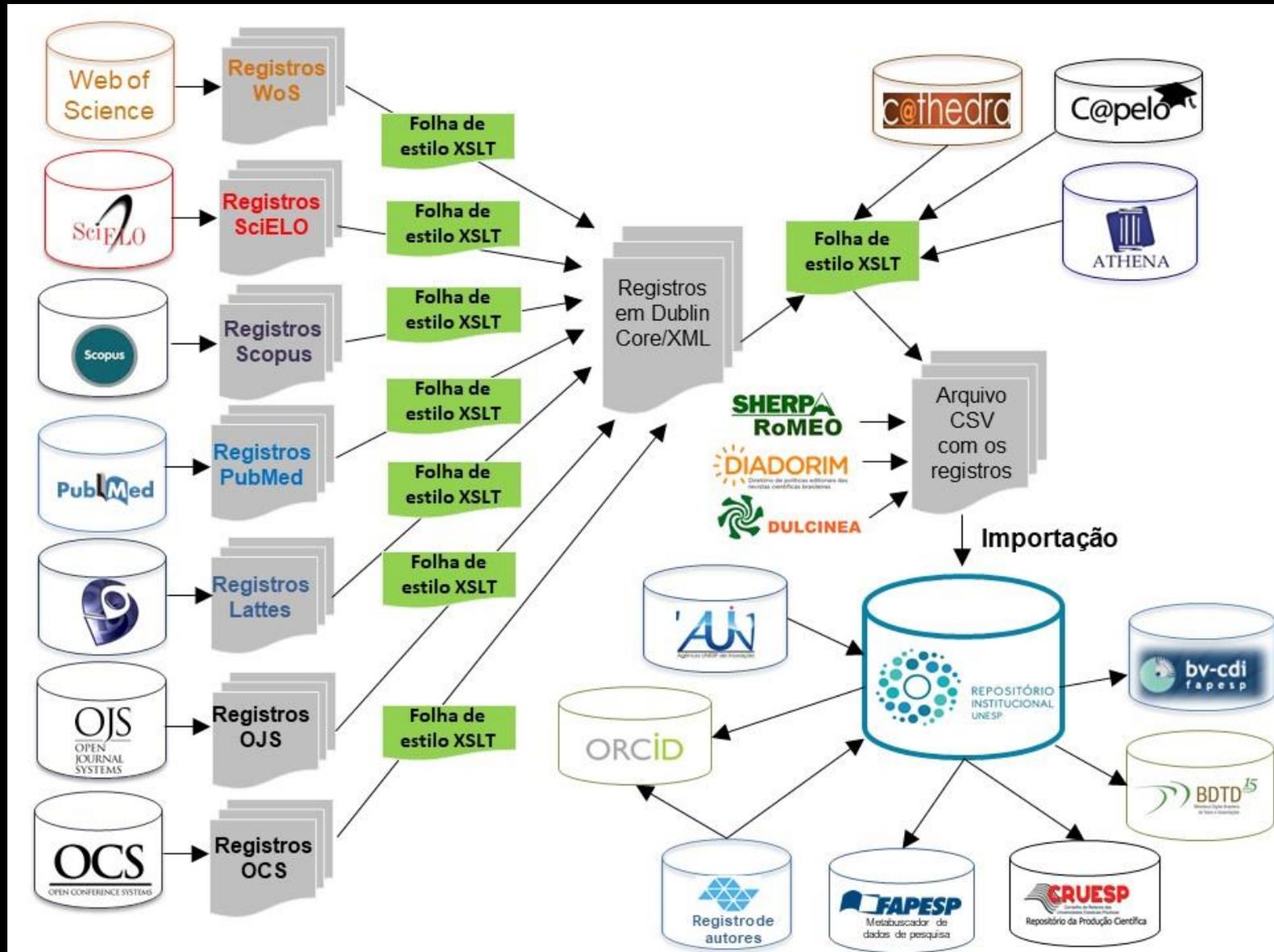
Repositório Digital (RD)



I ENCONTRO DA REDE BRASILEIRA DE REPOSITÓRIOS DIGITAIS

TEMÁTICA: REPOSITÓRIOS DIGITAIS E CIÊNCIA ABERTA

Repositório Institucional Unesp



Procedimentos - RI Unesp

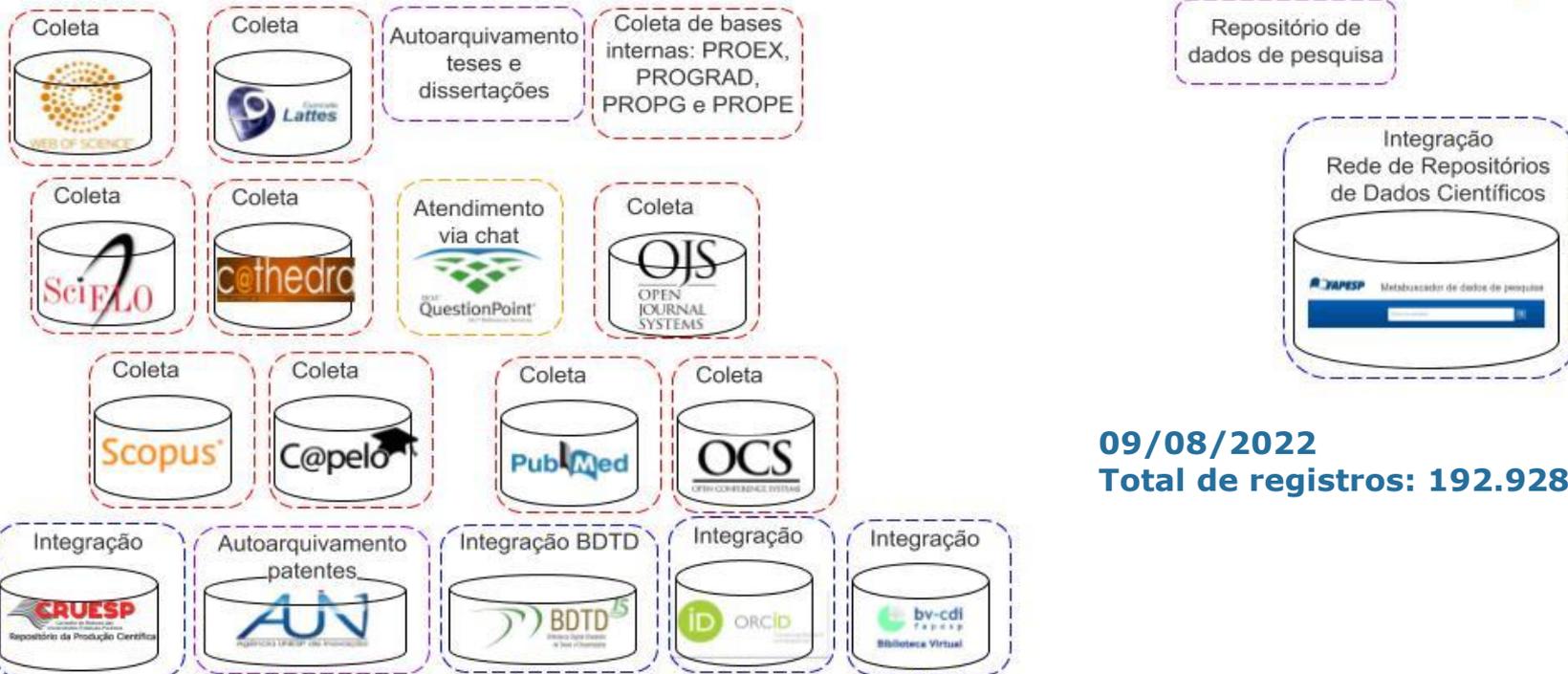
1. Coleta dos registros (via interface ou em lote)
2. Estudo da estrutura do XML
3. Conversão dos registros XML em DSpace-XML
4. Conversão dos registros DSpace-XML para DSpace-CSV
5. Etapa de deduplicação dos registros
6. Verificação manual de cada registro
7. Coleta do PDF
8. Mapeamento dos registros com as coleções
9. Importação dos registros no Repositório

Repositório Institucional Unesp

Linha do tempo dos serviços RI UNESP

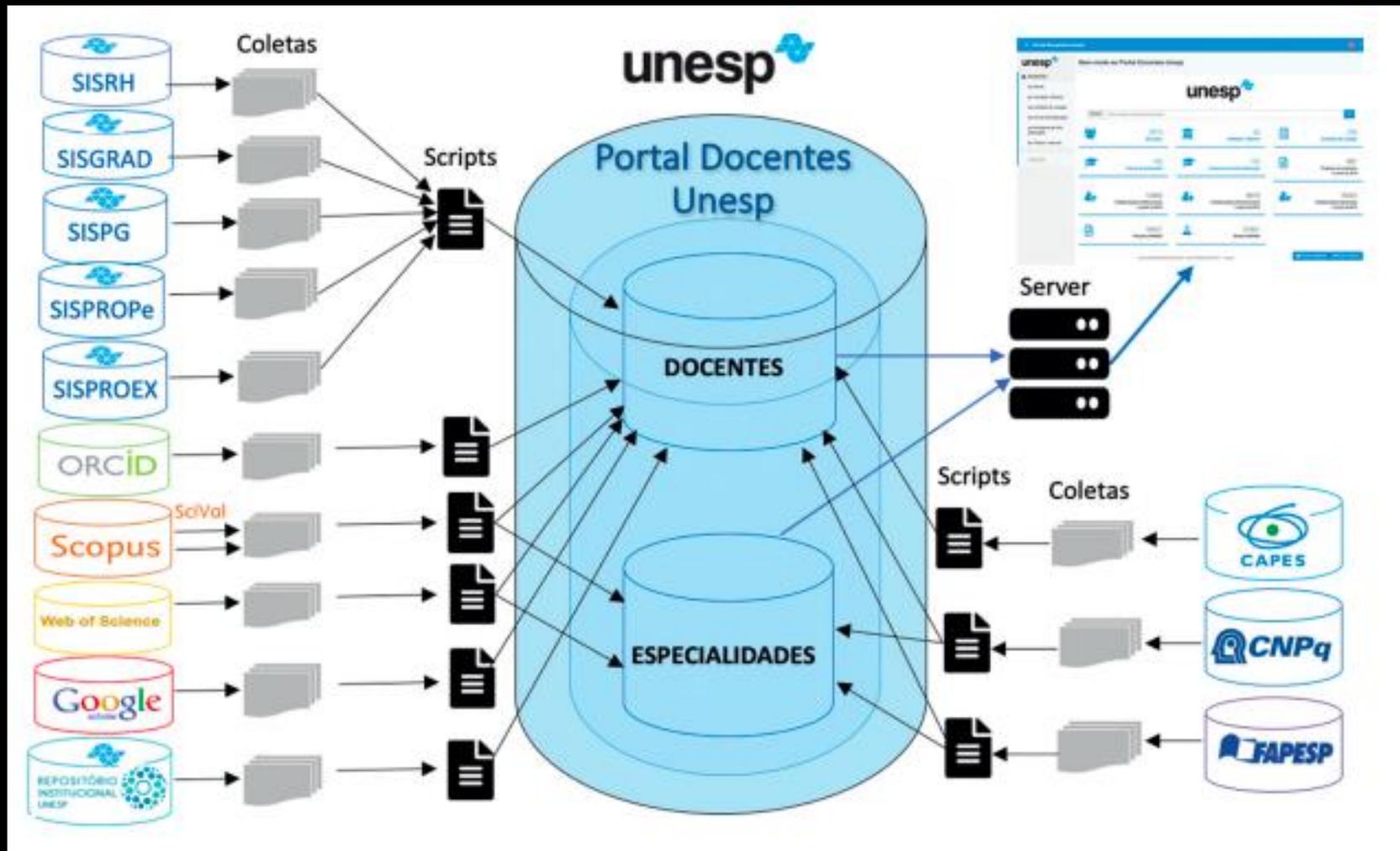


| 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------------------|------------------|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 13.933 registros | 72.000 registros | 91.000 registros | 104.571 registros | 111.426 registros | 138.066 registros | 140.719 registros | 150.340 registros |

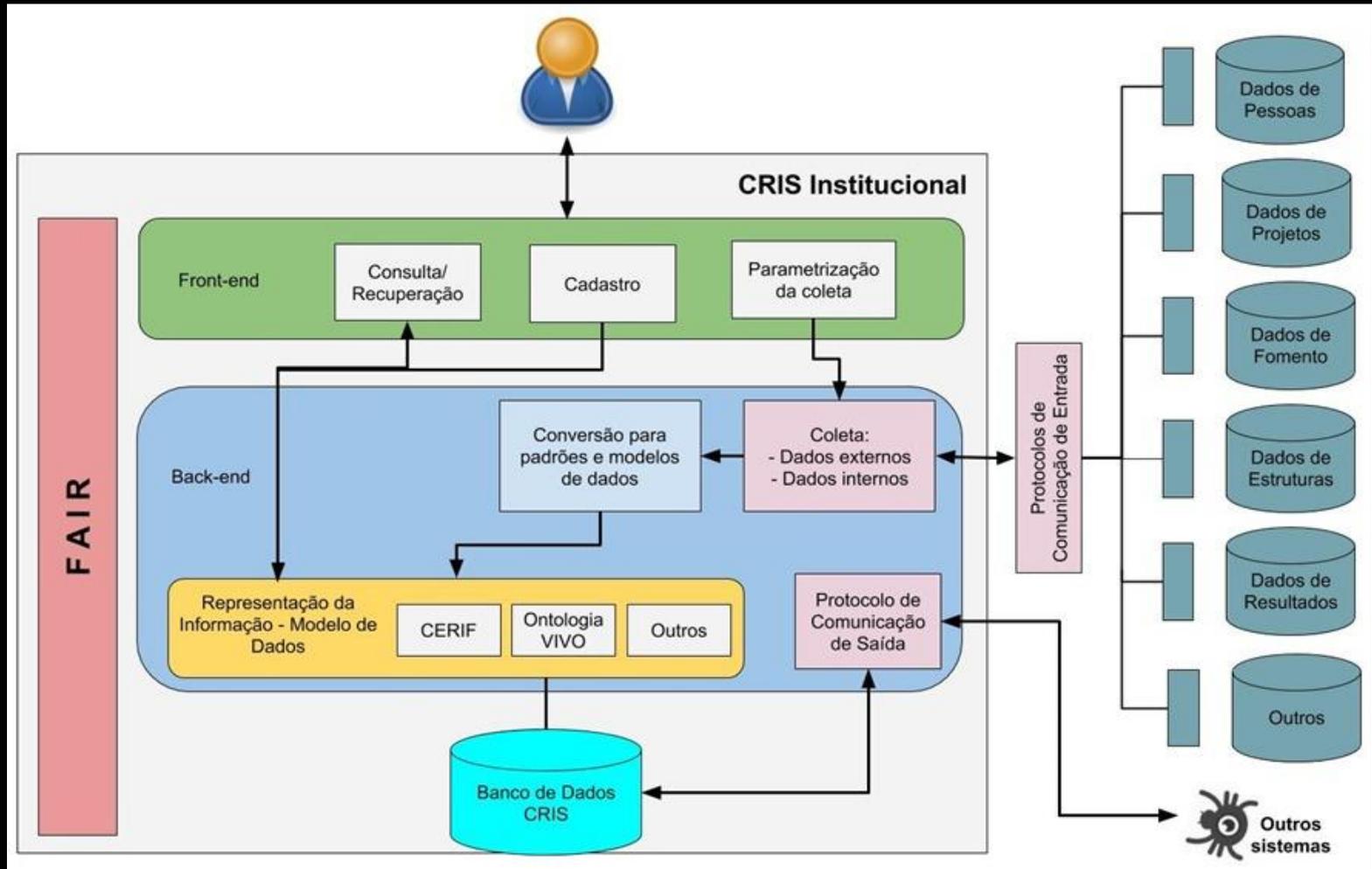


09/08/2022
Total de registros: 192.928

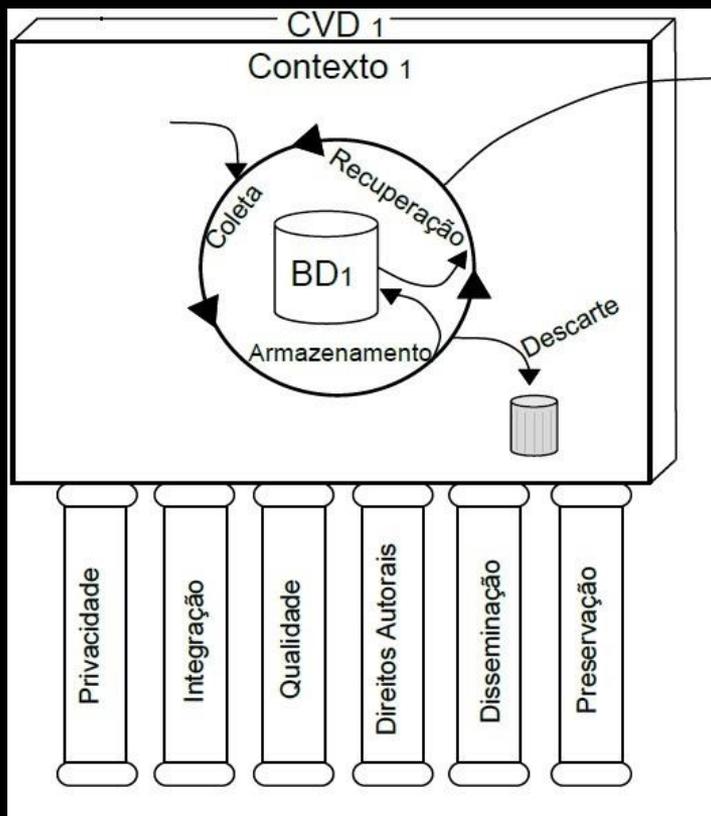
Ecossistema de dados



Ecosistema de pesquisa



CVD – coleta e reuso de dados



- Mapeamento das fontes de coleta
- Qualidade dos metadados
- Crosswalk de metadados
- Compatibilização com perfil de aplicação
- Preparação de scripts, folhas de estilo, webservices e APIs
- Conversão dos registros
- Verificação das questões legais (LGPD e direitos autorais)
- Definição de protocolos de comunicação (p. ex. OAI-PMH)
- Preservação digital

Referências

SANTANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, v. 21, n. 2, p. 116-142, 2016. DOI: [10.5433/1981-8920.2016v21n2p116](https://doi.org/10.5433/1981-8920.2016v21n2p116)

TORINO, Emanuelle; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Representação da informação e interoperabilidade entre ambientes científicos: um enfoque ao papel do autor. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 8. 2019, Londrina. **Anais** [...] Londrina: Universidade Estadual de Londrina, 2019. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2019/secin2019/paper/viewFile/602/410>

TORINO, Emanuelle; VIDOTTI, Silvana Aparecida Borsetti Gregorio; SANT'ANA, Ricardo César Gonçalves. Ciclo de vida de dados no processo de publicação e acesso à produção científica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20. 2019, Florianópolis. **Anais** [...] Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/576/612>.

TORINO, Emanuelle; CONEGLIAN, Caio Saraiva; VIDOTTI, Aparecida Borsetti Gregorio Vidotti. Estruturas de representação para reuso de dados no contexto da ecologia de pesquisa: CRIS Institucional. *Inf. Inf.*, Londrina, v. 25, n. 3, p. 1-27, jul./set. 2020. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/41946>. Acesso em: 03 nov. 2020.

I ENCONTRO DA REDE BRASILEIRA DE REPOSITÓRIOS DIGITAIS

TEMÁTICA: REPOSITÓRIOS DIGITAIS E CIÊNCIA ABERTA

09 a 11/08/2022 | Horário: 9h às 13h



unesp 

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

GPnti
Grupo de Pesquisa
Novas Tecnologias em Informação

Grata!

Silvana Aparecida Borsetti Gregorio Vidotti
silvana.vidotti@unesp.br