

# Calidad en repositorios digitales en Argentina, estudio comparativo y cualitativo



**UNJu**

Universidad  
Nacional de Jujuy

***Dr. J. Federico Medrano***

@fedemedrano



## Repositorios digitales

- Los repositorios han experimentado gran desarrollo en los últimos años (64% en 6 años)
  - ROAR (Registry of Open Access Repositories)
  - 2707 en 2011
  - 4446 en 2017
- Las instituciones reconocen su utilidad:
  - como vía de difusión de resultados de investigación científica
  - como elemento de preservación digital



## Protocolo OAI-PMH

---

- ⦿ Open Archives Initiative Protocol for Metadata Harvesting
- ⦿ Permite el acceso a los metadatos que describen cada uno de los registros de un repositorio
- ⦿ Se trata de un acceso a través de HTTP
- ⦿ Se trata de un acceso abierto (puede limitarse en función de sobrecargas de servidores y circunstancias afines)
- ⦿ Para recuperar registros (documentos) por los usuarios



# Actores



Data providers



Services providers

*El protocolo OAI-PMH es probablemente el más utilizado para brindar interoperabilidad desde el repositorio hacia el exterior*



“



## Proceso de recolección

---



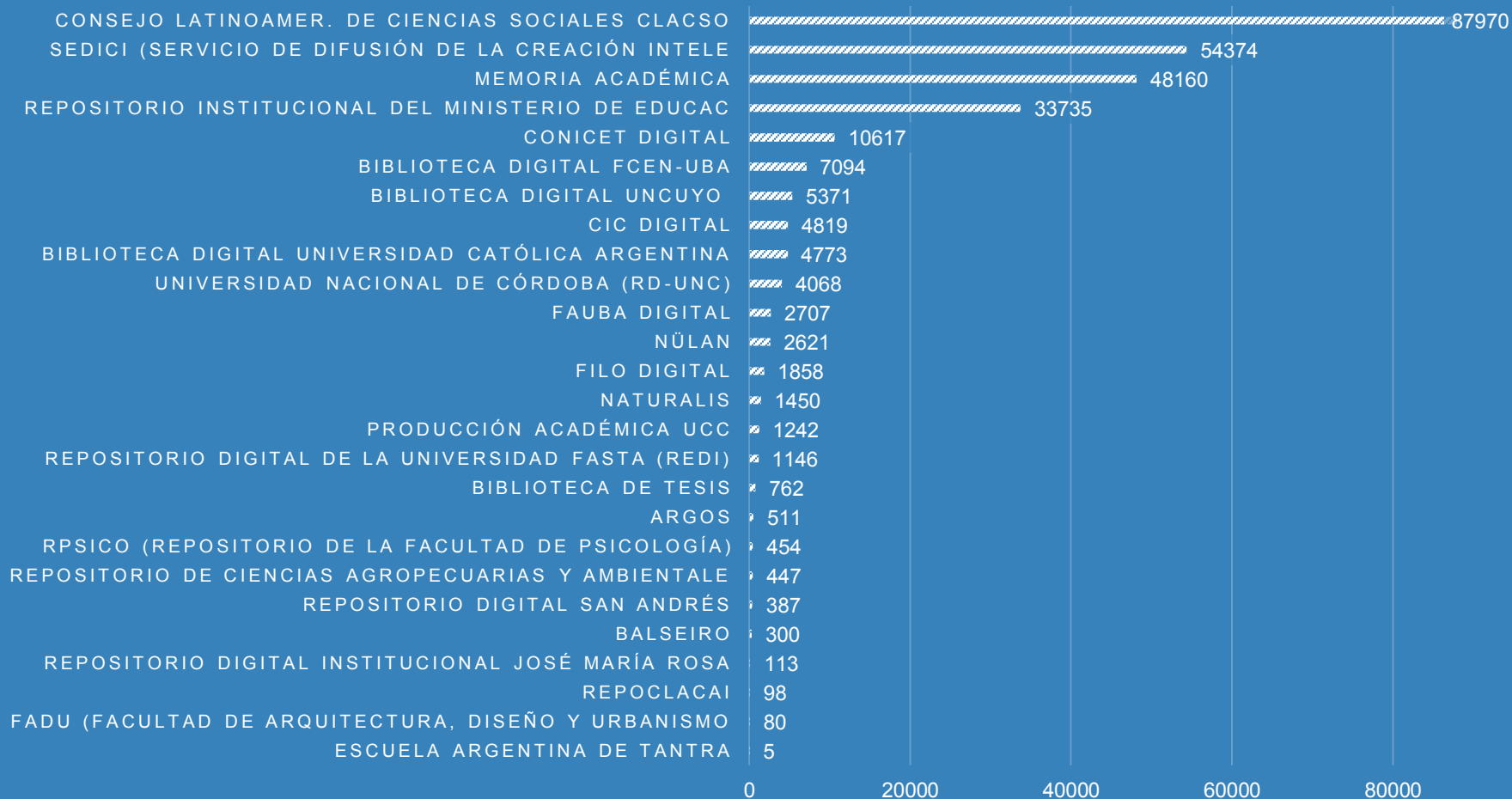


## Datos recolectados

---

- Los datos se recolectaron en Mar-2017
- ROAR posee alrededor de 4446 repositorios registrados
- De ellos, 47 pertenecen a instituciones Argentinas
- Eliminando duplicados, no operativos o simplemente inaccesibles, se han recolectado 26 repositorios
- Total de registros 275162

# MEDIA DE REGISTROS POR REPOSITORIO 10583



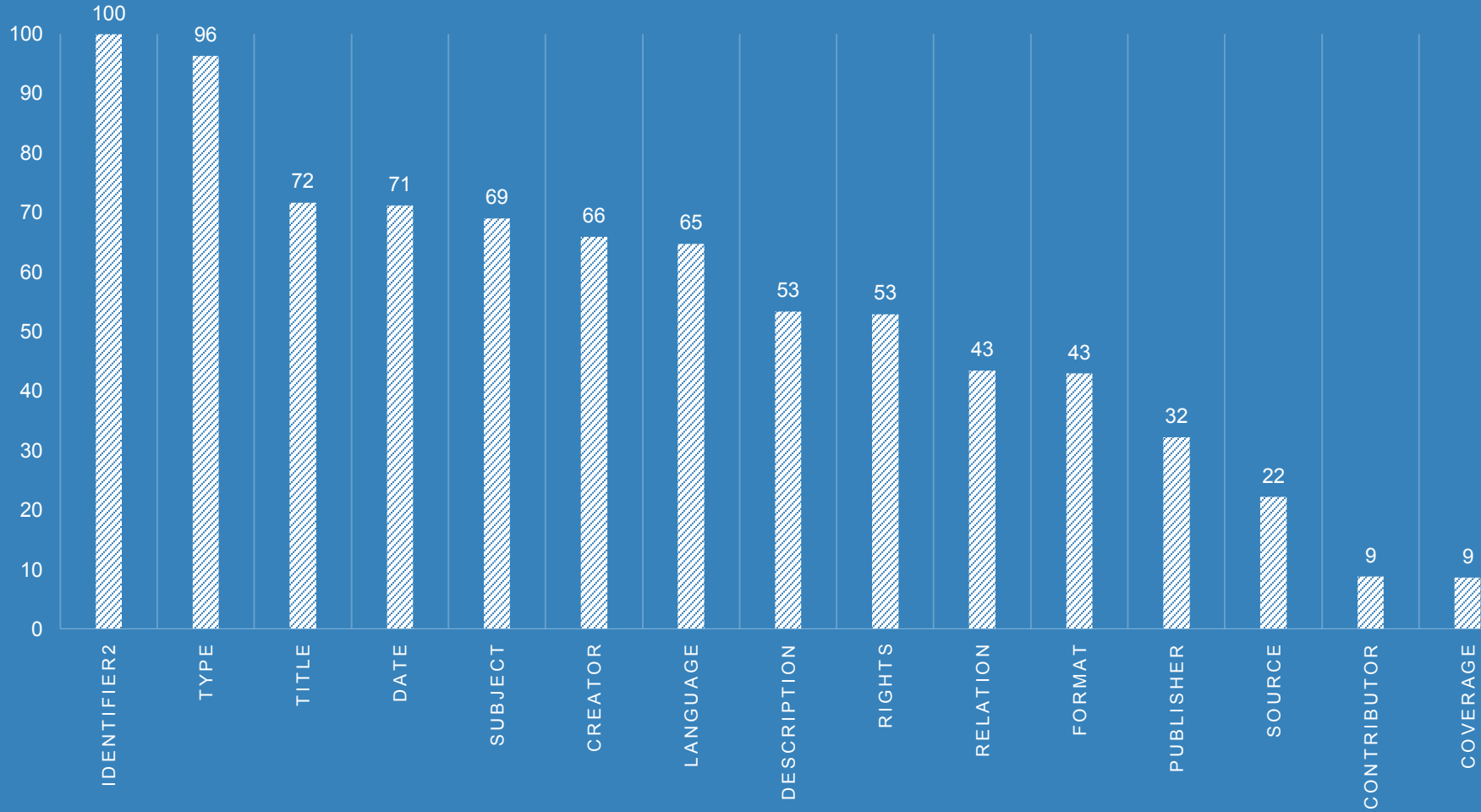


	title	creator	subject	description	publisher	contributor	date	type	format	source	language	relation	coverage	rights	identifier2
ESCUELA ARGENTINA DE TANTRA	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Fadu (Facultad de Arquitectura, Diseño y Urbanismo)	100	100	98	98	0	0	100	100	99	0	0	100	0	0	100
RepoCLACAI	100	98	98	90	66	1	100	93	0	0	99	23	0	0	100
Repositorio Digital Institucional José María Rosa	100	100	100	100	95	1	100	100	100	0	99	0	0	100	100
Balseiro	99	100	100	98	6	0	97	100	100	0	100	100	0	0	100
Repositorio Digital San Andrés	100	100	86	100	2	1	100	100	0	0	100	0	0	100	100
Repositorio de Ciencias Agropecuarias y Ambientale	100	100	12	46	98	45	100	100	100	0	0	100	0	0	100
RPsico (Repositorio de la Facultad de Psicología)	79	77	78	74	1	69	79	79	0	0	79	0	0	13	79
Argos	100	100	99	99	100	0	100	100	100	11	99	91	87	100	100
Biblioteca de tesis	100	100	100	100	0	100	100	100	100	0	100	0	0	100	100
Repositorio Digital de la Universidad FASTA (REDI)	100	99	62	79	1	44	100	100	82	38	85	0	4	98	100
Producción Académica UCC	100	100	100	78	32	2	93	100	98	0	98	100	0	83	100
Naturalis	100	100	100	13	24	14	100	100	100	62	100	12	0	100	100
Filo Digital	100	94	74	92	97	68	100	100	99	27	100	27	0	91	100
Núlan	100	98	79	87	44	18	100	99	86	74	86	100	0	86	100
FAUBA Digital	100	100	89	99	100	30	100	100	100	70	34	0	0	100	100
Universidad Nacional de Córdoba (RD-UNC)	100	98	95	97	45	39	89	99	11	0	100	3	11	100	100
Biblioteca Digital Universidad Católica Argentina	99	95	99	89	82	38	99	99	99	80	99	0	28	99	99
CIC Digital	100	100	100	100	29	20	100	0	100	0	100	0	0	100	100
Biblioteca Digital UNCuyo	95	91	99	92	97	28	90	98	100	71	100	7	9	100	100
Biblioteca Digital FCEN-UBA	100	81	52	77	93	51	99	100	100	33	93	3	7	56	100
CONICET Digital	100	100	100	100	99	0	100	100	100	2	100	99	0	100	100
Repositorio Institucional del Ministerio de Educac	100	61	98	40	28	0	100	99	56	0	64	75	36	0	100
Memoria Académica	100	99	98	63	63	8	100	100	90	92	93	49	0	90	100
SeDICI (Servicio de Difusión de la Creación Intele	100	100	100	95	3	12	100	100	0	2	100	86	0	85	100
Consejo Latinoamer, de Ciencias Sociales CLACSO	12	11	12	7	12	0	11	95	12	0	11	8	10	12	100

Referencias

■	0% a 19,99%
■	20% a 39,99%
■	40% a 59,99%
■	60% a 79,99%
■	80% a 100%

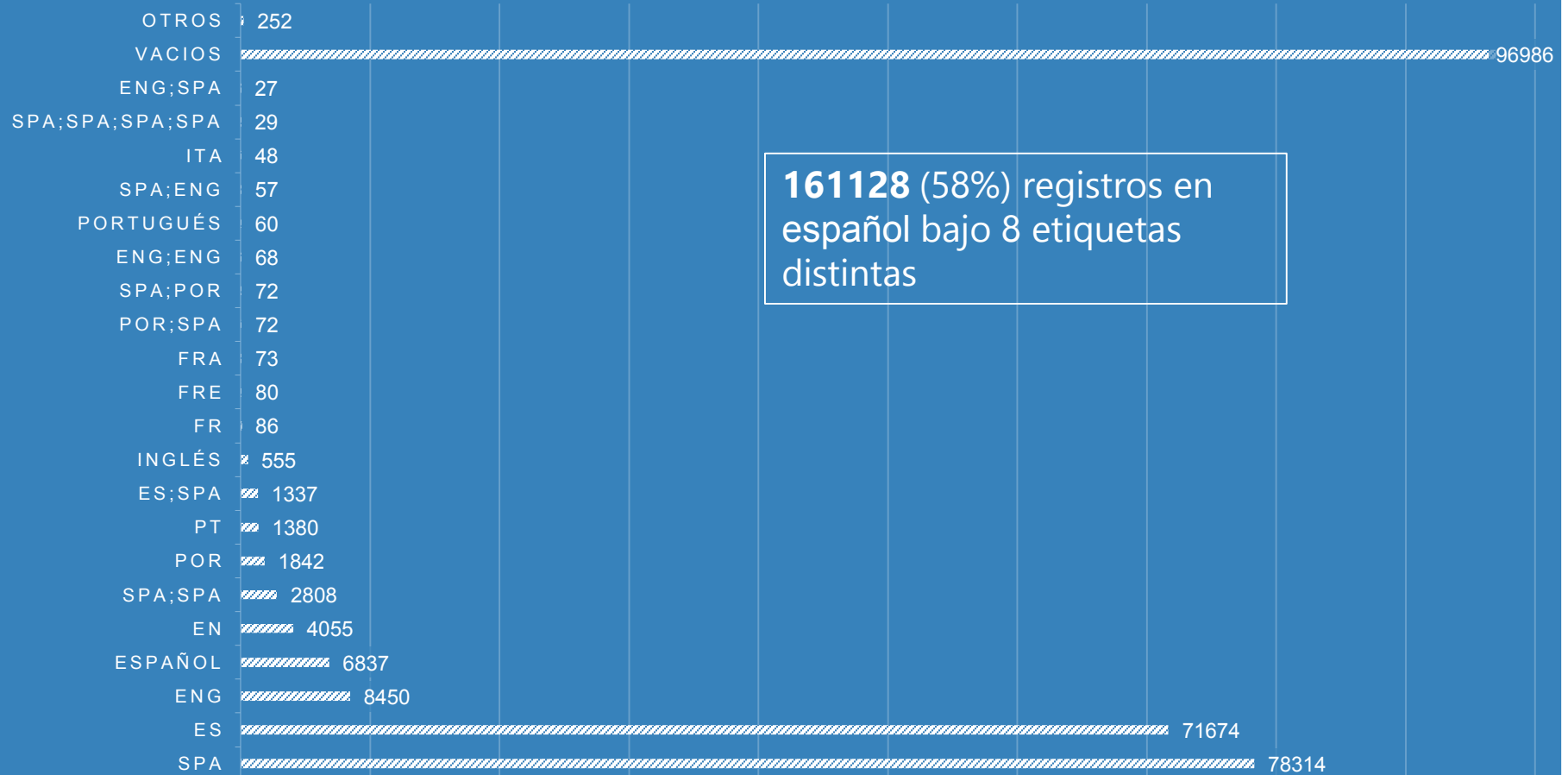
# UTILIZACIÓN DE CAMPOS DC



*¿Qué es la **calidad**?*  
*¿Qué medir cuando medimos?*  
*¿Cómo se mide la **calidad**?*

“

# LENGUAJE DEL REGISTRO



Campo type	cantidad	Campo type	cantidad
info:eu-repo/semantics/article;info:ar-repo/semantics/artículo;info:eu-repo/semantics/publishedVersion	27320	Imagen	4619
info:eu-repo/semantics/conferenceObject;info:ar-repo/semantics/documento de conferencia;info:eu-repo/semantics/publishedVersion	23948	info:eu-repo/semantics/review;info:ar-repo/semantics/visión literaria;info:eu-repo/semantics/publishedVersion	4108
info:eu-repo/semantics/article;info:eu-repo/semantics/publishedVersion	23245	article;info:ar-repo/semantics/artículo;info:eu-repo/semantics/article;info:eu-repo/semantics/publishedVersion	3724
Objeto de conferencia;Objeto de conferencia	19775	Tesis;Tesis de doctorado	3589
Artículo científico	19061	info:eu-repo/semantics/article;info:eu-repo/semantics/publishedVersion;Artículo revisado por pares	2993
Articulo;Articulo	18089	Reseña	2981
Article	12818	Text;draft;Capítulo de Libro	2951
legislation	12740	info:eu-repo/semantics/bachelorThesis;info:ar-repo/semantics/tesis de grado;info:eu-repo/semantics/acceptedVersion	2874
VACIOS	10119	info:eu-repo/semantics/article;artículo;info:eu-repo/semantics/publishedVersion	2424
text	7258	Artículo	2403

## Campo Type

1056 formas distintas, 443 formas distintas para referirse a un “Artículo”



## **Campo *format***

---

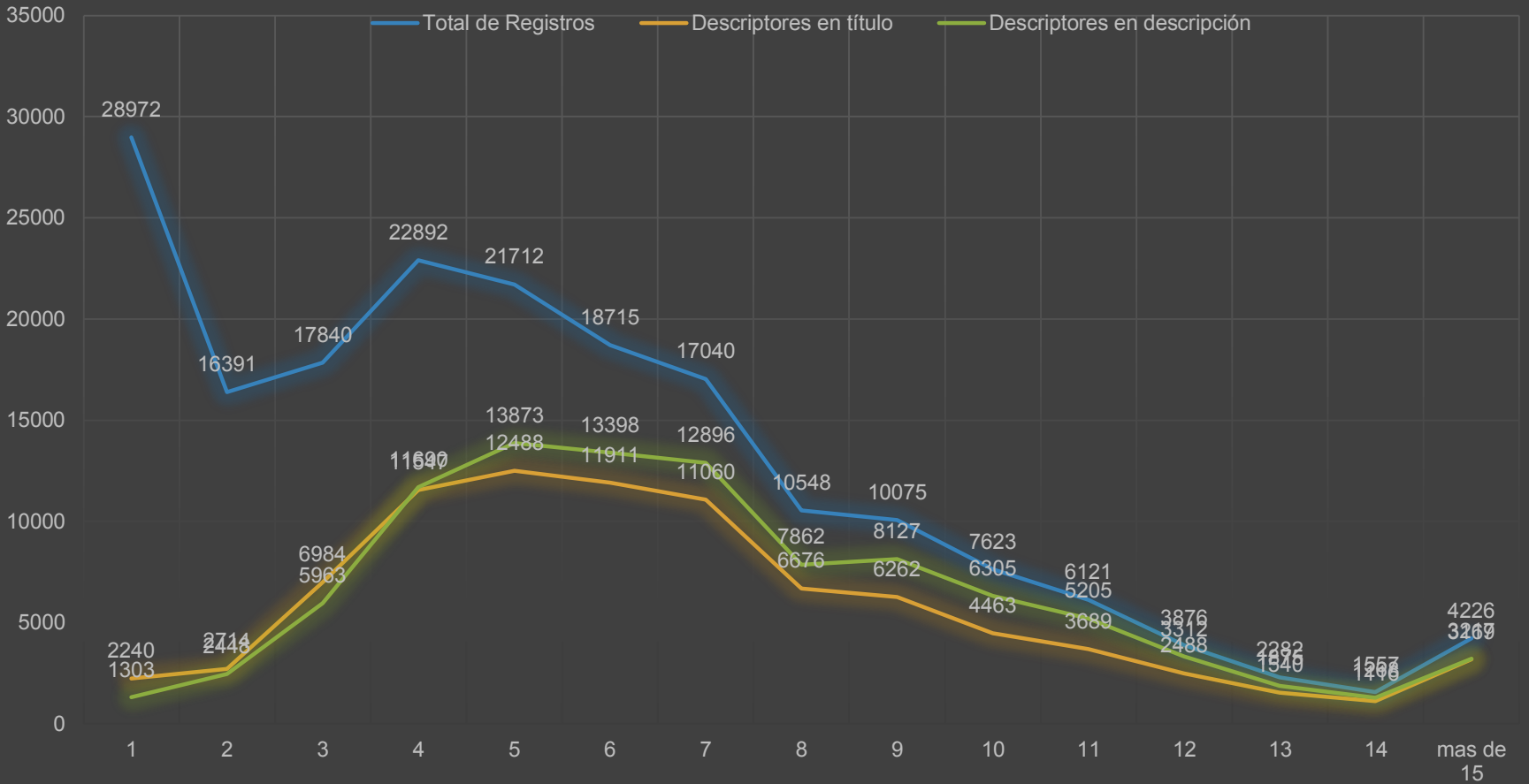
- Solo lo cumplimentan el 43% de los registros
- Existen 14142 formas distintas
- Existen 4281 formas distintas para referirse al formato PDF (35% del total de registros con este campo)
- El 4% para el formato HTML con solo 3 variantes



## Campo creator

- 66% de los registros contiene 135728 autores distintos
- el 92% se encuentra bajo la forma “Apellidos, Nombres”
- el resto en la forma “Nombres + Apellidos”

Cantidad de autores	Cantidad de registros
1	122020
2	22711
3	14046
4	8611
5	4855
6	3181
7	1825
8	1199
9	764
10	553
Mas de 10	1587



**Campo *Subject***  
 189870 registros lo cumplimentan, existen 814999 descriptores



# Conclusiones





## Conclusiones (1/3)

---

- Existe un problema serio de normalización de datos: campos *type*, *language*, *format*.
- Campo *creator* y formas del nombre
- Existencia de registros duplicados
- El uso desmedido de palabras claves y la incorporación de estas en título y resumen no cumplen su propósito



## Conclusiones (2/3)

---

- Es necesario establecer el origen del problema: falta de control, autoarchivo, límite de aplicación
- Aplicación de directrices DRIVER, OpenAIRE o SNRD, para facilitar la interoperabilidad
- Contar con esquemas formales y revisiones periódicas



## Conclusiones (3/3)

---

- ¿Cuál es la visibilidad/alcance real de los repositorios?
- Es necesario profundizar en la desambiguación de autores y de registros



# Gracias!

*Alguna pregunta ?*

Pueden encontrarme

- @jfedemedrano
- jfedericomedrano@gmail.com

# Calidad en repositorios digitales en Argentina, estudio comparativo y cualitativo



**UNJu**

Universidad  
Nacional de Jujuy

***Dr. J. Federico Medrano***

@fedemedrano