

Gbif.es

Índice de Calidad Aparente (ICA) en Darwin Test

Nodo Nacional de Información de Biodiversidad

Unidad de Coordinación

GBIF España

Año 2017



Índice de Calidad Aparente

El **Índice de Calidad Aparente (ICA)** es un indicador creado para dar una idea de la calidad de un juego de datos de presencia (citas corológicas, observaciones, ejemplares de colecciones de historia natural) en formato Darwin Core. Este índice es calculado por la aplicación **Darwin Test**. El ICA se desglosa en tres componentes: Taxonómico, Geográfico y Temporal. En el contexto de GBIF España usamos el ICA para hacer un seguimiento de la evolución de la calidad de los datos de biodiversidad publicados en nuestro ámbito. El llamar a este índice “de calidad aparente” deriva del hecho de que –aunque sea improbable– teóricamente un juego de datos puede tener un ICA máximo al tiempo que sus registros sean un pobre reflejo de la realidad a la que se refieren.

- Cuando Darwin Test calcula el **componente Taxonómico**, lo hace sobre la base de las coincidencias entre el nombre científico y la jerarquía taxonómica que hay entre el juego de datos en cuestión y la base de datos de “Catalogue of Life”. Es posible utilizar otras bases de datos de nombres científicos como, por ejemplo, los “Archivos de Autoridad Taxonómica” de SiB Colombia, u otros desarrollados para contextos específicos.
- El **componente Geográfico** valora el número de registros con coordenadas y el porcentaje de éstas que son consistentes con los contenidos de los campos “Country” y “stateProvince”. En la implementación del ICA en Darwin Test –tal como se descarga de la web de GBIF.ES– el chequeo contra “stateProvince” está optimizado para el territorio español. Su adaptación a otros territorios implica introducir en Darwin Test las coordenadas límite de las unidades “stateProvince” del territorio de interés.
- Por último, el **componente Temporal**, está determinado por el número de registros con valores posibles en los campos que atañen a fechas.

Cómo se calcula el ICA

45 % Componente Taxonómico			
20 %	% con géneros en Species2000	A	
10 %	% con especies	B	
9 %	% con jerarquía taxonómica		
	3 %	reino	C
	3 %	clase u orden	D
	3 %	familia	E
6 %	% con nombre de identificador	F	
35 % Componente Geográfico			
20 %	% con coordenadas	G	
10 %	% con países	H	
5 %	% con radiopunto	I	
- 20 %	% con coordenadas incorrectas	J	
20 % Componente Temporal			
11 %	% con año correcto	K	
7 %	% con mes correcto	L	
2 %	% con día correcto	M	
- 15 %	% con fechas incorrectas	N	

Desglose de la fórmula:

45% componente Taxonómico / Taxonomic component [*ICA_t*]

20 %	% con géneros en Species2000 / % with genus in Species2000		A
10 %	% con especies / % with species		B
9 %	% con jerarquía OK / % with rank OK		
	3%	reino / kingdom	C
	3%	clase u orden class or order	D
	3%	familia / family	E
6%	6 % with identified by		F

35% componente Geográfico / Geographic component [*ICA_g*]

20%	% con coordenadas OK / % with coordinates OK	G
10%	% con países OK / % with countries OK	H
5%	% con radiopunto OK / % with pointradius OK	I
- 20%	% coordenadas no OK / % coordinates not OK	J

20% componente Temporal / Temporal component [*ICA_d*]

11 %	% año OK / % year OK	K
7 %	% mes OK / % month OK	L
2 %	% día OK / % day OK	M
-15%	% fechas no OK / % dates not OK	N

$$ICA_t = (20 \times A) + (10 \times B) + (3 \times C) + (3 \times D) + (3 \times E) + (6 \times F)$$

$$ICA_g = (20 \times G) + (10 \times H) + (5 \times I) + (20 \times J)$$

$$ICA_d = (11 \times K) + (7 \times L) + (2 \times M) + (15 \times N)$$

$$ICA = ICA_t + ICA_g + ICA_d$$

Gbif, Es

Apparent Quality Index (ICA) in Darwin Test

Nodo Nacional de Información de Biodiversidad

Unidad de Coordinación

GBIF España

Año 2017

Apparent Quality Index

The **Apparent Quality Index** is an indicator of the data base quality (records, observations, specimens of natural history collections) in Darwin Core format. One of the features of the software application **Darwin Test** is to calculate this index. The index is divided in three components: Taxonomic, Geographic and Temporal. In the GBIF Spain framework, we use the ICA to monitor the quality of the biodiversity data published in the GBIF network. We refer to this index as an "apparent index" because, although it is unlikely, one data set can have a maximum ICA while its records are a poor reflection of the reality to which they refer.

- When Darwin Test calculates the **Taxonomic component**, it does so on the basis of the similarities between the scientific name and the taxonomic hierarchy, between the dataset and the data base in "Catalogue of Life". It is possible to use other data bases of scientific names such as "Archivos de Autoridad Taxonómica" from SiB Colombia or others developed for specific contexts.
- The **Geographic component** values the number of records with coordinates and the percentage of cells that are consistent with the fields "Country" and "stateProvince". In the implementation of ICA in Darwin Test the checking against "StateProvince" is optimized for the Spanish territory. Its adaptation to other areas means to include in Darwin Test the boundary coordinates of the units "StateProvince" the territory of interest.
- Finally, the **Temporal component** is determined by the number of right records in the date fields.

A continuación se enuncia y desglosa formalmente el cálculo del ICA / Below you can find how to calculate the ICA Index.

How the ICA is calculated

45 % Taxonomic component			
20 %	% with genus in Species2000	A	
10 %	% with species	B	
9 %	% with taxonomic hierarchy		
	3 %	kingdom	C
	3 %	Class or order	D
	3 %	family	E
6 %	% with identified by	F	
35 % Geographic Component			
20 %	% with coordinates	G	
10 %	% with countries	H	
5 %	% with pointradius	I	
- 20 %	% with incorrect coordinates	J	
20 % Temporal Component			
11%	% with correct year	K	
7 %	% with correct month	L	
2 %	% with correct day	M	
- 15 %	% with incorrect dates	N	

Formula Breakdown:

45% componente Taxonómico / Taxonomic component [*ICA_t*]

20 %	% con géneros en Species2000 / % with genus in Species2000		A
10 %	% con especies / % with species		B
9 %	% con jerarquía OK / % with rank OK		
	3%	reino / kingdom	C
	3%	clase u orden class or order	D
	3%	familia / family	E
6%	6 % with identified by		F

35% componente Geográfico / Geographic component [*ICA_g*]

20%	% con coordenadas OK / % with coordinates OK	G
10%	% con países OK / % with countries OK	H
5%	% con radiopunto OK / % with pointradius OK	I
- 20%	% coordenadas no OK / % coordinates not OK	J

20% componente Temporal / Temporal component [*ICA_d*]

11 %	% año OK / % year OK	K
7 %	% mes OK / % month OK	L
2 %	% día OK / % day OK	M
-15%	% fechas no OK / % dates not OK	N

$$ICA_t = (20 \times A) + (10 \times B) + (3 \times C) + (3 \times D) + (3 \times E) + (6 \times F)$$

$$ICA_g = (20 \times G) + (10 \times H) + (5 \times I) + (20 \times J)$$

$$ICA_d = (11 \times K) + (7 \times L) + (2 \times M) + (15 \times N)$$

$$ICA = ICA_t + ICA_g + ICA_d$$

Gbif, Es

Nodo Nacional de Información
de Biodiversidad

Unidad de Coordinación

GBIF España

gbif.es

datos.gbif.es

Año 2017

Real Jardín Botánico - CSIC

Plaza de Murillo, 2

28014 Madrid - SPAIN

Tel. 91 420 30 17 ext.170-171

