



Desambiguação de Nomes de Autores: O Desafio dos Repositórios

Maria Eduarda Pereira Rodrigues
António Moitinho Rodrigues

Sumário

- ❖ **Contexto e Motivação**
- ❖ **Âmbito e objetivos**
- ❖ **Material e métodos**
- ❖ **Resultados**
- ❖ **Considerações finais (NP)**



❖ Contexto e Motivação

Nos Repositórios Institucionais a desambiguação dos nomes do autores e o controlo de autoridade constituem um enorme desafio no contexto da gestão da informação.



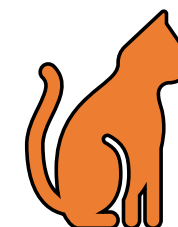
❖ Contexto e Motivação

Repositório Científico do
Instituto Politécnico de
Castelo Branco (RCIPCB)

Normalização da forma do
nome dos autores no
RCIPCB

Revisão bibliográfica

Desambiguação de nomes
Controlo de autoridade
Uso de Identificadores



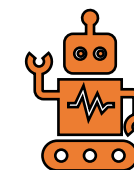
Verificar a situação em
outras IES congéneres



❖ Âmbito e Objetivos



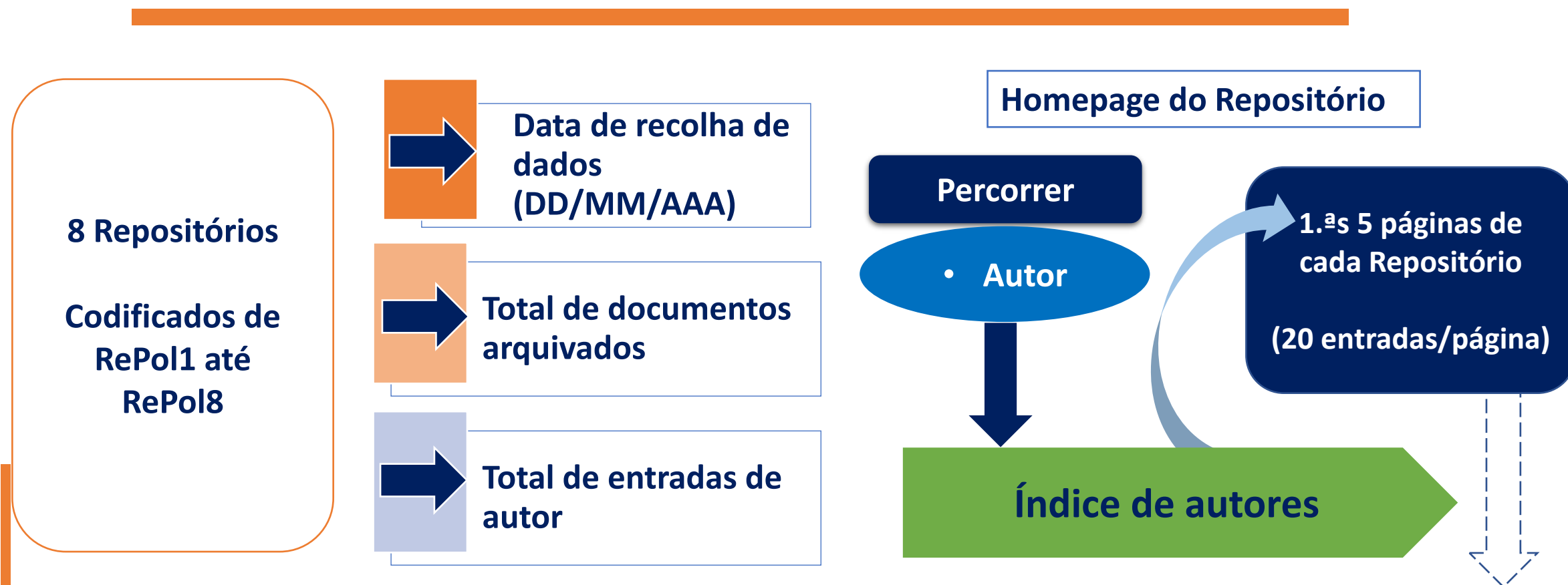
Âmbito – Índice de autores dos Repositórios de Instituições Portuguesas de Ensino Superior Politécnico portuguesas (IESP) presentes no portal do Repositório Científico de Acesso Aberto de Portugal (RCAAP).



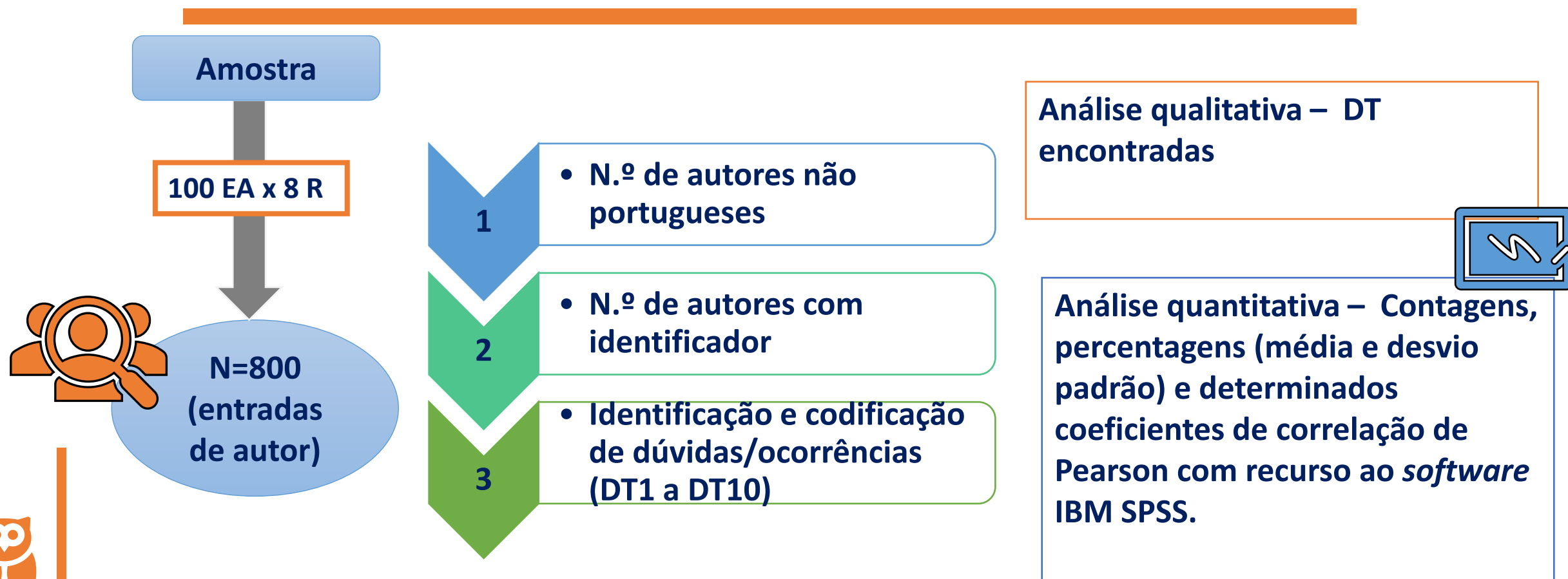
Objetivos – Através da consulta do índice de autores verificar se os dados disponíveis revelam ou não a existência de controlo de autoridade de nomes de autores, se existe efetiva desambiguação dos mesmos e propor eventuais soluções de melhoria.



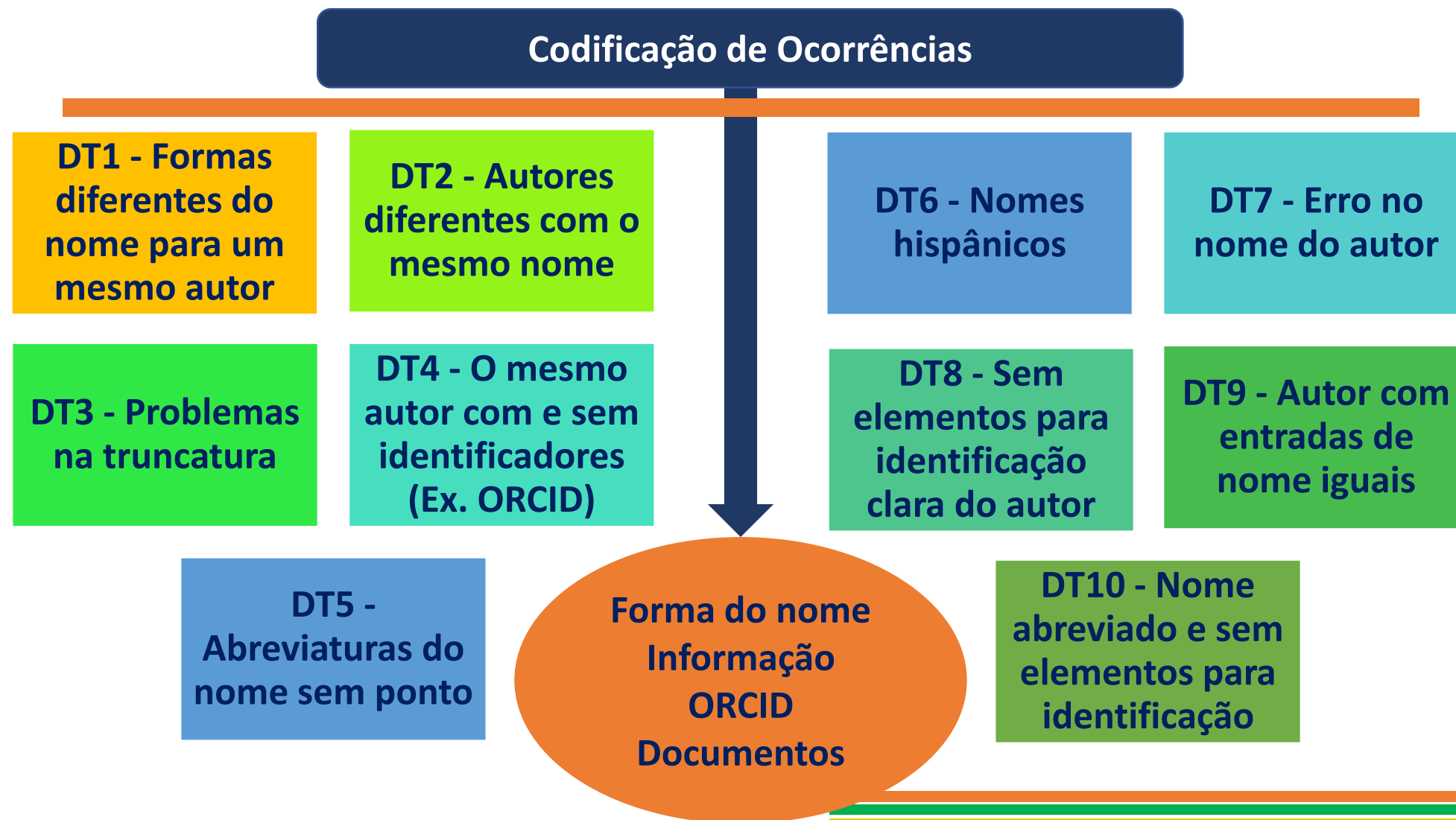
❖ Material e Métodos



❖ Material e Métodos



❖ Material e Métodos



❖ Resultados

Total de documentos arquivados/ RePol

Documentos por entradas de autor/RePol

Código RI	Data recolha dados	Registo total documentos	Número documentos por entrada autor
RePol1	02/04/2024	27031	1,55
RePol2	02/04/2024	8684	0,84
RePol3	02/04/2024	8533	1,29
RePol4	03/04/2024	16283	0,79
RePol5	03/04/2024	4257	0,78
RePol6	03/04/2024	6769	0,87
RePol7	04/04/2024	22833	0,91
RePol8	05/04/2024	2118	0,75
	Valor médio (\pm dp) dos 8 RI	12064	0,97 (\pm 0,29)

\pm dp – desvio padrão da amostra. Fonte: Autores



❖ Resultados

Código RI	% Autores sem ID	% Autores c/ ORCID	% Autores c/ CIENCIA ID	Entradas Autor com Nomes não Portugueses
RePol1	95,0%	100,0%	60,0%	68
RePol2	77,0%	95,7%	68,2%	30
RePol3	92,0%	100,0%	37,5%	15
RePol4	82,0%	100,0%	22,2%	33
RePol5	94,0%	100,0%	16,7%	28
RePol6	92,0%	87,5%	57,1%	29
RePol7	92,0%	100,0%	37,5%	33
RePol8	100,0%	0,0%	0,0%	19
	90,5%	85,4%	37,4%	31,9
	(±0,07)	(±0,35)	(±0,24)	(±15,97)



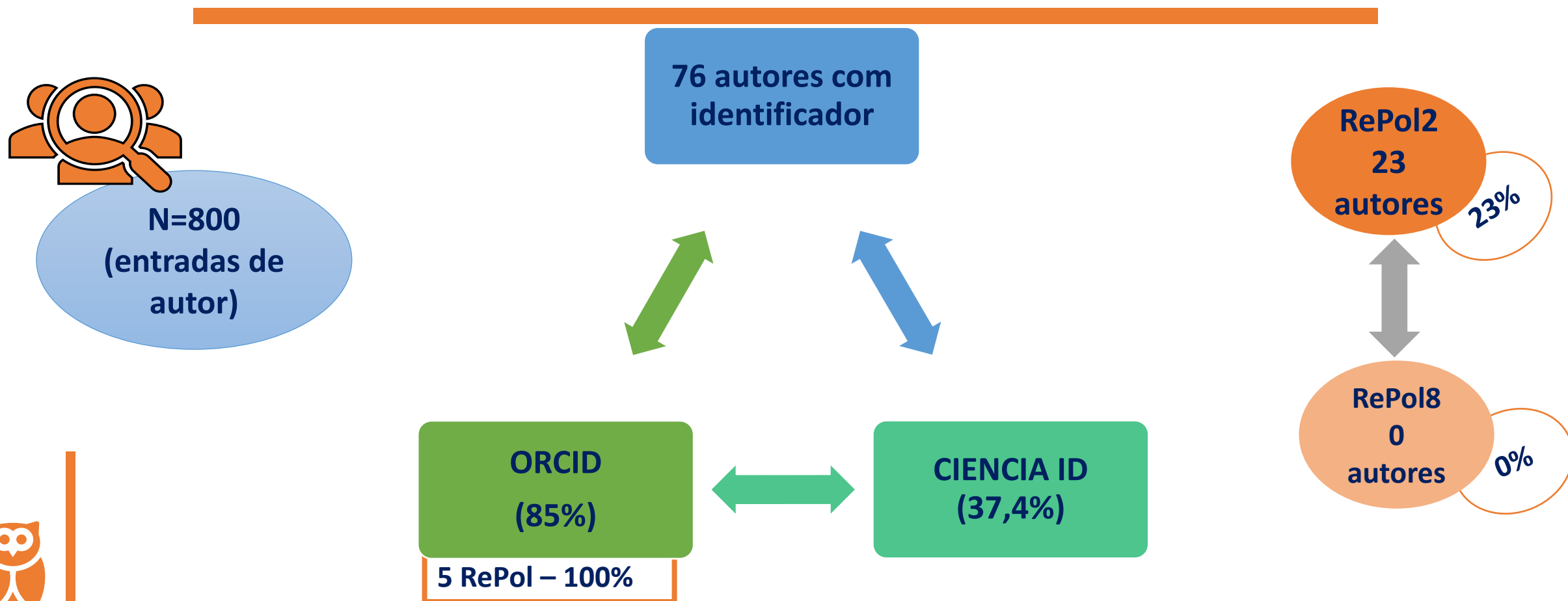
N=800
(entradas de autor)

255



±dp – desvio padrão da amostra. Fonte: Autores

❖ Resultados



❖ Resultados

Nomes Não Portugueses



N=800
(entradas
de autor)

Em média 90,5%
dos autores
amostrados não
possui qualquer
identificador

255 autores

Valor
médio
31,9/RePol

RePol1

68

RePol3

15

Menor número de
DT identificadas



❖ Resultados

DT	RePol1	RePol2	RePol3	RePol4	RePol5	RePol6	RePol7	RePol8	Total
DT-1 – Formas diferentes do nome	1	3	2	3	11	9	6	15	50
DT-2 – Autores com o mesmo nome	0	1	1	0	0	0	1	0	3
DT-3 – Problema na truncatura	0	3	0	2	4	4	5	0	18
DT-4 – Autor com e sem identificador	0	5	0	2	3	3	4	0	17
DT-5 – Nome abreviado sem ponto	0	3	0	2	0	1	1	2	9
DT-6 – Nomes hispânicos	0	0	0	0	1	0	2	1	4
DT-7 – Erro no nome do autor	0	0	0	2	1	2	1	2	8
DT-8 – Sem elementos para identificação	0	1	0	1	1	1	2	2	8
DT-9 – Autor com entradas de nomes iguais	0	0	0	0	0	0	5	1	6
DT-10 – Nome abreviado e sem elementos	0	0	1	0	0	0	0	0	1
Total	1	16	4	12	21	20	27	23	124

Fonte: Autores



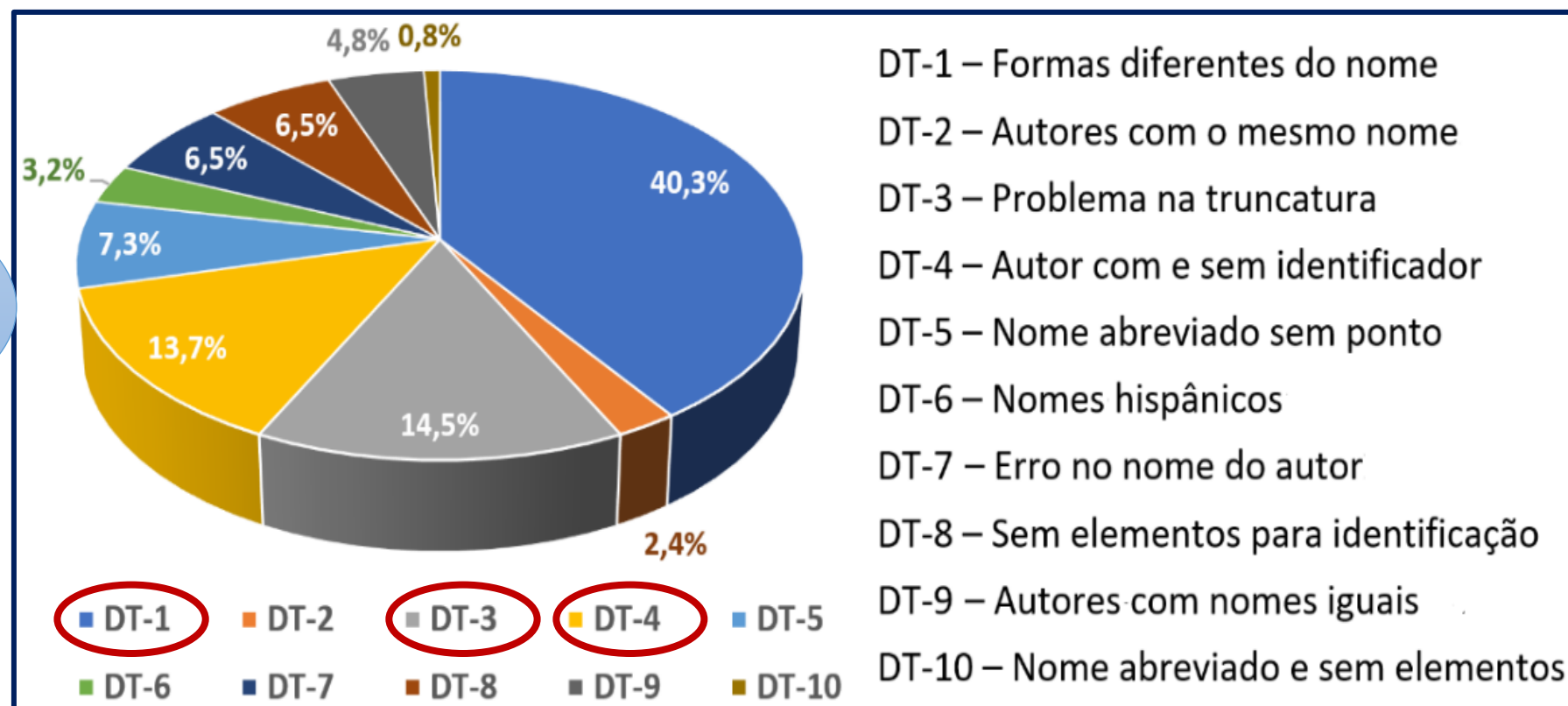
N=800
(entradas
de autor)



❖ Resultados



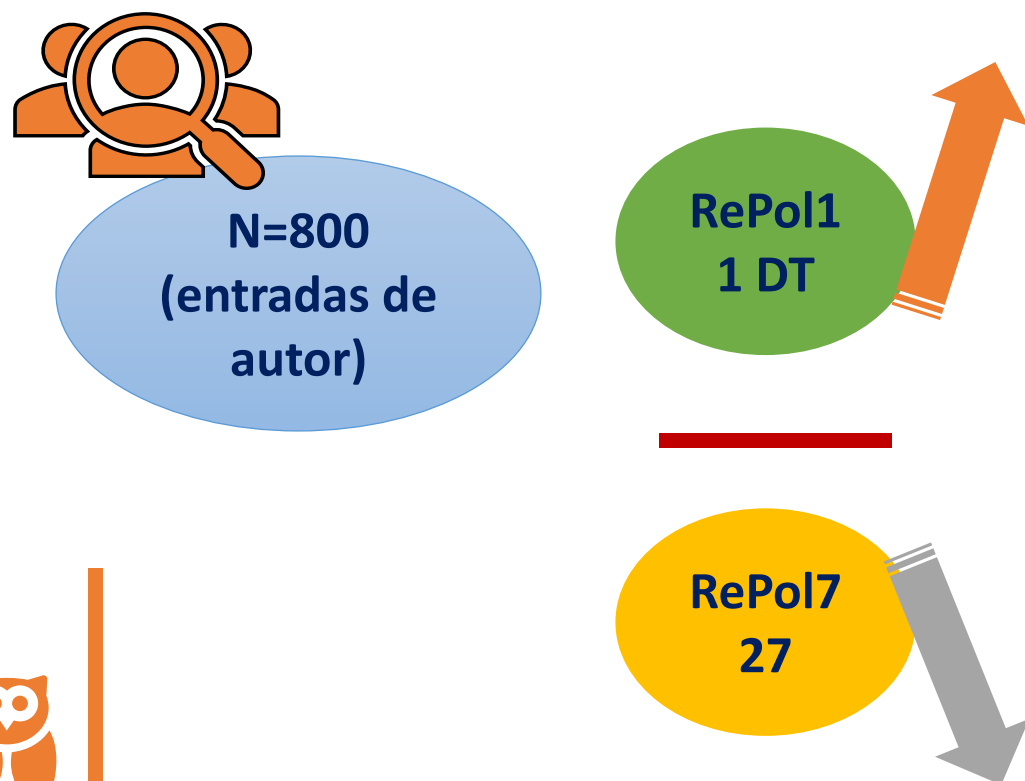
N=800
(entradas de autor)



Fonte: Autores



❖ Resultados



Baixa correlação entre a dimensão do repositório e o número de DT encontradas.

Os dados sugerem a existência de alguns mecanismos de controlo de autoridade em repositórios com maior número de documentos arquivados.

Parece haver maior grau de normalização nos registos de autores com nomes não portugueses.



❖ Resultados

✓ **Verificam-se dúvidas e ocorrências e todos os repositórios**

A maioria dos autores não possui identificador e quando existe o ORCID é o identificador mais utilizado seguido do Ciência ID

Existem muitos autores não afiliados às instituições donas dos Repositórios

Existem autores de registo único e outros que não é possível desambiguar através dos elementos disponíveis ou por falta de outros elementos ou documentos no repositório



❖ Considerações Finais

O estudo realizado permite afirmar que, no contexto do controlo de autoridade nos repositórios, a utilização de identificadores digitais para os autores é muito vantajosa porque, de um modo geral, permite desambiguar e identificar claramente os autores.

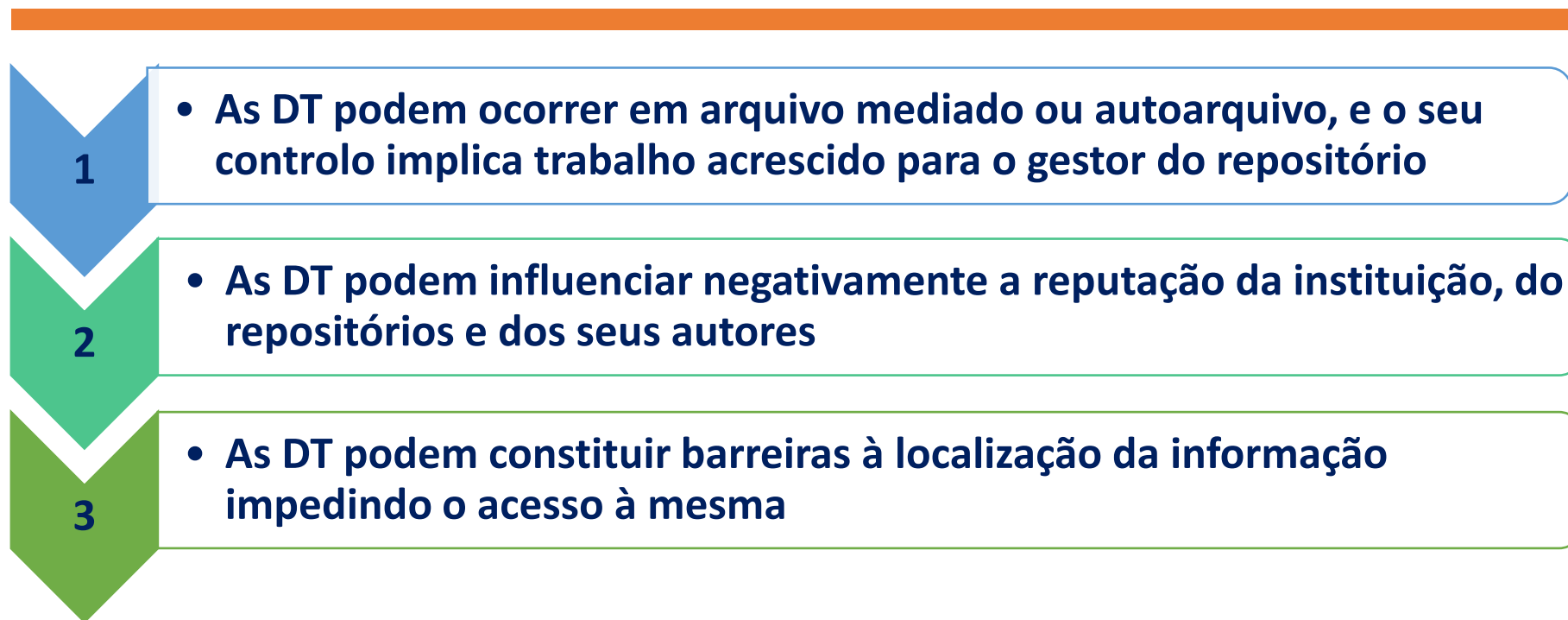
O ORCID é o identificador que parece adequar-se melhor a esta função.

A interoperabilidade entre os sistemas (ORCID, CIENCIA ID) é um fator muito importante no contexto do controlo de autoridade.

O controlo de autoridade garante transparência, rigor e confiabilidade no processo de arquivamento, pesquisa e extração de informação e documentos dos Repositórios.



❖ Considerações Finais



NP - Segundo alguma literatura consultada, a multiplicidade de operadores do repositório incluindo a prática de autoarquivo, poderá estar associada à dificuldade de estabelecer formas únicas para os nomes dos autores.



❖ Considerações Finais

DT1 - Formas diferentes do nome para um mesmo autor

1 Reis, Ana Paula Lopes | 2 Reis, A. Lopes
3 Reis A. P. Lopes | 4 Reis, A.P.L. | 5 Reis, Ana Paula
6 Reis, Ana Lopes | 7 Reis, A. , etc.

DT3 - Problemas na truncatura

1 Dias Castro, José Manuel | 2 Castro, José Manuel Dias | 3 Dias Castro, J.M.

1 Lopez Benitez, Joaquín | 2 L. Benitez, Joaquín | 3 Benitez, Joaquín Lopez
4. Lopez B., Joaquín

DT4 - O mesmo autor com e sem identificadores

Reis, Ana Paula Lopes · -----6
Reis, Ana Paula Lopes-----4





Desambiguação de Nomes de Autores: O Desafio dos Repositórios

Maria Eduarda Pereira Rodrigues
António Moitinho Rodrigues