



Análise de Producción Científica

BUGALICIA - Consorcio de Bibliotecas Universitarias de Galicia

Maio de 2008



CONTROL DE VERSIONS				
Versión	Revisión	Data	Autor	Descrición
1.0	1.0	01/08	José Triñanes Dieste	Documento base
2.0	1.0	03/08	José Triñanes Dieste	Modificaci3ns tras revisi3n
3.0	1.0	05/08	José Triñanes Dieste	Modificaci3ns tras revisi3n

Cláusula de confidencialidade:

Este documento é propiedade do Consorcio de Bibliotecas Universitarias de Galicia. Tódolos dereitos reservados. Non está permitida a reprodución total ou parcial destes documentos, nin o seu tratamento informático, nin a transmisión de ningunha forma ou por calquera medio, sen o permiso previo e por escrito do Consorcio de Bibliotecas Universitarias de Galicia. Queda expresa e terminantemente prohibida a súa comercialización.



Índice

1.- Introducción	6
2.- Fontes de datos	8
2.1.- ISI Web of Science.....	8
2.2.- Scopus	8
2.3.- Base de datos do IEDCYT	9
2.4.- Journal Citation Reports.....	9
2.5.- Grupos de Investigación	9
2.6.- Outras fontes	9
3.- ISI Web of Science	10
3.1.- Introducción	10
3.2.- Acceso á información	10
3.3.- Información dispoñible	11
3.4.- Estructura da información.....	12
3.5.- Consideracións a ter en conta	14
4.- Scopus	17
4.1.- Introducción	17
4.2.- Acceso á información	17
4.3.- Información dispoñible	17
4.4.- Estructura da información.....	18
4.5.- Consideracións a ter en conta	23
5.- IEDCYT	26
5.1.- Introducción	26
5.2.- Acceso á información	26
5.3.- Información dispoñible	26
5.4.- Estructura da información.....	29
5.5.- Consideracións a ter en conta	36
6.- Journal Citation Reports	38
6.1.- Introducción	38



6.2.- Acceso á información	38
6.3.- Información dispoñible	38
6.4.- Estructura da información.....	39
6.5.- Consideracións a ter en conta	41
7.- Grupos de Investigación	42
7.1.- Introducción	42
7.2.- Acceso á información	42
7.3.- Información dispoñible	42
7.4.- Estructura da información.....	42
8.- Outras fontes de datos.....	44
8.1.- Introducción	44
8.2.- Institucións Científicas.....	45
8.3.- Categorias ANEP	47
8.4.- Categorias UNESCO	48
8.5.- Categorias ISOC (IEDCYT)	49
8.6.- Publicacións de Scopus	50
8.7.- Publicacións de IEDCYT.....	52
9.- Modelo de Datos	53
9.1.- Introducción	53
9.2.- Dimensións	54
9.3.- Indicadores.....	61
10.- Informes e análise de datos.....	67
10.1.- Introducción	67
10.2.- Informes de cantidade de produción científica	67
10.3.- Informes de calidade da produción científica	68
10.4.- Informes de relacións estruturais	68
10.5.- Informes de colaboración	71
10.6.- Relación de informes.....	71
11.- Proceso de carga	74
11.1.- Características dos procesos de carga.....	74
12.- Outros procesos e ferramentas.....	76



12.1.-	Introducción	76
12.2.-	Aplicación de modificación e actualización de datos	76
12.3.-	Proceso de descarga de datos de ISI Web of Science.....	77
12.4.-	Proceso de descarga de datos de Scopus	78
12.5.-	Proceso de obtención do DOI (Digital Object Identifier)	79
13.-	Plataforma tecnolóxica	82
14.-	Conclusións.....	83
14.1.-	Consideracións sobre o Sistema de Análise.....	83
14.2.-	Plantexamento do Sistema de Análise.....	85
14.3.-	Configuración.....	85
14.4.-	Desenvolvemento.....	86
ANEXO.	Sitios de interés na Internet:	87

1.- Introducción

O Consorcio de Bibliotecas Universitarias de Galicia (BUGALICIA), como parte da súa labor de apoio continuo aos servizos bibliotecarios das tres universidades galegas, precisa dispor de información sobre a produción científica en Galicia.

O plantexamento inicial é integrar a información de produción científica procedente de diferentes fontes nunha base de datos sobre a que realizar os informes e consultas que se necesiten.

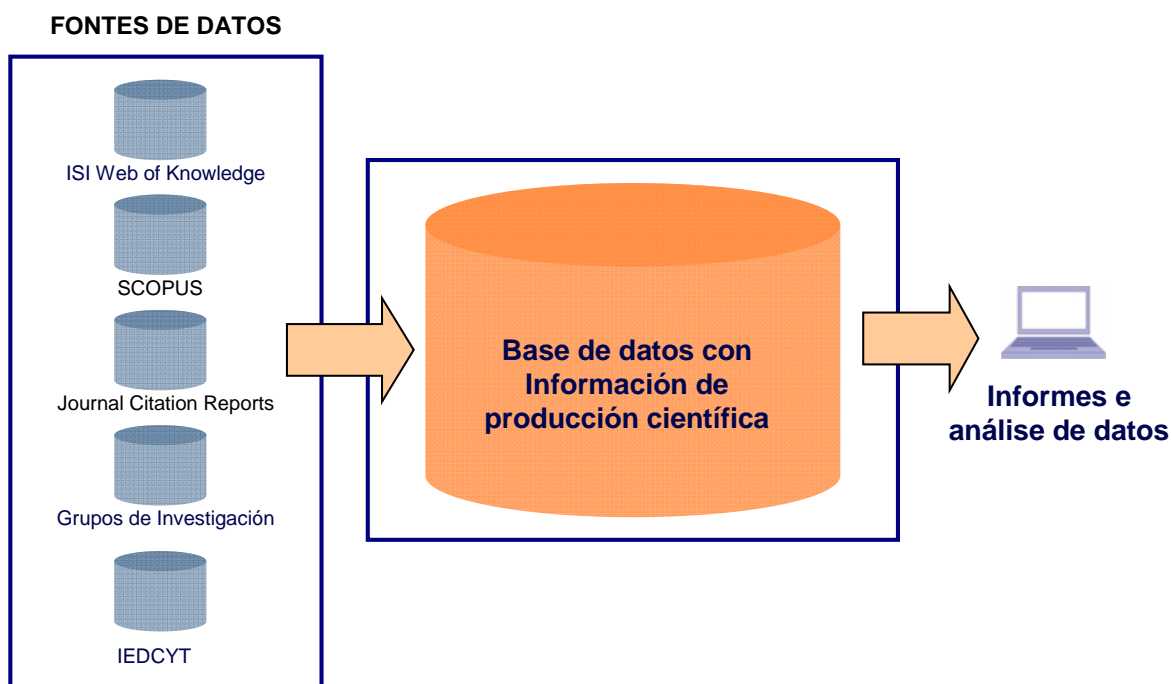


Figura 1: Plantexamento inicial do Sistema de Análise

Entre outros, os obxectivos que se perseguen coa implantación dun repositorio centralizado de datos son:

- Integrar nun único modelo de datos a información de produción científica provinte das diferentes fontes.
- Proporcionar maiores capacidades para a explotación e a análise de información.



- Facilitar a obtención de informes de producción científica, algúns dos cales estarán publicados nun portal web e poderán ser consultados polos usuarios que os precisen.

O presente documento é froito do traballo realizado polo Consorcio de Bibliotecas e Tecnocom para identifica-las características principais e a funcionalidade básica requerida para o deseño e construción dun **Sistema de Análise de Producción Científica**.

O documento está estruturado do seguinte xeito:

- Nos primeiros capítulos dase unha descrición básica das fontes de datos, coas características da información e as súas particularidades.
- Establécese unha primeira aproximación ao modelo de datos do Sistema de Análise, enumerando os indicadores necesarios para realizar informes e os elementos polos que realizar agrupacións de datos.
- Identifícanse un conxunto básico de informes e consultas a realizar sobre o Sistema.
- Defínense as características dos procesos de carga de datos sobre o Sistema.
- Identifícanse procesos e tarefas necesarios para o Sistema e unha breve reseña sobre a tecnoloxía a empregar.
- Finalmente recóllese o plantexamento que se considera máis axeitado e as conclusións ás que se chegou despois do traballo realizado.



2.- Fontes de datos

Para a análise de produción científica en Galicia, o Consorcio de Bibliotecas Universitarias dispón de diferentes fontes de datos:

- ISI Web of Science: Documentos publicados.
- SCOPUS: Documentos publicados.
- CINDOC (CSIC): Documentos publicados:
 - o ICYT - Ciencia e tecnoloxía
 - o ISOC - Ciencias sociais e humanas
 - o IME - Biomedicina
- JCR - ISI Journal Citation Reports. Información sobre publicacións.
- Grupos de Investigación SUG: Información sobre os Grupos de Investigación do Sistema Universitario Galego.
- Outras fontes: Relación de Institucións científicas, categorías ANEP, categorías UNESCO, etc.

Un mesmo documento pode estar referenciado en cada unha das fontes (ISI, Scopus e IEDCYT), polo que se poderían chegar a ter tres visións diferentes para un único documento publicado.

2.1.- ISI Web of Science

ISI (Institute for Scientific Information) Web of Science forma parte do ISI Web of Knowledge, unha plataforma integrada de Thomson Scientific, con acceso vía web e deseñada para axudar en tódolos niveis da investigación científica e académica que, entre outros contidos, ofrece datos de publicacións científicas.

2.2.- Scopus

Scopus é unha base de datos que engloba unha colección a nivel mundial de resumos, referencias e índices de literatura científica, técnica e médica.



2.3.- Base de datos do IEDCYT

O Instituto de Estudos Documentales sobre Ciencia y Tecnología (IEDCYT) do CSIC (Consejo Superior de Investigaciones Científicas) proporciona ao Consorcio de Bibliotecas unha serie de arquivos de texto que conteñen información referencial e bibliográfica que recolle a literatura científica contida en publicacións españolas.

Esta información agrúpase en:

- Ciencia e Tecnoloxía (ICYT)
- Ciencias sociais e humanas (ISOC)
- Biomedicina (IME)

2.4.- Journal Citation Reports

ISI Journal Citation Reports é parte do ISI Web of Knowledge. Presenta datos estatísticos cuantificables que permiten avaliar e determina-la importancia relativa das revistas por áreas de coñecemento.

Ademais permite coñece-las revistas científicas de maior impacto baseándose na análise das citacións dos artigos publicados nas mesmas revistas.

2.5.- Grupos de Investigación

Esta base de datos contén información sobre os grupos de investigación do Sistema Universitario Galego (SUG) e almacena datos de investigadores e artigos e revistas nas que publicaron algún texto.

2.6.- Outras fontes

O Consorcio de Bibliotecas precisa ademáis dispor de información actualizada da relación de Categorias ANEP, Categorias UNESCO, Institucións Científicas ou outras posibles listas normalizadas relacionadas coa produción científica que faciliten a explotación de datos.



3.- ISI Web of Science

A continuación detallarase a información dispoñible, estrutura e características dos datos procedentes do ISI Web of Science.

3.1.- Introducción

Dispoñible a través da plataforma ISI Web of Knowledge, Web of Science contén información sobre investigación multidisciplinaria de alta calidade publicada en revistas líderes mundiais nas ciencias, ciencias sociais, artes e humanidades.

ISI Web of Science identifica e indexa tódolos documentos significativos contidos en ditas publicacións. Isto quere dicir que ademáis de artigos poden buscarse cartas, correccións, adicionais, editoriais ou revisións que teñan aparecido na revista.

Web of Science é unha base de datos bibliográfica, polo tanto os rexistros conteñen información tal como títulos, autores, palabras clave, resumos, referencias citadas e outros detalles.

3.2.- Acceso á información

A información facilitada polo ISI Web of Knowledge é accesible vía web a través da Internet; deste xeito pódense obter os datos bibliográficos e de citas.

A descarga de datos a través da ferramenta en Internet non é moi operativa. ISI tamén proporciona acceso á súa información a través de Web Services e unha API (*Application Programmer Interface*). Este servizo permite establecer criterios de busca para filtra-los datos que se queren descargar.

Sería preciso programar un proceso que, facendo uso da API facilitada por ISI, obtivese os arquivos de texto coa información bibliográfica requerida.

No apartado 12.3.- *Proceso de descarga de datos de ISI Web of Science* dase máis información sobre este punto.



3.3.- Información dispoñible

A información dispoñible en ISI Web of Science inclúe máis de 16.000 publicacións, libros, actas e anais internacionais. Parte destes datos proveñen das 8.700 publicacións de investigación en diferentes disciplinas que ISI cubre anualmente (en Febreiro de 2008).

Web of Science está composta por tres bases de datos, que poden consultarse individual ou conxuntamente:

	Información dispoñible de publicacións	Revistas incluídas	Novos rexistros semanais	Novas referencias citadas semanais
Science Citation Index	1900 - actualidade	6.034	18.800	431.000
Social Sciences Citation Index	1956 - actualidade	1.764	2.600	59.100
Arts and Humanities Citation Index	1975 - actualidade	1.121	2.100	15.900

ISI Web of Science proporciona información de **Documentos** publicados. Para cada documento especificase o Tipo de Documento, Título, Idioma, Ano de publicación, Número de citas, Número de referencias citadas, etc...

Ademais cada documento ten asociada a seguinte información que é preciso ter en conta para incorporar ao Sistema de Análise:

- **Autores:** Autor/es do documento.
- **Referencias citadas:** Referencias que se fan nun documento a outros documentos.
- **Institucións:** Institución ou conxunto de Institucións que colaboran na elaboración do documento.
- **Palabras clave:** Conxunto de palabras (ou frases) clave que indicarían o contido dun documento.

ISI Web of Science contén dous grupos de palabras clave: definidas polo autor e definidas por ISI.



3.4.- Estructura da información

Neste apartado detállase a estrutura da información bibliográfica e de citas pertencente ao ISI Web of Science.

O formato aquí definido correspóndese cos arquivos de texto que se poden descargar a través da páxina web do ISI Web of Science e pretende servir como referencia da información contida na base de datos de ISI.

No caso de obte-los datos a través dun proceso que faga uso dos *Web Services* e da API (*Application Programming Interface*) de Thomson Scientific o formato podería variar.

Independentemente de como se obteña a información, xa sexa a través da páxina web ou empregando os *Web Services*, os datos dispoñibles e a súa estrutura sería a que aquí se indica.

Cada etiqueta do arquivo identifica un elemento de datos.

Etiqueta	Descrición
FN	Tipo de ficheiro.
VR	Número de versión do formato de ficheiro.
PT	Tipo de publicación (p.e.: libro, diario, serie de libro).
AU	Autor/es. Cada autor aparece nunha liña.
TI	Título do artigo.
SO	Título completo da fonte.
LA	Idioma.
DT	Tipo de documento (p.e.: Book Review, Editorial Material, Fiction, Note, Review)
NR	Número de referencias citadas.
SN	ISSN
PU	Editor.
C1	Dirección da institución ou institucións investigadoras que participaron na elaboración do documento. De existir máis de unha aparecerán en liñas distintas. Se varios departamentos dunha institución participan no documento, aparecerán en liñas distintas, como institucións diferentes.
DE	Palabras clave do autor.
ID	Palabras clave de ISI.
AB	Resumo.



Etiqueta	Descripción
CR	Referencias citadas. Cada referencia citada aparece nunha liña, co seguinte formato: Autor, Ano, Título abreviado da publicación, Volumen, Páxina. Por exemplo: BATTEN M, 2006, NAT IMMUNOL, V7, P929 No caso dalgúns xornais electrónicos emprégase o DOI (Digital Object Identifier) no canto dos números de páxina e números de artigo. Se non hai información de número de páxina ou do número de artigo citado aparecerá o DOI.
TC	Número de citas; número de veces que o documento foi citado.
BP	Páxina inicial.
EP	Páxina final.
PG	Número de páxinas.
JI	Abreviatura en formato ISO do título da fonte.
SE	Título da serie de libros.
BS	Subtítulo da serie de libros.
PY	Ano de publicación
PD	Data de publicación.
VL	Volumen.
IS	Publicación.
PN	Número de parte.
SU	Suplemento.
SI	Publicación especial.
GA	Número ISI de entrega de documento.
PI	Cidade do editor.
WP	Dirección web do editor.
RP	Dirección do autor.
CP	Patente citada.
J9	Abreviatura do título da fonte (ata 29 caracteres).
PA	Dirección do editor.
UT	Identificador único de ISI para cada artigo.
ER	Fin do rexistro.

O seguinte é un exemplo de información obtida do ISI Web of Science:

FN ISI Export Format
VR 1.0
PT S
AU Daniels, GM
Amara, SG
AF Daniels, Gwynn M.
Amara, Susan G.



TI Selective labeling of neurotransmitter transporters at the cell surface
 SO NEUROTRANSMITTER TRANSPORTERS
 SE METHODS IN ENZYMOLOGY
 LA English
 DT Review
 ID POLARIZED EPITHELIAL-CELLS; EXPRESSION
 C1 Oregon Hlth & Sci Univ, Vollum Inst Adv Biomed Res, Dept Cell & Dev Biol, Portland, OR 97201 USA.
 Oregon Hlth & Sci Univ, Howard Hughes Med Inst, Portland, OR 97201 USA.
 RP Daniels, GM, Oregon Hlth & Sci Univ, Vollum Inst Adv Biomed Res, Dept
 Cell & Dev Biol, Portland, OR 97201 USA.
 CR AHN J, 1996, J BIOL CHEM, V271, P6917
 GOTTARDI CJ, 1993, SCIENCE, V260, P552
 GOTTARDI CJ, 1995, AM J PHYSIOL, V268, P285
 GU HH, 1996, J BIOL CHEM, V271, P18100
 MATTER K, 1994, CURR OPIN CELL BIOL, V6, P545
 NGUYEN TT, 1996, J NEUROCHEM, V67, P645
 PIETRINI G, 1994, J BIOL CHEM, V269, P4668
 POVLOCK SL, 1998, METHODS ENZYMOL, V296
 SARGIACOMO M, 1989, J MEMBRANE BIOL, V107, P277
 NR 9
 TC 29
 PU ELSEVIER ACADEMIC PRESS INC
 PI SAN DIEGO
 PA 525 B STREET, SUITE 1900, SAN DIEGO, CA 92101-4495 USA
 SN 0076-6879
 J9 METH ENZYMOLOGY
 JI Methods Enzymol.
 PY 1998
 VL 296
 BP 307
 EP 318
 PG 12
 SC Biochemical Research Methods; Biochemistry & Molecular Biology
 GA BGJ33
 UT ISI:000247400300021
 ER

3.5.- Consideracións a ter en conta

A continuación recóllense certas consideracións sobre os datos do ISI Web of Science:



- **Institucións:** As institucións que colaboran no documento veñen tra-la etiqueta "C1", na que se indica a dirección da institución investigadora.

Cada institución ou departamento dunha mesma institución aparece nunha liña separada, pero a dirección está nunha mesma liña separada por comas (,) o que dificulta a identificación do texto, xa que ademáis empréganse abreviaturas.

Por exemplo:

C1 Univ Magdeburg, Fak Med, Zentrum Innere Med, Klin Gastroenterol Hepatol & Infektiol, D-39120 Magdeburg, Germany.
 INSERM, U853, Bordeaux, France.
 Univ Dublin Trinity Coll, Adelaide & Meath Hosp, Dublin, Ireland.
 Univ Bologna, Bologna, Italy.
 Univ Aberdeen, Aberdeen AB9 1FX, Scotland.
 VA Med Ctr, Houston, TX USA.
 McMaster Univ, Hamilton, ON, Canada.
 Henry Dunant Hosp, Athens, Greece.
 Univ Wisconsin, Sch Med, Milwaukee, WI 53201 USA.
 Erasmus Univ, Med Ctr, Rotterdam, Netherlands.

- **Referencias citadas:** A información de referencias citadas é imprescindible para realiza-las análises de cocitacións. Para elo é preciso identificar correctamente os documentos aos que fai referencia un documento concreto.

Cada referencia citada aparece na forma: Autor, Ano, Título abreviado da publicación, Volumen e Páxina.

Esta información aparece tra-la etiqueta "CR".

Por exemplo:

CR ABEYSINGHE SS, 2004, HUM MUTAT, V23, P219
 ALLIS CD, 1980, CELL, V20, P609
 AN W, 2004, CELL, V117, P735
 BANNISTER AJ, 2001, NATURE, V410, P120
 BANNISTER AJ, 2005, J BIOL CHEM, V280, P17732
 BERNSTEIN BE, 2002, P NATL ACAD SCI USA, V99, P8695
 BERNSTEIN BE, 2005, CELL, V120, P169
 BERNSTEIN BE, 2006, CELL, V125, P315
 BERNSTEIN BE, 2007, CELL, V128, P669
 BOYER LA, 2006, NATURE, V441, P349
 BRUCE K, 2005, NUCLEIC ACIDS RES, V33, P5633

Nalgúns casos, e se non existe información de páxina ou volumen amósase o DOI (Digital Object Identifier).



- **Materias:** A materia (ou materias) sobre a que trata un documento publicado non está directamente referida no arquivo que se obtén de ISI, senón que dita materia derivase a partires da revista donde o documento en cuestión foi publicado.

- **Palabras clave:** É preciso rexistra-lo conxunto de palabras clave asociado a cada documento.
Agás que o deseño do modelo de datos así o aconsellara, non sería preciso rexistrar por separado cada unha das palabras clave que aparecen asociadas a un documento.

- **Citas e Referencias citadas:** O número de citas que ten o documento correspondese coas veces que foi citado por outros documentos. Aparece tra-la etiqueta "TC" (Times Cited).
O número de Referencias citadas está despois da etiqueta "NR" e representa o número de referencias que un documento fai a outros documentos. Correspóndese co número de liñas que hai tra-la etiqueta CR de referencias citadas.



4.- Scopus

A continuación detallarase a información dispoñible, estrutura e características dos datos procedentes de Scopus.

4.1.- Introducción

Scopus dispón dunha colección a nivel mundial de resumos, referencias e índices de literatura científica, técnica e médica e tamén de información publicada na Internet.

4.2.- Acceso á información

Scopus pon a disposición dos seus usuarios unha ferramenta web para acceder á súa información a través da Internet, aínda que a descarga de datos deste xeito non é operativa para o o volumen de datos que precisa o Consorcio de Bibliotecas.

Dispón tamén dunha API (*Application Programmer Interface*) que permite obter os resultados dunha búsqueda en Scopus. Para elo é preciso programar un proceso que faga uso da API e obteña a información requirida.

No apartado 12.4.- *Proceso de descarga de datos de Scopus* dase máis información sobre este punto.

4.3.- Información dispoñible

A información dispoñible en Scopus actualízase diariamente e inclúe resumos e referencias citadas (dende 1996) de máis de 15.000 publicacións procedentes de aproximadamente 4.000 editores internacionais que cubren as áreas de ciencias sociais e humanas, saúde e ciencias físicas.

De Scopus obtense información de **Documentos** publicados. Para cada documento especificase o Tipo de Documento, Título, Idioma, Ano de publicación, Número de citas, etc...



Ademais cada documento ten asociada a seguinte información que é preciso ter en conta para incorporar ao Sistema de Análise:

- **Autores:** Autor/es do documento.
- **Referencias citadas:** Referencias que se fan nun documento a outros documentos.
- **Institucións:** Institución ou conxunto de Institucións que colaboran na elaboración do documento.
- **Palabras clave:** Conxunto de palabras (ou frases) clave que indicarían o contido dun documento.

4.4.- Estructura da información

A continuación detállase o formato dos datos procedentes de Scopus e que conteñen información bibliográfica.

O formato aquí definido correspóndese cos arquivos de texto que se poden descargar a través da páxina web de Scopus. No caso de obte-los datos a través dun proceso que faga uso da API (*Application Programming Interface*) de Scopus o formato podería variar.

A estrutura aquí definida pretende identifica-lo conxunto de información que se pode recoller de Scopus.

Cada etiqueta dentro do arquivo identifica un elemento de datos.

Etiqueta	Descrición
Etiquetas de título e referencia	
TY	Tipo de referencia (Tipo de Documento) Debe conter un dos valores definidos na táboa Tipo de Referencia (ve táboa Tipos de Referencia)
ER	Fin de referencia. Debe ser a última etiqueta nunha referencia.
ID	Identificador de referencia. Pode consistir en calquera carácter alfanumérico ata unha lonxitude máxima de 20.
T1	Título primario.
TI	A etiqueta BT emprégase únicamente para libros completos e
CT	referencias a traballos non publicados.
BT	Este campo pode conter caracteres alfanuméricos; practicamente non existe limitación na lonxitude deste campo.



Etiqueta	Descrición
T2 BT	Título secundario. A etiqueta BT emprégase para tódolos tipos de referencias agás para libros completos e referencias a traballos non publicados; practicamente non existe limitación na lonxitude deste campo.
T3	Título da serie. Este campo pode conter caracteres alfanúmericos; practicamente non existe limitación na lonxitude deste campo.
Etiquetas de autor	
A1 AU	Autor principal. Cada autor debe estar nunha liña separada, precedido desta etiqueta. Cada referencia pode conter un número ilimitado de autores, cunha lonxitude de ata 255 caracteres para cada valor. O nome do autor debe respecta-la seguinte sintaxis: Apellido, Nome, Sufixo. Como nome poden aparecer nomes completos, iniciais ou ámbolos dous. O sufixo é opcional; p.e.: Jr/Sr/II/MD, etc.
A2 ED	Autor secundario. Cada autor debe estar nunha liña separada, precedido desta etiqueta. Terá a mesma sintaxis que o autor principal (A1, AU) e pode ter unha lonxitude de ata 255 caracteres.
A3	Autor da serie. Cada autor debe estar nunha liña separada, precedido desta etiqueta. Terá a mesma sintaxis que o autor principal (A1, AU) e pode ter unha lonxitude de ata 255 caracteres.
Etiquetas de ano e texto libre	
Y1 PY	Data principal. Debe estar no seguinte formato: YYYY/MM/DD/outra información. O ano, mes e día son datos numéricos. O campo "outra información" pode conter letras, espazos e guións. Cada unha da información específica da data é opcional, pero non as barras (/) non son opcionais. p.e.: se aparece únicamente ano e outra información: 1998///Spring.
Y2	Data secundaria. Co formato indicado para as etiquetas Y1 e PY.
N1 N2	Notas. Conteñen texto libre con caracteres alfanuméricos; practicamente non hai limitación para a lonxitude deste campo. Detrás desta etiqueta atópase información importante para o sistema. As referencias citadas a outro documento aparecerán coa etiqueta "N1" seguida da palabra "References". Por exemplo: N1 - References: Hall, S.R., Allen, F.H., Brown, I.D., (1991) Acta Cryst., A47, p. 655 As citas feitas a un documento teñen, despois do valor "N1", a palabra "Cited By". Por exemplo: Cited By (since 1996): 39425



Etiqueta	Descripción
AB	Resumo. Texto libre con caracteres alfanuméricos; practicamente non ten limitación na súa lonxitude.
Etiquetas de palabras clave e estado de reimpresión	
KW	Palabras clave. Cada palabra clave ou frase estará nunha liña nova, precedida da correspondente etiqueta. Poden ser palabras ou frases completas, cunha lonxitude de ata 255 caracteres. Prácticamente non hai limitación ao número de palabras clave.
RP	Estado de reimpresión. Campo opcional. Pode toma-los valores "IN FILE", "NOT IN FILE", "ON REQUEST (MM/DD/YY) [Non se emprega]
Etiquetas de publicación	
JF	Nome completo de publicación periódica.
JO	Valor alfanumérico de ata 255 caracteres.
JA	Abreviatura estándar de nome de publicación periódica. Indica a publicación na que o artigo foi (ou vai ser, no caso de referencias en prensa) publicado. Valor alfanumérico de ata 255 caracteres. Si é posible, as abreviaturas seguirán o estilo do <i>Index Medicus</i> , con periodos despois das abreviaturas.
J1	Nome de publicación periódica: abreviatura de usuario 1. Lonxitude de ata 255 caracteres.
J2	Nome de publicación periódica: abreviatura de usuario 2. Lonxitude de ata 255 caracteres.
VL	Número de volumen.
IS	Número de publicación.
SP	Páxina inicial.
EP	Páxina final.
CP	Cidade de publicación.
PB	Editor
SN	ISSN/ISBN
AD	Dirección completa da institución asociada ao documento referenciado. Se se trata dunha colaboración e polo tanto hai máis dunha institución, cada unha delas ven codificada nunha nova liña tra-la etiqueta "AD".
Etiquetas varias	
AV	Disponibilidade.
M1 M2 M3	Miscelánea (de 1 a 3 valores).
U1 U2 U3 U4 U5	Valores definidos polo usuario (de 1 a 5).



Etiqueta	Descrición
UR	Web/URL. Practicamente non hai límite á lonxitude deste campo. As direccións URL poden atoparse de xeito individual, unha por etiqueta ou múltiples direccións separadas por punto e coma. There is no practical limit to the length of this field. URL addresses can be entered individually, one per tag or multiple addresses can be entered on one line using a semi-colon as a separator.
L1	Enlace a PDF.
L2	Enlace a texto completo.
L3	Rexistros relacionados.
L4	Imaxes

Táboa de Tipos de Referencia

Nome	Tipo de referencia
ABST	Abstract
ADVS	Audiovisual material
ART	Art Work
BILL	Bill/Resolution
BOOK	Book, Whole
CASE	Case
CHAP	Book chapter
COMP	Computer program
CONF	Conference proceeding
CTLG	Catalog
DATA	Data file
ELEC	Electronic Citation
GEN	Generic
HEAR	Hearing
ICOMM	Internet Communication
INPR	In Press
JFULL	Journal (full)
JOUR	Journal
MAP	Map
MGZN	Magazine article
MPCT	Motion picture
MUSIC	Music score
NEWS	Newspaper
PAMP	Pamphlet
PAT	Patent
PCOMM	Personal communication
RPRT	Report
SER	Serial (Book, Monograph)
SLIDE	Slide



Nome	Tipo de referencia
SOUND	Sound recording
STAT	Statute
THES	Thesis/Dissertation
UNBILI	Unenacted bill/resolution
UNPB	Unpublished work
VIDEO	Video recording

O seguinte é un exemplo de información obtida de Scopus:

TY - JOUR
T1 - Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction
JF - Analytical Biochemistry
JA - ANAL. BIOCHEM.
VL - 162
IS - 1
SP - 156
EP - 159
PY - 1987///
SN - 00032697 (ISSN)
AU - Chomczynski, P.
AU - Sacchi, N.
AD - Laboratory of Biochemistry and Metabolism, NIDDK, National Institutes of Health, Bethesda, MD 20892, United States
AB - A new method of total RNA isolation by a single extraction with an acid guanidinium thiocyanate-phenol-chloroform mixture is described. The method provides a pure preparation of undegraded RNA in high yield and can be completed within 4 h. It is particularly useful for processing large numbers of samples and for isolation of RNA from minute quantities of cells or tissue samples.
KW - messenger rna
KW - rna
KW - gene expression
KW - genetic engineering
KW - human
KW - human cell
KW - priority journal
KW - Animal
KW - Cell Line
KW - Chloroform
KW - Guanidine
KW - Guanidines
KW - Human
KW - Hydrogen-Ion Concentration



KW - Mammae
 KW - Phenol
 KW - Phenols
 KW - Rats
 KW - RNA
 KW - Solutions
 N1 - Cited By (since 1996): 39425
 N1 - Export Date: 19 December 2007
 N1 - Source: Scopus
 N1 - CODEN: ANBCA
 N1 - Language of Original Document: English
 N1 - Chemicals/CAS: RNA, 63231-63-0; Chloroform, 67-66-3; Guanidine, 113-00-8; Guanidines; Phenol, 108-95-2; Phenols; RNA, 63231-63-0; Solutions
 UR - <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0023277545&partnerID=40&rel=R7.0.0>
 ER -

4.5.- Consideracións a ter en conta

Neste apartado recóllense certas consideracións sobre os datos de Scopus:

- **Obtención de datos:** Como se ten dito nun punto anterior, o ideal para obter a información de Scopus é facer uso da API que proporciona, xa que deste xeito podería automatizarse a tarefa de descarga de datos.
Segundo a documentación da API de Scopus, non se poden recuperar as referencias citadas nin as palabras clave, o que limitaría as explotacións de datos: As referencias citadas son imprescindibles para facer análise de citas.
- **Autores:** Scopus dispón de etiquetas diferentes para identificar a autoría dun documento, segundo se trate de autores principais, secundarios ou autor da serie.
- **Institucións:** Nos ficheiros de texto, cada institución que colabora na elaboración dun documento ven tra-la etiqueta "AD". En cada unha das liñas aparece dirección completa da institución.

Por exemplo:

AD - Baylor Inst. for Immunol., Research, Baylor Research Institute, 3409 Worth Street, Dallas, TX 75246, United States

AD - Lab. of Cell. Physiol. and Immunol., Rockefeller University, 1230 New York Avenue, New York, NY 10021-6399, United States



Scopus habilitou na súa ferramenta unha utilidade chamada *Scopus Affiliation Identifier* que permite identificar de xeito único unha institución.

A API facilitada por Scopus permite obter o identificador de *Affiliation* dunha institución.

- **Materias:** A materia dun documento en Scopus non está directamente referida no arquivo de texto; para identifica-la materia dun documento hai que facelo a través da revista na que dito documento foi publicado.

No caso de Scopus, unha revista só ten asociada unha materia

- **Referencias citadas:** En Scopus, a información de referencias citadas ven identificada pola etiqueta "N1" e a palabra References (N1 - References:)

Cada referencia a outro documento ven nunha liña nova separada por punto e coma (;).

Os datos relativos a unha referencia citada dependerán da publicación, pero un patrón xenérico de cada referencia sería:

Autores, Título, Ano de publicación entre paréntesis, volumen, páxinas;

Por exemplo:

- N1 - References:** Leslie, A., Data Collection and Processing (1993) Proceedings of the CCP4 Study Weekend, pp. 44-51. , 29-30 January compiled by L Sawyer, N. Isaac, S. Bailey;
 Leslie, A.G.W., (1987) Proceedings of the Daresbury Study Weekend at Daresbury Laboratory, pp. 39-50. , 23-24 January compiled by J. R. Helliwell, P. A. Machin, and M. Z. Papiz;
 Kabsch, W., (1988) J. Appl. Cryst., 21, pp. 67-81;
 Kabsch, W., (1988) J. Appl. Cryst., 21, pp. 916-924;
 Rossmann, M.G., (1984) Proceedings of the Study Weekend at Daresbury Laboratory, pp. 28-40. , Biological Systems: Structure and Analysis, 24-25 March, (G. P. Diakun and C. D. Garner, eds.), Science and Engineering Research Council, Daresbury;
 Rossmann, M.G., Leslie, A.G.W., Abdel-Meguid, S.S., Tsukihara, T., (1979) J. Appl. Cryst., 12, pp. 570-581;
 Rossmann, M.G., (1979) J. Appl. Cryst., 12, pp. 225-238;
 Rossmann, M.G., Erickson, J.W., (1983) J. Appl. Cryst., 16, pp. 629-636;

O feito de que a referencias citadas non veñan codificadas dun xeito perfectamente identificable dificulta a súa interpretación e por tanto explotación.



- **Palabras clave:** Agás que o deseño do modelo de datos así o aconsellara, non sería preciso rexistrar por separado cada unha das palabras clave que aparecen asociadas a un documento.

- **Citas e Referencias citadas:** O número de citas que ten o documento correspondese coas veces que foi citado por outros documentos. Aparece tra-la etiqueta "N1" e a frase "Cited By", indicándose ademáis o ano dende o que se contan as citas. Por exemplo: N1 - Cited By (since 1996): 39425.
Scopus non facilita o reconto de Referencias citadas nunha liña ou etiqueta, pero correspondese co número de liñas (referencias citadas) que veñen tra-la etiqueta N1– References.

- **DOI:** Nalgúns casos, Scopus facilita o Digital Object Identifier dun documento, aparecendo tra-la etiqueta "N1" e a palabra "doi"
Por exemplo: **N1 - doi:** 10.1016/S0076-6879(97)76066-X



5.- IEDCYT

Neste apartado detállase a información dispoñible, estrutura e características dos datos procedentes do IEDCYT (*Instituto de Estudios Documentales sobre Ciencia y Tecnología*).

5.1.- Introducción

O IEDCYT proporciona ao Consorcio de Bibliotecas información bibliográfica básica contida nas seguintes bases de datos:

- ICYT (Ciencia e Tecnoloxía)
- ISOC (Ciencias Sociais e Humanidades)
- IME (Biomedicina)

Estas bases de datos bibliográficas conteñen a produción científica publicada en España dende os anos 70. Recollen fundamentalmente artigos de revistas científicas e de forma selectiva actas de congresos, series, compilacións, informes e monografías.

5.2.- Acceso á información

O IEDCYT dispón dun acceso web na Internet que permite descarga-la información de publicacións, aínda que esta opción é pouco operativa.

Outra posibilidade é solicita-la información en formato electrónico. Neste caso, o CD-ROM facilitado polo IEDCYT contén os datos bibliográficos rexistrados nun formato propio e unha aplicación desenvolvida polo IEDCYT para explota-los.

En calquera caso, o IEDCYT non dispón polo momento da posibilidade de obte-la información vía *Web Services*, como no caso de Scopus ou ISI.

5.3.- Información dispoñible

Neste apartado identifícase a información de publicacións científicas proporcionada polo IEDCYT.



5.3.1.- ICYT (Ciencia e tecnoloxía)

Trátase dunha base de datos referencial e bibliográfica que recolle a literatura científica contida en publicacións españolas de ciencia e tecnoloxía.

Entre outros temas, contén información de: Agronomía, Astronomía, Astrofísica, Ciencias da vida, Ciencias da terra e do espazo, Farmacoloxía, Física, Matemáticas, Química e tecnoloxías tales como: Aeronáutica, Alimentación, Bioquímica, Biotecnoloxía, Electrónica, Industrial, Instrumentación, Médica, Medio ambiente, Naval, Nuclear, Ordenadores, etc.

As fontes do IEDCYT son 747 publicacións periódicas editadas en España (en Febreiro de 2008) especializadas nas áreas citadas, fundamentalmente revistas ademais de monografías, actas de congresos, informes e tesinas.

Os datos facilitados conteñen información dende o 1979 ata a data actual, cunha actualización diaria e un crecemento anual de 8.000 referencias. En Febreiro de 2008 o volume de información é de 173.094 rexistros.

Os datos contidos no ICYT proporciona información de **Documentos** publicados. Para cada documento indícase o Tipo de Documento, Título, Idioma, Ano de publicación, etc... Ademais cada documento ten asociada a seguinte información que é preciso ter en conta para incorporar ao Sistema de Análise:

- **Autores:** Autor/es do documento.
- **Institucións:** Institución ou conxunto de Institucións que colaboran na elaboración do documento.
- **Palabras clave:** Conxunto de palabras (ou frases) clave que identifican o contido dun documento.

5.3.2.- ISOC (Ciencias sociais e humanidades)

Base de datos referencial e bibliográfica que fundamentalmente recolle referencias de artigos publicados en revistas españolas do ámbito das ciencias sociais e humanas.



Entre outros temas, contén información de: Antropoloxía, Arqueoloxía, Belas Artes, Biblioteconomía e Documentación, Ciencias Políticas, Dereito, Economía, Educación, Filosofía, Xeografía, Historia, Literatura, Psicoloxía, Socioloxía, etc.

As fontes do ISOC son 2.077 publicacións periódicas editadas en España (Febreiro de 2008) especializadas nas áreas citadas, fundamentalmente revistas ademais de monografías, actas de congresos, informes e tesinas cun crecemento anual de 23.000 referencias. En Febreiro de 2008 o volumen de información é de 497.854 rexistros.

A base de datos ISOC pode consultarse completa ou por Áreas Temáticas, que son as seguintes:

- América Latina
- Antropoloxía Social
- Arqueoloxía
- Arte
- Biblioteconomía e Documentación
- Dereito
- Economía
- Educación
- Filosofía
- Xeografía, Urbanismo e Arquitectura
- Historia
- Lingua e Literatura
- Psicoloxía
- Socioloxía e Ciencias Políticas

Para cada unha das áreas temáticas emprégase unha clasificación diferente.

Na base de datos do ISOC rexístrase información de **Documentos** publicados. Para cada documento hai asociada a seguinte información:

- **Autores:** Autor/es do documento.



- **Institucións:** Institución ou conxunto de Institucións que colaboran na elaboración do documento.
- **Materias:** Un documento pode ter asociadas ata catro materias temáticas.
- **Palabras clave:** Conxunto de palabras (ou frases) clave que identifican o contido dun documento.

5.3.3.- IME (Biomedicina)

Base de datos referencial e bibliográfica que contén información de publicacións de Biomedicina (Administración sanitaria, Farmacia clínica, Medicina experimental, Microbioloxía, Psiquiatría, Saúde pública).

En Febreiro de 2008 as fontes do IME son 321 publicacións periódicas editadas en España e conteñen información dende o ano 1971 cun volume de información de 263.206 rexistros.

Os datos procedentes de IME conteñen información de **Documentos** publicados. Para cada documento hai asociada a seguinte información:

- **Autores:** Autor/es do documento.
- **Institucións:** Institución ou conxunto de Institucións que colaboran na elaboración do documento.
- **Palabras clave:** Conxunto de palabras (ou frases) clave que identifican o contido dun documento.

5.4.- Estructura da información

Neste apartado descríbese a estrutura dos arquivos de texto proporcionados polo IEDCYT e que conteñen a información de publicacións científicas.

5.4.1.- ICYT (Ciencia e tecnoloxía)

A estrutura dos arquivos ICYT pode verse na seguinte táboa:



Etiqueta	Descrición
Documento	Número que identifica a cada un dos documentos na base de datos.
Autor	Personas físicas ou xurídicas que firman o documento. O nome propio aparece desenvolvido. Os apelidos preceden ao nome. Cada autor ven separado por un punto e coma (;).
Título	Título do documento en español.
Inglés-Título	Título do documento en inglés.
Otro-Idioma-Título	Título do documento en idioma diferente ao español ou inglés.
Lugar-Trabaja	Indica a/s institución/s, pública ou privada, na que traballa cada un dos autores da publicación. Cada institución ou departamento ven separado por un punto e coma (;). Exprésase mediante abreviaturas dacordo coa norma ISO 4/1984.
Obra-Colectiva-Título	Título dos libros de autoría colectiva en que publicara o documento.
Revista	Identificación da publicación na que se atopa o texto orixinal do documento referenciado.
Datos-Fuente	Datos de ano, volume, número, páxinas que identifican o documento referenciado nunha publicación. Soamente se pode buscar por ano ou intervalo de anos. Tras esta etiqueta está o ano de publicación do documento.
ISSN	Número internacional que identifica unívocamente a cada revista.
ISBN	Número internacional que identifica unívocamente a cada libro.
Idioma	Lingua orixinal en que está escrito o documento.
Tipo-Documento	Indica o tipo de documento: Artigo de revista, Acta de congreso, Capítulo de monografía, Libro ou Monografía completo, Informe técnico ou Tesis doutoral.
Modo-Documento	Campo que indica o enfoque ou perspectiva que caracteriza ao documento. Poden ser: Biografía, Carta ao editor, Conferencia ou Discurso, Debate, Directorio, Estado da cuestión ou Revisión, Editorial, Entrevista, Necroloxía, Homenaxe, Informe, Memoria, Mesa redonda, Nota, Norma, Ponencia ou Comunicación, Proxecto, Resumo, Reunión e Traducción.
Clasificación	Cada documento clasifícase segundo a súa temática ou materia. Emprégase a Nomenclatura Internacional da UNESCO para os campos de Ciencia e Tecnoloxía. Exprésase mediante un código de seis díxitos organizado segundo unha estrutura xerárquica na que os dous primeiros díxitos corresponden ao campo temático, os dous seguintes á disciplina e os dous últimos á subdisciplina. Por exemplo: 120000 ... 129999 Matemáticas 210000 ... 219999 Astronomía y Astrofísica 220000 ... 229999 Física 230000 ... 239999 Química



Etiqueta	Descripción
Congreso	Nome, ano e lugar do congreso, seminario, xornada, etc. ao que pertence o documento.
Descriptoros	Palabras clave que representan o contido principal do documento. Poden ser unitermos ou termos compostos.
Identificadores	Nomes propios de persoas, institucións, obras artísticas ou literarias, empresas, partidos políticos aos que se refire o documento.
Topónimos	Identificadores xeográficos, físicos, históricos ou xeopolíticos aos que se refire o documento.
Resumen	Síntesis do contido do documento polo autor.
url	Campo específico que presenta a URL de cada artigo de revista electrónica.
Serie	Nome da publicación periódica cando non é revista.
Editor	Lugar de edición, nome do editor e ano de publicación do libro.

Exemplo dun rexistro no arquivo de texto ICYT:

Base: ICYT
Documento: 173468
Autor: Rodríguez Bugarín, Miguel; Novales Ordax, Margarita; Orro Arcay, Alfonso
Título: Alta velocidade y territorio. Algunas experiencias internacionales
Inglés-Título:
Otro-Idioma-Título:
Lugar-Trabaja: Univ.A Coruña,A Coruña,España
Obra-Colectiva-Título:
Revista: Ingeniería y territorio
Datos-Fuente: 2005, (70): 4-11, 8 Ref
ISSN: 1695-9647
ISBN:
Idioma: Español
Tipo-Documento: Artículo de revista
Modo-Documento:
Clasificación: 332304 TRANSITO RAPIDO
Congreso:
Descriptoros: Ferrocarril de alta velocidade; Transporte ferroviario
Identificadores:
Topónimos: Japón; Francia; Alemania
Resumen:
Ing-Resumen:
Url:



5.4.2.- ISOC (Ciencias sociais e humanidades)

A estrutura xenérica dos arquivos ISOC é a seguinte:

Etiqueta	Descrición
Documento	Número que identifica a cada un dos documentos na base de datos.
Autor	Personas físicas ou xurídicas que firman o documento. O nome propio aparece desenvolvido. Os apelidos preceden ao nome. Cada autor ven separado por un punto e coma (;).
Título	Título do documento en español.
Inglés-Título	Título do documento en inglés.
Otro-Idioma-Título	Título do documento en idioma diferente ao español ou inglés.
Lugar-Trabaja	Indica a/s institución/s, pública ou privada, na que traballa cada un dos autores da publicación. Cada institución ou departamento ven separado por un punto e coma (;). Exprésase mediante abreviaturas dacordo coa norma ISO 4/1984.
Revista	Identificación da publicación na que se atopa o texto orixinal do documento referenciado.
Obra-Colectiva	Título dos libros de autoría colectiva en que publicara o documento.
Datos-Fuente	Datos de ano, volume, número, páxinas que identifican o documento referenciado nunha publicación. Soamente se pode buscar por ano ou intervalo de anos. Tras esta etiqueta está o ano de publicación do documento.
ISSN	Número internacional que identifica unívocamente a cada revista.
ISBN	Número internacional que identifica unívocamente a cada libro.
Idioma	Língua orixinal en que está escrito o documento.
Tipo-Documento	Indica o tipo de documento: Actas de congresos, Artigos de monografía, Artigo de revista, Informe técnico, Libros, Monografías.
Modo-Documento	Antoloxía, Bibliografía, Carta ao editor, Conferencia ou Discurso, Críticas de arte, Crónica, Debate, Dictame, Directorio, Documentación gráfica, Editorial, Entrevista, Estado da cuestión, Revisión, Informe, Xurisprudencia, Lexislación, Léxico, Memoria, Mesa redonda, Necroloxía, Homenaxe, Norma, Nota, Ponencia. Comunicación, Proxecto, Recensión, Resumo, Sesión clínico patolóxica, Táboas estatísticas, Texto orixinal, Traducción.



Etiqueta	Descrición
Clasificación	Código de seis díxitos seguido do texto correspondente que identifica a materia do documento. Ata un máximo de catro por documento códigos diferentes por documento. Segundo sexa a área temática dos arquivos ISOC emprégase unha clasificación diferente.
Compilador-Director	Nome do compilador dunha obra ou do director dunha tese.
Congreso	Nome, ano e lugar do congreso, seminario, xornada, etc. ao que pertence o documento.
Editor	Lugar de edición, nome do editor e ano de publicación do libro.
Serie	Nome da publicación periódica cando non é revista.
Signatura	Código identificativo.
Notas	Campo de texto libre no que se recolle información diversa, p.e.: si o documento inclúe resumos en inglés, apéndices, táboas, gráficos, mapas, etc.
Descriptorios	Palabras clave que representan o contido principal do documento. Poden ser unitermos ou termos compostos.
Identificadores	Nomes propios de persoas, institucións, obras artísticas ou literarias, empresas, partidos políticos aos que se refire o documento.
Topónimos	Identificadores xeográficos, físicos, históricos ou xeopolíticos aos que se refire o documento.
Legislación	Elementos que identifican as normas xurídicas. Presentan a seguinte estrutura: Rango, Tipo, Título, Lugar, Data, Artigo.
Jurisprudencia	Datos identificativos das sentencias dictadas por tribunais de calquera instancia. A súa forma: Tipo, Órgano, Sala, Número, Data (ano-mes-día).
Periodo-Histórico	Anos a que fai referencia o documento.
Siglo	Expresanse con dúas cifras arábegas. p. e.: 19, 20 (século XIX, XX).
Resumen	Síntesis do contido do documento elaborado polo autor.
Localización	Siglas que indican a institución en cuos fondos se atopa o documento.
Url	Campo específico que presenta a URL de cada artigo de revista electrónica.

Exemplo dun rexistro procedente do arquivo ISOC:

Base: ISOC
Documento: 543181
Autor: Acuña Fariña, Juan Carlos
Título: Aspectos de la relación entre teorías de gramática y teorías de procesamiento
Inglés-Título: Aspects of the relationship between theories of grammar and theories of processing



Otro-Idioma-Título:
Lugar-Trabajo: Univ. Santiago de Compostela, España
Revista: Atlantis. Revista de la Asoc. Esp. de Est. Anglo-Norteamericanos
Obra-Colectiva:
Datos-Fuente: 2005, 27(1): 11-27, 83 Ref
ISSN: 0210-6124
ISBN:
Idioma: Inglés
Tipo-Documento: Artículo de revista
Modo-Documento:
Clasificación: 550201 Modelos. Conceptos teóricos y metodológicos
550700 Psicolingüística. Obras generales
550905 Lingüística e informática
Compilador-Director:
Congreso:
Institución-Patrocinadora:
Editor:
Serie:
Signatura:
Notas: Resumen (En)
Descriptor: Teorías gramaticales; Psicolingüística; Investigación científica; Procesamiento del lenguaje natural; Análisis lingüístico; Interdisciplinariedad
Identificadores:
Topónimos:
Legislación:
Jurisprudencia:
Periodo-Histórico:
Décadas:
Siglo:
Resumen:
Ing-Resumen:
Localización:
Url: http://www.atlantisjournal.org/Papers/27_1/011-027%20Acuña.pdf

5.4.3.- IME(Biomedicina)

A estrutura dos arquivos IME pode verse na seguinte táboa:

Etiqueta	Descrición
Documento	Identifica a cada documento.



Etiqueta	Descripción
Autor	Indica os autores xa sexan persoas físicas ou xurídicas. Cada autor vai separado por un punto e coma (;).
Título	Título do artigo en castelán.
Otro-Idioma-Título	Título en idioma diferente ao castelán.
Lugar-Trabaja	Institución en que traballan os autores. Cada institución ou departamento ven separado por un punto e coma (;).
Revista	Título da revista que publica o documento.
Datos-Fuente	Datos complementarios para a identificación da publicación. Esta etiqueta identifica o ano de publicación do documento, páxinas, volume, etc. Aquí codifícase o ano de publicación do documento.
ISSN	ISSN da publicación.
Idioma	Idioma do documento.
Tipo-Documento	Código que indica o tipo de documento.
Modo-Documento	Código que indica o contido formal do documento.
Notas	Notas
Descriptor	Palabras clave que expresan os conceptos fundamentais contidos no documento.
Resumen	Resumo.
Título obra colectiva	Título da obra colectiva.
Título en inglés	Título en inglés.
Topónimos	Identificadores xeográficos.
Congreso	Congreso.
Descriptor Sec.	Palabras clave que expresan aspectos secundarios contidos no documento.
Identificadores	Nomes de persoas, organismos, etc.. aos que se refire o documento.
Localización	Código da Institución que indexou o documento.

Exemplo dun rexistro no arquivo de texto IME:

<p>Base: IME Documento: 279384 Autor: Barreiro, M; Núñez, C; Domínguez Muñoz, JE; Lorenzo, A; Barreiro, F; Potel, J; Peña, AS Título: Asociación entre las mutaciones del gen NOD2/CARD15 con los antecedentes quirúrgicos por enfermedad de Crohn Otro-Idioma-Título: Lugar-Trabaja: Hospital Clínico Universitario. Departamento de Gastroenterología. Santiago de Compostela. La Coruña. España;Vrije Universiteit Medish Centrum. Laboratory of Inmunogenetics. Amsterdam. Holanda;Hospital Clínico Universitario. Departamento de Cirugia. Santiago de Compostela. La Coruña. España;Vrije Universiteit Medish Centrum. Department of Gastroenterology. Amsterdam. Holanda Revista: Revista Española de Enfermedades Digestivas</p>
--



Datos-Fuente: 2005, 97 (8): 547-553, 25 ref.

ISSN: 1130-0108

Idioma: Inglés

Tipo-Documento: Artículo de revista

Modo-Documento:

Notas:

Descriptoros:

Resumen:

Ing-Resumen:

Url:

5.5.- Consideracións a ter en conta

Aspectos a ter en conta sobre os datos de IEDCYT:

- **Materias:** Nos datos procedentes do IEDCYT a materia temática dun documento ven codificada no arquivo de texto tra-la etiqueta "Clasificación:", cando para ISI e Scopus esta clasificación en materias está asociada á publicación á que pertence o artigo ou documento.
No caso de documentos ICYT a clasificación empregada é a da UNESCO e un documento ten asignada unicamente unha materia.
En documentos ISOC, segundo se trate dunha área temática emprégase unha clasificación diferente.
Os documentos de IME non teñen unha clasificación asociada.
- **Institucións:** A institución ou institucións que colaboran na elaboración dun documento veñen tra-la etiqueta "Lugar-Trabajo:".
- **Citas e Referencias Citadas:** Nos datos procedentes do IEDCYT non están rexistradas as citas nin as referencias citadas dun documento.
- **Publicacións:** O IEDCYT dispón dun directorio de revistas que pode ser de utilidade para cargar no sistema, si ese fora o caso, aquelas revistas que non estén rexistradas en Journal Citation Reports.



Este directorio está organizado do mesmo xeito que a información de documentos:

- Ciencia e tecnoloxía
- Ciencias sociais e humanidades
- Biomedicina



6.- Journal Citation Reports

Nos próximos apartados detállase a información dispoñible, estrutura e características dos datos de Journal Citation Reports que se incorporarán ao Sistema de Análise.

6.1.- Introducción

Dentro dos servizos proporcionados polo ISI Web of Knowledge atópase Journal Citation Reports, que ofrece información de publicacións científicas a nivel mundial e datos do seu impacto e influencia na comunicade científica.

6.2.- Acceso á información

A través do portal web de ISI Web of Knowledge pode facerse unha consulta sobre JCR e descarga-la información necesaria a un ficheiro de texto, donde cada columna está separada por un punto e coma (;).

En principio non é posible a descarga de información de JCR a través dunha API con Web Services, coma no caso do ISI Web Of Science.

6.3.- Información dispoñible

Os datos de citas de JCR proveñen de 7.600 revistas de todo o mundo relativas a 220 disciplinas.

Cada edición anual contén os datos da publicación do ano anterior e amosa a relación entre as revistas citadoras e as revistas citadas.

Estes datos permiten facer comparacións entre revistas, xa que JCR acumula o número de citas e artigos de practicamente tódalas especialidades das ciencias, tecnoloxía e ciencias sociais.

Journal Citation Reports ten dúas edicións:

- **JCR Science Edition:** Contén os datos de máis de 5.900 revistas en 172 categorías temáticas.
- **JCR Social Sciences Edition:** Contén máis de 1.700 revistas en 55 categorías temáticas.



Journal Citation Reports contén, para un ano, información de **Publicacións** científicas, indicando o seu título, título abreviado, factor de impacto, etc.

Cada publicación ten asociada a seguinte información, que é preciso ter en conta:

- **Materia:** Materia ou materias científicas que cubre unha determinada publicación.

Ademais e para cada rexistro contido en Journal Citation Reports, o Consorcio de Bibliotecas añade unha serie de valores calculados sobre os datos básicos.

6.4.- Estructura da información

Na seguinte táboa identifícanse os datos que poden obterse de Journal Citation Reports e o seu significado.

Columna	Descrición
Issn	Issn da publicación.
tituloCorto	Título corto da publicación
tituloLargo	Título largo da publicación.
TipoMateria	Materias ISI. Unha publicación pode estar asociada a máis dunha materia.
fimp	Factor de impacto. Identifica a frecuencia con que se cita un artigo medio dunha revista nun ano en particular. Pode empregarse esta cifra para avaliar ou compara-la importancia relativa dunha revista coas outras do mesmo campo. Calculase dividindo o número de citas rexistradas a artigos publicados nos dous anos anteriores entre o número total de artigos publicados nos dous anos anteriores.
totalCites	Número total de citas. Indica o número de veces que cada revista foi citada en tódalas revistas incluídas na base de datos de JCR para o ano ao que pertencen os datos.
immediacyIndex	Índice de inmediatez. Mide a frecuencia con que se cita un artigo promedio dunha revista dentro do mesmo ano de publicación. Este índice é útil para evalua-las revistas que publican investigacións de vanguardia. Calculase dividindo o número de citas a artigos nun ano dado entre o número de artigos publicados no mesmo ano.
Articles	Número de artigos. O número de artigos publicados nunha revista para un ano dado.



Columna	Descrición
citedHalfLife	Vida media das citas recibidas. Número de anos a partires do ano en curso que representan o 50% do número total de veces que se citou a revista no ano en curso. Indica a idade dos artigos citados. É un valor alfanumérico que pode conter cadeñas como ">10.0"
citingHalfLife	Vida mediadas citas incluídas. Número de anos a partires do ano en curso que representa o 50% das referencias citadas en artigos publicados nunha revista no ano en curso.
Información añadida polo Consorcio de Bibliotecas	
fimp_orig	Información de control
totalG	Para unha publicación, número de publicacións existentes no JCR dun ano coa mesma materia.
1tG	Primeiro tercio do valor de totalG ($1/3 * totalG$)
2tG	Segundo tercio do valor de totalG ($2/3 * totalG$)
posicionG	Posición da publicación respecto do valor de totalG e segundo o seu factor de impacto. A maior factor de impacto menor valor de posicionG.
ABC	Indica en qué tercio se sitúa unha publicación, dependendo do seu valor de posicionG. A: $1tG \leq posicionG$ B: $1tG < posicionG \leq 2tG$ C: $2tG < posicionG$
MIX	Valor de ABC concatenado coa materia da publicación. Por exemplo: B // SCI-MATHEMATICS
posicionABC	Posición da publicación respecto do valor de ABC e segundo o seu factor de impacto. A maior factor de impacto menor valor de posicionABC.
deABC	Calculase como o número de publicacións que teñen o mesmo valor ABC (A, B ou C) dentro dunha mesma materia e por tanto co mesmo valor de posicionG.
fiMaxABC	Factor de impacto máximo dentro do mesmo valor de ABC e Materia.
fiMinABC	Factor de impacto mínimo dentro do mesmo valor de ABC e Materia.

O seguinte é un exemplo de información rexistrada en Journal Citation Reports:

Issn	0960-7692
tituloCorto	ULTRASOUND OBST GYN
tituloLargo	ULTRASOUND IN OBSTETRICS & GYNECOLOGY
fimp	2,288
TipoMateria	SCI-ACOUSTICS
totalG	28
1tG	9



2tG		18
posicionG		1
ABC	A	
MIX	A // SCI-ACOUSTICS	
posicionABC		1
deABC		9
fiMaxABC		2,288
fiMinABC		1,189
totalCites		4352
immediacyIndex		0.583
Articles		216
citedHalfLife		5.1
fimp_orig		2.288

6.5.- Consideracións a ter en conta

A continuación recóllense certas consideracións relacionadas cos datos de Journal Citation Reports:

- **Factor de impacto:** O factor proporciona unha medida da frecuencia coa que un artigo medio dunha publicación científica se cita nun ano en particular. Este valor pode variar para cada ano.
Representa o ratio entre citas e documentos citables publicados recentemente.
O factor de impacto para un ano dado calcúlase dividindo o número de citas para os dous anos anteriores entre o número de documentos citables dos dous anos anteriores. Por exemplo:
Factor de impacto 2007 = (Citas feitas no 2007 a artigos publicados en 2006 e 2005) / (Número de artigos publicados en 2006 e 2005).
- **Materias:** Unha publicación pode ter asignada máis dunha materia ISI e, ademais, esta relación de materias asociadas a unha publicación pode variar co paso dos anos.



7.- Grupos de Investigación

Nos seguintes puntos identificase a información dispoñible, estrutura e características dos datos pertencentes á Base de Datos de Grupos de Investigación do Sistema Universitario Galego.

7.1.- Introducción

O Sistema Universitario Galego rexistra nesta base de datos a información dos Grupos de Investigación, cos investigadores pertencentes a cada grupo e os artigos que publican.

7.2.- Acceso á información

O Consorcio de Bibliotecas dispón de acceso á base de datos de Grupos de Investigación.

7.3.- Información dispoñible

Da información rexistrada na base de datos, o Sistema de Análise necesita incorpora-los datos dos **Grupos de Investigación**, **Investigadores** e os **Artigos** que éstos publicaron.

Ademáis, os datos almacenados neste sistema permitirán completa-la información de autores e institucións.

7.4.- Estructura da información

Na seguinte imaxe amósase a estrutura da base de datos de Grupos de Investigación.

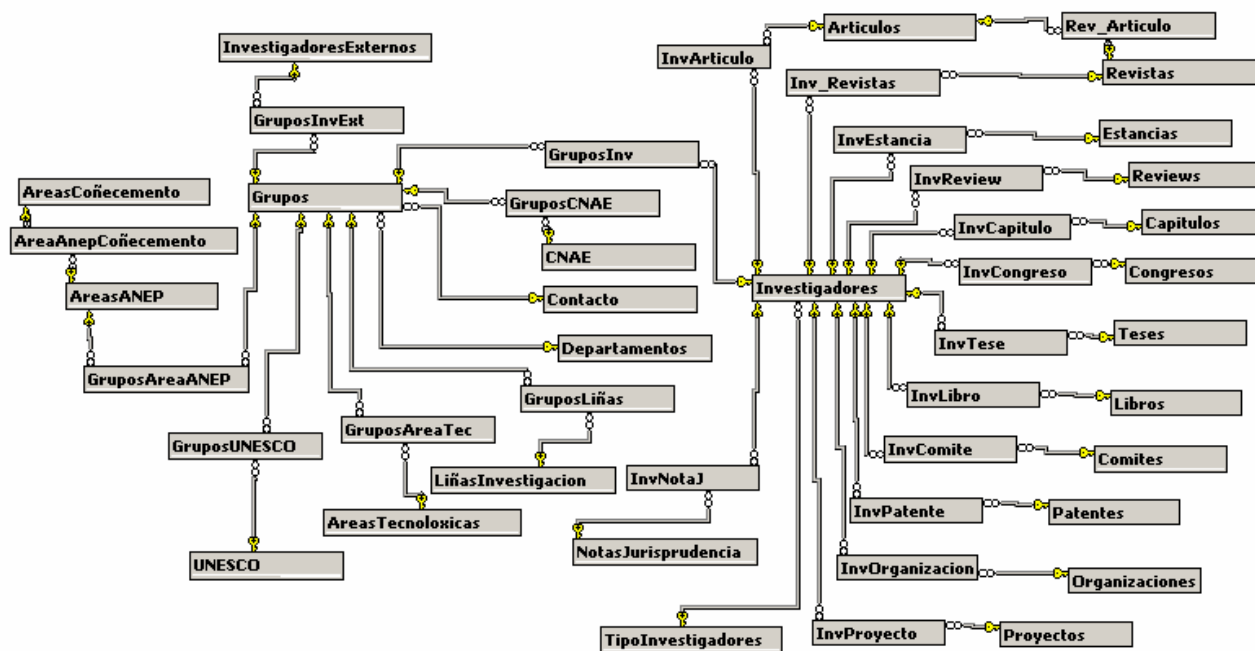


Figura 2: Modelo de datos da BD de Grupos de Investigación



8.- Outras fontes de datos

8.1.- Introducción

Ademáis de toda a información proporcionada por cada unha das diferentes fontes que alimentará o Sistema de Análise de Producción Científica, deben preverse os mecanismos necesarios para incorporar ao sistema outros datos, como listas actualizadas de materias científicas, institucións ou calquera outra información que sexa necesaria.

En moitos casos, estes datos definirán unha lista normalizada de valores, o que facilitará a carga de datos e tamén as análises comparativas de información procedente de fontes distintas.

Entre esta información independente das fontes e que será preciso cargar no sistema, identifícase:

- **Institucións Científicas:** Relación de institucións investigadoras españolas elaborada polo proxecto SCImago da Universidade de Granada.
- **Categorías ANEP:** Relación de materias científicas definidas pola *Agencia Nacional de Evaluación y Prospectiva (ANEP)*.
- **Categorías UNESCO:** Nomenclatura internacional de materias científicas proposta pola UNESCO para os campos da ciencia e a tecnoloxía.
- **Categorías ISOC (IEDCYT):** Os documentos de Ciencias sociais e humanas pertencentes ao IEDCYT están organizados en categorías de materias propias para cada base de datos.
- **Listado de publicacións de Scopus:** Incorpora-lo listado de publicacións de Scopus é preciso porque aquí é donde se especifican as materias de Scopus e tamén porque permitiría rexistrar no sistema publicacións ás que non se fai referencia en Journal Citation Reports.



- **Listado de revistas rexistradas en IEDCYT:** Conxunto de revistas sobre as que se rexistra información nas bases de datos do IEDCYT (ICYT, ISOC e IME). Este listado complementaríase as publicacións recollidas en Journal Citation Reports e Scopus.

8.2.- Institucións Científicas

A Universidade de Granada, a través do proxecto Atlas of Science, incorpora a unha base de datos propia información procedente do ISI Web of Knowledge.

Como resultado deste proxecto dispoñen nas súa base de datos dunha relación de institucións científicas españolas e estranxeiras. Inicialmente esta lista está conformada polas institucións que aparecen nos ficheiros do ISI Web of Science (etiqueta "C1") que posteriormente se depura e actualiza con información máis precisa.

O Consorcio de Bibliotecas ten acceso a estes datos e polo tanto empregará a lista depurada de institucións como fonte de datos ao sistema e referencia nas cargas de datos.

Esta información será incorporada nunha carga inicial do Sistema e podería empregarse posteriormente para complementala información rexistrada, pero sería preciso establecer os procesos necesarios para engadir institucións non rexistradas pola Universidade de Granada e que estas poidan ser modificadas e actualizadas polo propio Consorcio.

Estes datos están na táboa *insti* (institucións españolas) e *instituciones* (institucións estranxeiras) da súa base de datos, que ten a seguinte estrutura (Febreiro de 2008):

Nome	Descrición
Idins	
IdInsa	
IdInsNor	
rs_organization	Nome da institución procedente de ISI.
organizacion	Nome da institución modificada pola U. Granada.
country	Código do país
rs_city	Nome da cidade procedente de ISI



Nome	Descrición
ciudad	Nome da cidade modificado pola U. Granada
ccaa	Código da Comunidade Autónoma.
sector	Sector a que se dedica a institución
normalizada	
aceptada	
idssinsti	
observaciones	
usuario	
fecha	
hora	
fecham	
fechacarga	
elentic	
aborg	
borrado	
activa	
ncat	

A continuación amósase un exemplo da información contida na táboa de institucións:

Idins	1177
IdInsa	2613
IdInsNor	1177
rs_organization	Ctr Supercomputac Galicia
organizacion	
country	1
rs_city	Santiago De Compostela
ciudad	Santiago de Compostela
ccaa	11
sector	4
normalizada	True
aceptada	True
idssinsti	212039593
observaciones	
usuario	
fecha	
hora	
fecham	
fechacarga	
elentic	
aborg	CSGAL
borrado	False
activa	True
ncat	8



No caso das institucións estranxeiras (táboa instituciónese), esta é a estrutura cun exemplo de datos:

idinse	idinsnor	rs_organization	rs_city	fechacarga	numcountry	origen
10	10	106 Ippocratous St	Athens	18/04/2007	32	
32	32	Sir Alister Hardy Fdn Ocean Sci Lab	Plymouth	18/04/2007	6	
33	33	Sir Charles Gairdiner Hosp	Nedlands	18/04/2007	36	

8.3.- Categorías ANEP

Relación de materias científicas definidas pola *Agencia Nacional de Evaluación y Prospectiva* (ANEP) e que permiten clasifica-la produción científica en diferentes categorías.

Constituirá a clasificación de referencia empregada polo Consorcio de Bibliotecas como lista normalizada, permitindo deste xeito facer uso das materias ANEP para clasifica-los indicadores rexistrados para cada unha das diferentes fontes de datos.

Para poder emprega-las categorías ANEP como lista normalizada é preciso definir e manter as relacións entre cada unha das materias rexistradas para un documento na súa orixe (materias ISI, Scopus, UNESCO e materias ISOC) coa súa correspondente materia ANEP.

A clasificación ANEP evoluciona e cambia co paso do tempo. É preciso empregar sempre a última versión, o que require a actualización das relacións e dos datos xa cargados no sistema.

As 25 áreas científicas de ANEP agrúpanse en diferentes ámbitos, como se amosa na seguinte táboa:

Ábito de coñecemento	Área Científica
CIENCIAS DE LA SALUD	MEDICINA
	PSICOLOGÍA Y CIENCIAS DE LA SALUD
CIENCIAS EXPERIMENTALES	AGRICULTURA
	BIOLOGIA MOLECULAR Y CELULAR Y GENÉTICA
	BIOLOGÍA VEGETAL, ANIMAL Y ECOLOGÍA
	FÍSICA Y CIENCIAS DEL ESPACIO



Ámbito de coñecemento	Área Científica
	FISIOLOGÍA Y FARMACOLOGÍA
	GANADERIA Y PESCA
	GEOCIENCIAS
	MATEMATICAS
	QUIMICA
	TECNOLOGIA Y CIENCIA DE LOS MATERIALES
CIENCIAS SOCIALES Y JURIDICAS	CIENCIAS SOCIALES
	DERECHO
	ECONOMIA
ENSEÑANZAS TECNICAS	CIENCIAS Y TECNOLOGIA DE LA COMPUTACION
	INGENIERÍA CIVIL Y ARQUITECTURA
	INGENIERIA ELECTRICA, ELECTRONICA Y AUTOMATICA
	INGENIERIA MECANICA, NAVAL Y AERONAUTICA
	TECN. Y CIENCIAS DE LA ALIMENTACION
	TECNOLOGIA ELECTRONICA Y DE LAS TELECOMUNICACIONES
	TECNOLOGIA QUIMICA
HUMANIDADES	FILOLOGIA Y FILOSOFIA
	HISTORIA Y ARTE
MULTIDISCIPLINAR	MULTIDISCIPLINAR

8.4.- Categorías UNESCO

Nomenclatura proposta pola UNESCO para establecer unha clasificación nos campos da ciencia e a tecnoloxía.

Esta clasificación codifícase mediante un código de seis díxitos organizado segundo unha estrutura xerárquica na que os dous primeiros valores corresponden ao campo temático, os dous seguintes á disciplina e os dous últimos á subdisciplina. Por exemplo:

120000 ... 129999 Matemáticas

210000 ... 219999 Astronomía y Astrofísica

- **Campo temático.** Apartados máis xenéricos. Codificado con dous díxitos. Comprende varias disciplinas.
- **Disciplinas:** Descrición xeneral de grupos de especialidades en Ciencia e Tecnoloxía. Son apartados codificados con catro díxitos.
- **Subdisciplinas:** Son as entradas máis específicas da nomenclatura e representan as actividades que se realizan dentro dunha disciplina. Están codificadas con seis



díxitos. Deben corresponderse coas especialidades individuais en Ciencia e Tecnoloxía.

As áreas de Ciencia e Tecnoloxía definidas pola UNESCO son:

Código	Nome Campo
11	Lógica
12	Matemáticas
21	Astronomía y Astrofísica
22	Física
23	Química
24	Ciencias de la Vida
25	Ciencias de la tierra y del espacio
31	Ciencias agrarias
32	Ciencias médicas
33	Ciencias tecnológicas
51	Antropología
52	Demografía
53	Ciencias económicas
54	Geografía
55	Historia
56	Ciencias jurídicas y Derecho
57	Lingüística
58	Pedagogía
59	Ciencia política
61	Psicología
62	Ciencias de las artes y las letras
63	Sociología
71	Ética
72	Filosofía

8.5.- Categorías ISOC (IEDCYT)

Algunhas das áreas temáticas nas que se agrupa a información de Ciencias Sociais e Humanas (ISOC) do IEDCYT teñen unha clasificación temática propia. Por exemplo:

ISOC-América Latina: Clasificación temática propia. Códigos de 6 díxitos e nome.

Por exemplo: 850200; 200300, América Latina. Documentación y política científica; Fuentes documentales.



ISOC-Arte: Clasificación temática propia con 17 categorías. Códigos de 6 dígitos e nome, ata un máximo de catro por documento.

Por exemplo: 850200; 851130; 200301; 151428, América Latina. Documentación y política científica; América Latina. Música y danza; Documentos primarios: material impreso y no impreso.; Música popular

ISOC-Psicología: 16 categorías, baseada na clasificación de *Psychological Abstracts* elaborada pola *American Psychological Association (APA)*.

Por exemplo: 651203; 651505, Trastornos de aprendizaje y retraso mental; Educación especial y preventiva.

8.6.- Publicacións de Scopus

Scopus mantén unha relación propia de publicacións, que ademáis pode obterse a través da súa páxina web.

No arquivo que se obtén de Scopus a estrutura cambia lixeiramente se os datos se refiren a publicacións procedentes das bases de datos CSA.

Publicacións CSA:

Campo	Comentario
ISSN	ISSN
E-ISSN	e-ISSN
Source Title (CSA excl.)	Título (non inclúe títulos CSA).
Medline-sourced Title	Indica se se trata dun título procedente de Medline.
Open Acces status	Indica o estado de Open Acces. Posibles valores: Not OA OA but not registered OA but not registered, formerly OA registered OA registered
Added to list since Nov. 2007 (CSA excl.)	Identifica se foi añadida ao listado dende a versión anterior de publicacións Scopus.
Source Type	Tipo de fonte. Posibles valores: Book Series Conference Proceedings Journal Trade Journal
Active or Inactive	Indica se a publicación está activa ou inactiva.



Campo	Comentario
Coverage	Indica a cobertura temporal. Ano inicial - ano final.
Title history indication	Permite identifica-la evolución histórica dun título. Debe interpretarse en conxunción cos seguintes campos (Related title #).
Related title 1	Títulos relacionados.
Related title 2	
Related title 3	
Related title 4	
Publisher's Name	Nome do editor.
Country	País.
Classification Top Level	Clasificación (Materias).
Classification (Primary) Level 1	
Classification 2	
Classification 3	
Classification 4	
Classification 5	
Classification 6	
Classification 7	
Classification 8	
Classification 9	
Classification 10	
Classification 11	
Classification 12	
Classification 13	
Classification 14	

Publicacións non CSA:

Campo	Comentario
ISSN	ISSN
E-ISSN	e-ISSN
CSA Source Title	Título.
CSA databases	Nome da base de datos CSA á que pertence a publicación.
Added since last version	Identifica se foi añadida ao listado dende a versión anterior de publicacións Scopus.
Source Type	Tipo de fonte.
Active or Inactive	Indica se a publicación está activa ou inactiva.
Coverage	Indica a cobertura temporal. Ano inicial - ano final.



Campo	Comentario
Open Acces status	Indica o estado de Open Acces. Posibles valores: Not OA OA but not registered OA but not registered, formerly OA registered OA registered
Title history indication	Permite identifica-la evolución histórica dun título. Debe interpretarse en conxunción cos seguintes campos (Related title #).
Related title 1	Títulos relacionados.
Related title 2	
Related title 3	
Related title 4	
Publisher's Name	Nome do editor.
Country	Páís.
Region	Rexión
Classification Top Level	Clasificación.
Classification (Primary) Level 1	
Classification 2	
Classification 3	
Classification 4	
Classification 5	
Classification 6	
Classification 7	
Classification 8	

8.7.- Publicacións de IEDCYT

Co obxectivo de incorporar ao Sistema de Análise aquelas publicacións que non estén recollidas en Journal Citation Reports e nos listados de publicacións de Scopus, sería preciso obte-los listados de revistas das que se rexistra información no IEDCYT.



9.- Modelo de Datos

9.1.- Introducción

O modelo de datos que se plantexa para o Sistema de Análise ten que dar resposta ás necesidades de información identificadas polo Consorcio de Bibliotecas. Neste sentido, debe buscarse un deseño que facilite a consulta de información e a integración de datos procedentes de diferentes fontes.

Non é obxectivo deste documento establecer un deseño de datos concreto, pero neste apartado empréganse os conceptos de *Dimensións* e *Indicadores* como un xeito de identificar a información que deberá estar dispoñible no Sistema de Análise, aínda estando estes conceptos máis próximos a deseños dimensionais.

Así, poderíamos dar as seguintes definicións:

- **Dimensións:** Representan as diferentes entidades ou elementos de negocio identificados no sistema. Supoñen o punto de entrada á información e permiten explora-los indicadores existentes agrupandoos polas diferentes dimensións dispoñibles.
Cada Dimensión está composta dun conxunto de Atributos ou propiedades que caracterizan a cada un dos seus elementos membro.
- **Indicadores:** Por indicador entendemos unha propiedade da información que pode ser expresada nun valor cuantitativo e que é preciso medir.

En calquera caso, a seguinte é unha relación inicial que debe ser revisada e modificada, se fora preciso, no momento do deseño e desenvolvemento do sistema.



9.2.- Dimensións

A continuación propóñense un conxunto de dimensións para o Sistema de Análise de Producción Científica derivado do estudo das fontes de datos e das necesidades de información do Consorcio de Bibliotecas.

En cada caso defínese a información que contencrán e certas consideracións a ter en conta no seu deseño ou na carga de datos.

Esta primeira aproximación debe confirmarse e refinarse no momento de deseña-lo modelo de datos tendo en conta, entre outras cuestións, as posibilidades de integración de datos procedentes das distintas fontes.

9.2.1.- Data

Contén as datas para as que existe información no sistema o que permite, entre outras cousas, facer análise por ano de publicación dun documento.

Atributos:

- o Ano
- o Mes

9.2.2.- Documento

Documentos dispoñibles no Sistema (artigos pertencentes a unha publicación, actas, teses, etc.) e a información que os caracteriza: título, volume, serie, etc.

Atributos:

- o Título
- o Título secundario
- o Título noutro idioma
- o Volume
- o Número de publicación
- o Número de parte
- o Páxinas
- o Páxina inicial
- o Páxina final
- o Publicación especial
- o Suplemento



- o Nome congreso
- o Data
- o Idioma
- o DOI (Digital Object Identifier)

Integrar non sistema os documentos procedentes de cada unha das fontes é unha tarefa complexa. No momento do deseño do modelo de datos debe decidirse se é factible traballar cunha lista normalizada de documentos donde se especifique a través dun atributo a súa procedencia, ou é mellor manexar relacións de documentos independentes para cada unha das fontes.

9.2.3.- Publicación

Relación das distintas publicacións para as que se ten recollido información de produción científica.

Consideranse como publicacións revistas, libros ou aquelas publicacións definidas como tal en Journal Citation Reports ou en Scopus.

Atributos:

- o ISBN
- o ISSN
- o Título Corto
- o Título Largo
- o Título da serie
- o Editor
- o País
- o Rexión
- o Activa (Si/Non)
- o Factor de impacto

Habería que busca-lo xeito de dispoñer dunha única relación normalizada de publicacións. Neste caso poderían empregarse como valores de referencia as listas de publicacións existentes en Journal Citation Reports e en Scopus.



Nesta dimensión será preciso articula-los mecanismos de carga e actualización de datos necesarios para coñece-la evolución das publicacións ao longo do tempo: cando unha publicación cambia de nome, cando se fusiona con outra, cambia de temática, etc.

O Factor de Impacto asociado a unha publicación (con orixe en ISI) é un atributo que se establece para un ano e cambia co paso do tempo. Neste caso concreto, é preciso saber, para cada ano e publicación, cal era o seu factor de impacto.

9.2.4.- Materia

Relación de materias científicas nas que pode clasificarse un documento publicado. Por exemplo: Economía, Química, Enxeñería Civil e Arquitectura, etc.

Cada unha das fontes de datos emprega unha clasificación por materias diferente:

- **ISI**: Clasificación de materias ISI.
Esta clasificación procede da base de datos de Journal Citation Reports e contén 172 categorías en Science Citation Index e 55 categorías en Social Sciences Citation Index.
- **Scopus**: Clasificación Scopus derivada das materias asociadas a cada unha das publicacións.
Esta clasificación está organizada en 27 áreas temáticas con ata 13 niveis de detalle.
- **IEDCYT.ICYT**: Nos datos de Ciencia e Tecnoloxía do IEDCYT empregase a clasificación da UNESCO.
Trátase dunha codificación de materias con 6 díxitos organizadas por Campo temático, Disciplinas e Subdisciplinas (con 2 díxitos para cada unha delas).
Son 24 categorías e 248 disciplinas.
- **IEDCYT.ISOC**: Clasificación temática específica para ISOC.
Cada unha das áreas temáticas dispoñibles en ISOC ten unha clasificación temática propia.



No Sistema de Análise empregarase a **Clasificación ANEP** como relación normalizada de materias. A clasificación de materias ANEP propón 25 Áreas científicas pertencentes a 6 Ámbitos do coñecemento. Os ámbitos de coñecemento non pertencen estritamente ás categorías ANEP e son establecidos pola *Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA)*.

Será preciso por tanto establece-las correspondencias entre a clasificación ANEP e as clasificacións empregadas en cada unha das fontes de datos, de xeito que a explotación de datos poida facerse por categorías ANEP independentemente da codificación empregada en orixe. En calquera caso deberá manterse tamén a codificación orixinal para cada unha das fontes.

Asimesmo, para tódalas clasificacións de materias empregadas en cada unha das fontes de datos, debe contemplarse a actualización e reorganización das mesmas: por exemplo desglosando unha materia en máis detalle ou incorporando novas categorías.

No caso das materias ANEP e materias ISI é preciso manter sempre a última versión da mesma, sendo preciso para elo actualiza-las relacións entre materias e os datos xa cargados no sistema.

Atributos Materias ANEP:

- o Código
- o Ámbito de coñecemento
- o Área científica

Atributos Materias ISI:

- o Código
- o Descrición

Atributos Materias Scopus:

- o Código
- o Categoría
- o Área Temática



- o Nivel clasificación 1
- o ...
- o Nivel clasificación 13

Atributos Materias UNESCO:

- o Código
- o Campo
- o Disciplina
- o Subdisciplina

9.2.5.- Institución

Conxunto de Institucións Científicas para as que se pode consultar datos de produción científica.

Para esta dimensión empregarase como fonte principal e referencia a relación de institucións manexada polo proxecto SCImago da Universidade de Granada. Tamén deben ser incorporadas ao sistema todas aquelas institucións ás que se faga referencia noutras fontes (ISI, Scopus e IEDCYT) que non existan na lista da Universidade de Granada.

Atributos:

- o Código
- o Nome
- o Organización
- o Cidade
- o Dirección postal
- o Páxina web

No momento de deseña-los procesos de carga e o modelo de datos haberá que determinar si se actualizan os nomes das institucións cando estes cambien ou se mantén unha lista de tódolos nomes antigos, se esta opción é posible.

9.2.6.- Sector

Sectores nos que se poden encadra-las diferentes institucións rexistradas no Sistema. Por exemplo:

- Administración.



- Centros mixtos (dobre adscrición institucional, Universidade-CSIC; Administración-Universidade)
- Empresa.
- Organismos Públicos de Investigación.
- Sistema sanitario.
- Universidade.

Atributos:

- o Código
- o Nome

9.2.7.- Autor

Relación de autores que teñan publicado algún dos documentos dispoñibles no Sistema de Análise.

Non existe unha lista de referencia ou normalizada dende a que obter os autores. Cada unha das fontes de datos que alimenta o sistema identifica aos autores que publicaron un documento empregando unha notación propia sen un valor identificativo que sexa único. Para cada documento ven rexistrado o autor ou autores mediante o apelido ou apelidos e o nome (ás veces unicamente a inicial do nome).

Así, atoparemos nomes distintos que fan referencia ao mesmo autor e nomes idénticos que fan referencia a diferentes autores.

Dispor dentro do Sistema de Análise dunha lista normalizada de Autores facilitaría a comparación dos resultados obtidos segundo a orixe da información, pero a elaboración de dita lista é unha tarefa moi complexa que require a intervención de persoas para decidir ante posibles conflitos de nomes.

Na fase de deseño deben terse en conta as dificultades plantexadas na xestión de autores. Pode optarse por manexar unha lista independente para cada fonte de datos ou deseñar unha relación común de autores donde se indique a orixe dos mesmos.



En calquera caso debe habilitarse unha ferramenta de fácil manexo para que un usuario cos permisos axeitados poida actualiza-los datos de autores e decidir se un mesmo nome fai referencia á mesma persoa ou dous nomes diferentes son o mesmo autor.

Tamén sería bo que o proceso de carga fixera un “barrido” sobre os autores a incorporar ao sistema presentando ao usuario unha lista reducida de datos a revisar.

Atributos:

- Apelido 1
- Apelido 2
- Nome
- Apelidos e Nome
- Sexo
- Galego (Si/Non)

9.2.8.- Tipo de Documento

Identifica os diferentes tipos de documento nos que clasifica a información de produción científica cada unha das fontes de datos. Por exemplo: Resumo, Artigo, Acta, Tese, etc.

Atributos:

- Código de Tipo de Documento
- Descrición de Tipo de Documento

Cada unha das fontes codifica os tipos de documento dun xeito diferente. En principio non se considera necesaria a definición dunha lista normalizada de Tipos de Documento.

A modo de exemplo, na seguinte táboa amósanse algúns dos valores de Tipo de Documento para cada unha das fontes:

ISI Web of Science	Scopus	IEDCYT
Book Review	Abstract	Articulo de revista
Correction	Book, Whole	Acta de congreso
Correction, Addition	Book chapter	Capítulo de monografía
Note	Conference proceeding	Libro o Monografía completo
Editorial Material	Newspaper	Informe técnico
Review	Thesis/Dissertation	Tesis doctoral



9.2.9.- Xeografía

Localización xeográfica asociada a cada unha das Institucións Científicas dispoñibles no Sistema de Análise.

Habilitar a información de localización de xeito independente da relación de Institucións facilita a explotación de datos.

Atributos:

- País
- Comunidade Autónoma
- Provincia

9.2.10.- Tipo de Colaboración

Para aqueles documentos que foron elaborados en colaboración por varias institucións identifica, por ámbito xeográfico, o tipo de colaboración. Posibles valores: Galega, Interrexional e Estranxeira.

Atributos:

- Código de Tipo de Colaboración
- Descrición de Tipo de Colaboración

O obxecto de incluír esta agrupación, que se deriva da autoría dos documentos, é facilitar os estudos por colaboración.

9.3.- Indicadores

Neste apartado establécese unha primeira aproximación ao conxunto de indicadores do Sistema de Análise.

Ésta é unha relación inicial que debe completarse nas fases de análise e deseño do Sistema en función das necesidades de información e dos datos obtidos das fontes.

No Sistema, e soportado polo modelo de datos, deben existir unha serie de indicadores básicos, algúns dos cales serían:



Indicadores cuantitativos:

- **Documentos:** Reconto de documentos recollidos no Sistema.
- **Documentos citables:** Reconto de documentos susceptibles de ser citados
Un documento considérase citable en función do Tipo de Documento: Artigos.

Indicadores cualitativos:

- **Factor de impacto:** Factor de impacto dunha publicación establecido por *Journal Citation Reports*.
O factor de impacto obtense do seguinte xeito:

$$FI(Ano_i) = \frac{\text{Citas en Ano}_i \text{ a documentos publicados en Ano}_{i-1} \text{ e Ano}_{i-2}}{\text{Items citables publicados en Ano}_{i-1} \text{ e Ano}_{i-2}}$$

- **Documentos citables:** Reconto de documentos susceptibles de ser citados
Un documento considérase citable en función do Tipo de Documento: Artigos.

Indicadores de estrutura e relacións:

- **Documentos citados:** Reconto de documentos que teñen alomenos unha cita.
- **Documentos non citados:** Reconto de documentos que non teñen ningunha cita.
- **Citas recibidas:** Reconto de citas feitas a un documento por outros documentos.
- **Citas realizadas:** Reconto de referencias citadas a outros documentos.
- **Autocitas:** Reconto de documentos donde o autor se cita a sí mesmo.
- **Documentos en coautoría:** Número de documentos que foron elaborados por máis dun autor.

Indicadores de colaboración:

- **Documentos en colaboración:** Número de documentos que foron elaborados en colaboración por autores pertencentes a máis dunha institución.

Ademáis, e a partires destes indicadores básicos, será preciso facer estudos empregando outros indicadores calculados.



No momento do desenvolve-lo sistema haberá que valorar as vantaxes e inconvenientes de dispoñer no modelo de datos dalgúns destes indicadores xa calculados ou se é preferible calcula-los nos informes.

Algúns dos indicadores derivados e estudos que se obterán do Sistema son:

Indicadores derivados cuantitativos:

- **% Documentos:** Porcentaxe de documentos a un nivel de detalle determinado respecto do total de documentos.
- **% Documentos citables:** Porcentaxe de documentos citables a un nivel de detalle determinado respecto do total de documentos.
- **Taxa de variación interanual:** Aumento cuantitativo do número de documentos dun ano respecto do anterior.

$$TV_n = \frac{Ndoc_n - Ndoc_{n-1}}{Ndoc_{n-1}} * 100$$

- **Índice de esforzo temático:** Actividade relativa nunha área temática determinada a través do nivel de especialización: esforzo relativo que unha institución adica a unha área temática.

$$IET_{cat1} = \frac{Ndoc_{cat1(institución)} / Ndoc \sum_{cat}(institucion)}{Ndoc_{cat1(xeografía)} / Ndoc \sum_{cat}(xeografía)}$$

- **Índice de esforzo temático relativo:** Índice de especialización temática relativizado, de xeito que poidan establecerse comparacións entre distintas institucións.

$$IER = \frac{IET - 1}{IET + 1}$$



Indicadores derivados cualitativos:

- **Factor de impacto normalizado:** Normalización do factor de impacto co fin de homoxeneizar as escalas de diferentes categorías temáticas.

$$FITM_{jc} = \frac{FI_{kc} - FI_c}{\sigma FI_c}$$

- **Factor de impacto medio esperado relativo:** Emprégase para comparar os factores de impacto normalizados (FITM) de diferentes dominios.

$$FIR_{jc} = \frac{FITM_{insti - x}}{FITM_{insti - xeografía}}$$

- **Factor de impacto medio das revistas donde se producen as citas:** Indica o factor de impacto medio das revistas donde se citan os artigos dunha revista. Refírese ao promedio do factor de impacto ponderado das revistas donde foi citada unha revista.
Calculase multiplicando o factor de impacto de cada revista (nese ano) na que se citan artigos dos anos anteriores da revista analizada polo número de artigos citados en cada revista. A suma total divídese polo número total de artigos citados.

$$FIMRC_{revista} = \frac{\sum FI_{revista_i} * \text{Num artigos}_{revista_i}}{\text{Num artigos}}$$

- **Factor de impacto ponderado:** Calculase a partires do factor de impacto (FI) dunha revista e do factor de impacto medio das revistas donde se producen as citas (FIMRC). Obtense calculando o promedio de ambos factores.

$$FIP = \frac{FIMRC + FI}{2}$$



- **Potencial investigador:** Matiza a información cualitativa coa cantidade de traballos que se publican nunha determinada revista.
Trata de relativizar o binomio calidade-cantidade, de maneira que a partires do factor de impacto normalizado se pondera o número de traballos mediante a función:

$$PI = \sum (Ndoc * (FITM + k))$$

- **Potencial investigador comparado:** Reflicte a actividade relativa nunha área temática a través do nivel de especialización.
Calcúlase como unha porcentaxe do valor observado respecto do esperado (O/E).

$$PIC = \left(\frac{(O - E)}{E} \right) * 100$$

Indicadores derivados de estrutura e relacións:

- **% de documentos citados:** Porcentaxe de documentos citados a un nivel de detalle determinado sobre o total.
- **% de Documentos non citados:** Porcentaxe de documentos non citados a un nivel de detalle determinado sobre o total.
- **Índice relativo de citación:** O número de citas recibidas entre o número total de citas.
- **Taxa de documentos en coautoría:** Proporción de documentos asinados por máis dun autor calculado sobre o total da produción analizada.
- **Índice de coautoría:** Número medio de autores por documento.

Indicadores derivados de colaboración:

- **Taxa de documentos en colaboración:** Porcentaxe de documentos realizados en colaboración por máis dunha institución calculada sobre o total da produción analizada.



- **Potencial investigador comparado:** Calcúlase para tódolos tipos de colaboración por anos e para un período determinado.
- **Índice de internacionalización:** Aмосa información sobre o maior ou menor grao de participación internacional para o total da produción científica.

$$ii = \frac{ei}{pa} * 100$$

ei = número de enlaces internacionais

pa = produción total por xeografía

- **Índice de coautoría asimétrica:** Permitiría caracteriza-la importancia relativa dos enlaces dun país con respecto a outro.

$$asi = \frac{cop}{co(m - p)} * 100$$



10.- Informes e análise de datos

10.1.- Introducción

O obxectivo final do Sistema de Análise é realizar informes e facer consultas empregando os datos de produción científica almacenados no Sistema.

Tendo en conta o xeito de acceder á informaición e como se elaboran as consultas de datos no Sistema, poderían establecerse dous grupos de informes:

- **Informes predefinidos:** Conxunto de informes cun formato específico que van a ser consultados de xeito habitual.
Estes informes estarían dispoñibles para os usuarios a través do portal de informes *Reporting Services* de SQL Server.
- **Informes específicos:** Serían consultas puntuais ou informes avanzados realizados por usuarios técnicos.
Ademáis, algún informe pode precisar de cálculos complexos sobre os datos ou dunha representación gráfica non soportada por *Reporting Services*. Un exemplo poderían se-los informes de cocitación.
Con este fin é indispensable que no sistema esté dispoñible toda a información que poidera precisarse para dar resposta ás consultas plantexadas.

Ademáis, e segundo sexa a natureza da información consultada, poderíamos falar das seguintes categorías de informes:

10.2.- Informes de cantidade de produción científica

Análise dos aspectos cuantitativos de produción científica, realizados a partires dos indicadores cuantitativos dispoñibles no Sistema.

Algúns dos informes sobre a cantidade da produción científica que sería preciso publicar son:



- Número de documentos por autor
- Número de documentos por revista
- Número de documentos por institución galega
- Número de revistas nas que publican os investigadores/autores galegos
- Número de documentos por ano
- Tipos de documentos por ano, institución ou autor
- Documentos por sectores (administración, sistema sanitario, empresa...)
- Número de documentos nunha área temática (totais e por anos)
- Autores que máis documentos publican nunha área temática
- Institucións que publican nunha área temática, número de documentos e títulos
- Revistas nas que se publican documentos dunha determinada área temática

10.3.- Informes de calidade da produción científica

Este conxunto de informes pretenden medir a calidade da produción científica, empregando para elo indicadores cualitativos como *Factor de impacto* ou o *Potencial investigador*.

Como exemplo de informes de calidade da produción científica publicados no Sistema estarían:

- Revistas con factor de impacto máis alto nunha área temática
- Autores que publican nas revistas de maior factor de impacto
- Potencial Investigador por institución
- Potencial investigador nunha área temática

10.4.- Informes de relacións estruturais

O obxectivo destes informes é amosar unha imaxe da estrutura e relacións existentes no conxunto da produción científica rexistrada no Sistema, establecida a través dos contidos temáticos das publicacións.

Entre os informes desta categoría estarían as representacións multivariadas, que establecen comparacións entre os datos empregando para elo os indicadores de produción e calidade agrupados por área temática e sector, por institución e sector, por xeografía e area temática, etc.



Como exemplos de informes desta categoría estarían:

- Documentos citables e porcentaxe dos mesmos das institucións que publican nunha área temática
- Factor de Impacto Relativo con respecto a Galicia e a España por institución nunha área temática
- Autores por institución
- Autores máis citados nunha área temática
- Citas por revista
- Citas por autor
- Citas por institución
- Documentos máis citados da produción científica galega
- Porcentaxe de documentos citables das institucións galegas
- Táboa de cocitacións por área temática.

Cocitación

Os estudos de cocitación aportan un medio para obter a información relacional da produción científica e realízanse empregando os datos de referencias citadas de cada documento.

Se un documento X cita a dous documentos A e B, enténdese que estes dous documentos son *cocitados* polo documento X.

A modo de exemplo de cocitación inclúese o seguinte gráfico:

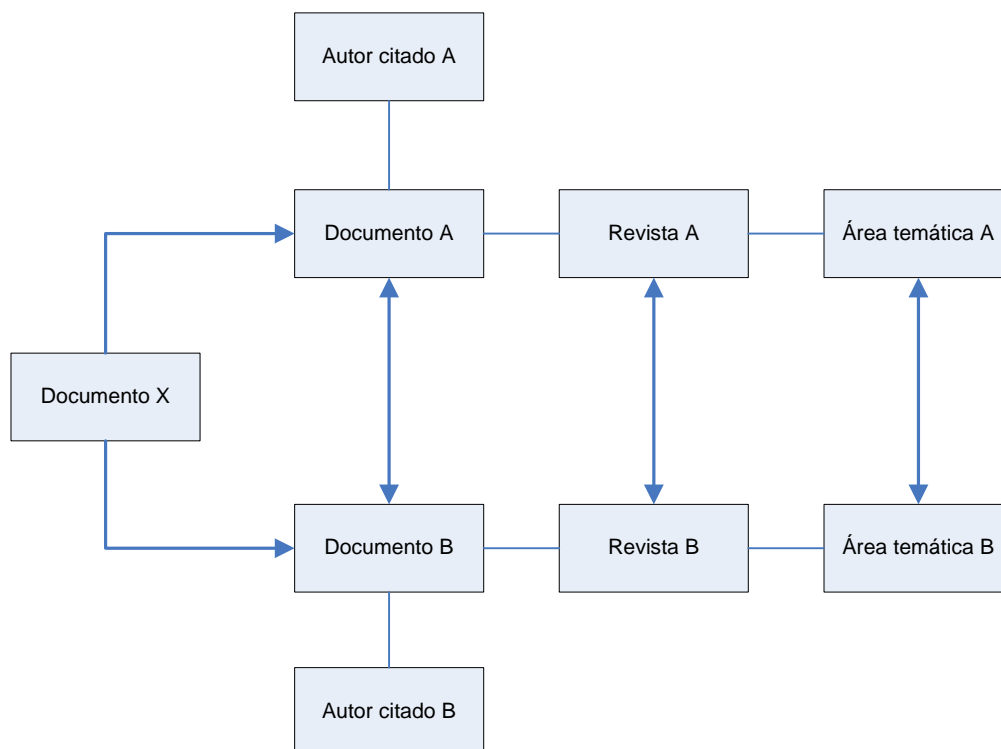


Figura 3: Cocitación entre documentos

Os documentos A e B son cocitados polo documento X. A forza ou intensidade coa que A e B se relacionan depende do número de veces que A e B son cocitados e permite establecer unha medida para representar a estrutura dun dominio.

A través desta relación podería compoñerse unha representación gráfica das relacións existentes entre as diferentes categorías ou áreas temáticas.

En calquera caso á hora de realizar unha representación gráfica da cocitación é importante que o conxunto de datos esté limitado dalgún xeito (ano de publicación, materias, xeografía) pois do contrario a gráfica non aportaría a información requerida.

Como xa se dixo, para elabora-los mapas de cocitación empréganse as referencias citadas. Tando ISI Web of Science como Scopus dispoñen desta información, aínda que a través da API de Scopus non é posible descargar estes datos (ver apartado 12.4.- *Proceso de descarga de datos de Scopus* para máis información sobre este punto).



Un punto a ter en conta que nos datos de ISI Web of Science as referencias citadas só mencionan ao autor principal do documento.

10.5.- Informes de colaboración

Os informes de colaboración tamén reflicten as relacións da produción científica, neste caso en base aos autores dos documentos.

Algúns dos informes de colaboración necesarios son:

- Colaboración por institución (interrexional, nacional, internacional) con porcentaxe de colaboración sobre os totais
- Colaboración interrexional, nacional e internacional nos documentos dunha área temática, xunto coa porcentaxe de colaboración sobre os totais

Colaboración

Considérase que un documento foi elaborado en colaboración cando os seus autores pertencen a máis dunha institución.

Non se considera colaboración cando os autores pertencen ao mesmo grupo de investigación ou a dous departamentos dunha mesma institución.

Cando o traballo foi realizado por autores de institucións dun mesmo país, considérase *colaboración nacional*. Se as institucións son de dous ou máis países diferentes considérase *colaboración internacional*.

Tamén se podería falar de *colaboración interrexional*, cando os autores son de institucións de diferentes Comunidades Autónomas.

10.6.- Relación de informes

A seguinte táboa identifica unha relación de informes a publicar no sistema:



Informes	Fonte		
	WOS	Scopus	IEDCYT
Número de documentos por autor	X	X	X
Número de documentos por revista	X	X	X
Número de documentos por institución galega	X	X	X
Número de revistas nas que publican os investigadores/autores galegos	X	X	X
Número de documentos por ano	X	X	X
Tipos de documentos por ano, institución ou autor	X	X	X
Documentos por sectores (administración, sistema sanitario, empresa...)	X	X	X
Número de documentos nunha área temática (totais e por anos)	X	X	X
Autores que máis documentos publican nunha área temática	X	X	X
Institucións que publican nunha área temática, número de documentos e títulos	X	X	X
Revistas nas que se publican documentos dunha determinada área temática	X	X	X
Documentos citables e porcentaxe dos mesmos das institucións que publican nunha área temática	X	X	
Autores por institución	X	X	X
Autores máis citados nunha área temática	X	X	
Citas por revista	X	X	
Citas por autor	X	X	
Citas por institución	X	X	
Documentos máis citados da produción científica galega	X	X	
Porcentaxe de documentos citables das institucións galegas	X	X	
Taxa de variación interanual	X	X	X
Índice de esforzo temático & relativo	X	X	X
% documentos citados e non citados & índice relativo citación	X	X	
Taxa de documentos en coautoría & índice coautoría	X	X	X
Taxa de documentos en colaboración	X	X	X
Índice de internacionalización	X	X	X
Índice de coautoría asimétrica	X	X	X
Revistas con factor de impacto máis alto nunha área temática	X		
Autores que publican nas revistas de maior fctor de impacto	X		
Potencial investigador por Insitución	X		
Potencial investigador nunha área temática	X		
Factor de impacto relativo con respecto a Galicia e a España por institución ou área temática	X		
Táboa de cocitacións por área temática	X		



Informes	Fonte		
	WOS	Scopus	IEDCYT
Evolución do factor de impacto normalizado	x		
Evolución do factor de impacto medio esperado relativo	x		
Informe de potencial investigador	x		
Informe de pootencial investigador comparado	x		



11.- Proceso de carga

11.1.- Características dos procesos de carga

Os procesos de carga de datos no Sistema de Análise de Producción Científica terán as seguintes características e funcionalidades:

- **Carga inicial de datos:** Inicialmente faráse unha carga de datos que incorporará ao sistema toda a información da que se dispoña no instante da carga.
Esta carga inicial comprenderá tanto a información de documentos publicados como os datos de listas normalizadas e dimensións do sistema: diferentes categorías de materias (ANEP, ISI, UNESCO, Scopus, ISOC), publicacións, institucións, autores, etc.
- **Cargas incrementais:** O sistema permitirá realizar cargas incrementais de datos, de xeito que en cada execución dos procesos de carga se incorpore ao sistema aquela información que é nova ou rectifica a información xa cargada previamente.
Debe se-lo proceso de carga quen determine a necesidade ou non de actualización de datos en cada caso, o mesmo que a estratexia a empregar; por exemplo na xestión das diferentes versións de Journal Citation Reports ou nos datos de institucións.
- **Dexeño modular:** Sería aconsellable establecer un deseño modular dos procesos de carga en base ás diferentes fontes de datos e fases da carga.
- **Parametrización das cargas:** Os procesos de carga deberían dispoñer dos parámetros necesarios para determina-lo seu funcionamento.
Algúns dos parámetros que deben contemplarse son:
 - o Orixe e destino da información (rutas de acceso, servidores, bases de datos, etc...)
 - o Módulos do proceso de carga que van a executarse: fontes de datos e fases ca carga.



- **Xestión de conflitos de datos:** Cando nalgún caso o proceso de carga non sexa quen de identificar axeitadamente un dato (por exemplo unha institución, unha publicación ou un autor), debe xerar unha saída de datos con información detallada do conflito, de xeito que posteriormente ao usuario ou operador de carga lle sexa máis sinxelo decidir qué facer en cada caso co dato que plantexou un conflito. Asimesmo debe buscarse na medida do posible a resolución automática deste tipo de conflitos de datos.
- **Resultados de cada execución:** Información detallada de cada execución da carga, validación da mesma e controis sobre os datos procesados.



12.- Outros procesos e ferramentas

12.1.- Introducción

Debido ás particularidades do Sistema de Análise de Producción Científica, sobor de todo no que atinxe á obtención de datos e aos procesos de carga, é preciso dispoñer de certos procesos e utilidades que, dalgún xeito, quedarían fóra do alcance dun sistema de análise pero que se fan necesarios neste caso.

Estes procesos ou servizos serían:

12.2.- Aplicación de modificación e actualización de datos

O habitual en solucións *DataWarehouse* é que sexan os procesos de carga os que comprobaban a información que se incorpora ao sistema, de xeito que a actualización e modificación de datos sexa transparente para os operadores e usuarios.

No caso do Sistema de Análise de Producción Científica, e debido ás diferentes fontes de datos e non estruturación dos mesmos (por exemplo institucións) e tamén á dificultade para identificar axeitadamente certos valores (por exemplo autores) faise necesario a intervención de usuarios que, de xeito manual, depuren e actualicen a información.

Por exemplo:

- Cando no sistema hai dous autores co mesmo nome, pero son distintas persoas.
- Cando existen dous autores con diferente nome pero son a mesma persoa.
- Para asignar o sector ao que pertence unha institución.
- Para manter actualizada e correxida a lista normalizada de institucións.

Para elo faise preciso unha aplicación, preferiblemente con acceso vía web e coa posibilidade de traballaren dous ou máis usuarios de xeito concurrente, que permita actualiza-los datos de Institucións e Autores: creación, modificación e borrado.

Sería de gran utilidade que fora o proceso de carga quen nunha primeira instancia identifique aqueles autores ou institucións que presenta un "conflicto", de xeito que sexan presentados



ao usuario para que estas discrepancias poidan resolverse a través da aplicación de actualización de datos.

Esta aplicación debería estar sustentada sobre un modelo de datos independente do modelo de datos do Sistema de Análise.

Co obxectivo de favorecer un maior acerto á hora de identifica-los autores, este modelo de datos debería relacionar os autores dun documento coas materias e institucións relacionadas con dito documento: Deste xeito sería máis sinxelo para o usuario saber cando dous nomes diferentes son a mesma persoa ou un mesmo nome fai referencia a persoas distintas.

12.3.- Proceso de descarga de datos de ISI Web of Science

Como se reseñou en apartados anteriores, Thomson Scientific, o proveedor de ISI Web of Knowledge e os seus produtos, dispón dunha API (*Application Programming Interface*) para, a través de *Web Services*, descarga-los datos rexistrados na súa base de datos de Web of Science.

Dado que a descarga de datos a través da páxina web de ISI sería un traballo tedioso e pouco operativo plantéxase a necesidade de codificar un proceso que, facendo uso da API, acceda á base de datos de Web of Science e obteña os datos de publicacións científicas.

Este proceso necesitaría parámetros de "Data" e "País" de xeito que se restrinxise o volumen de información a descargar.

Os campos polos que se poden facer búsquedas no Web of Science a través da API e, por tanto, poderían formar parte dos parámetros do proceso son:

Código	Descrición
TS	Palabra clave
TI	Título
AU	Autor
GP	Grupo de autores
SO	Fonte
AD	Dirección postal
OG	Organización



Código	Descrición
SG	Suborganización
SA	Rúa da dirección
CI	Cidade
PS	Provincia/Estado
CU	País
ZP	Código postal
UT	ut-Código identificador de documentos en ISI
DT	Tipo de documento
LA	Idioma
PY	Ano de publicación

En principio a través da API pode accederse a tódolos campos do Web of Science.

12.4.- Proceso de descarga de datos de Scopus

Scopus tamén dispón dunha API para obte-los datos de publicacións científicas da súa base de datos, de xeito que non haxa que facelo a través da súa páxina web.

Sería preciso, á hora de lanza-lo proceso que obtén os datos de Scopus, definir parámetros para "Data" e "Pais", de xeito que poidera acotarse o volumen de información a descargar.

Para usuarios non rexistrados en Scopus a documentación desta API é máis ben escasa, polo que sería preciso confirmar que se pode obter toda a información necesaria.

Segundo esta documentación, os campos que se poden consultar a través da API son:

Campo	Descrición
Article title	Título do artigo.
Document type	Tipo de documento.
Cited-by count	Número de citas.
Inward URL	URL.
Source title	Título da fonte.
ISSN	ISSN.
Vol	Volume.
Issue	Número.
Page	Páxina.
Publication date	Data de publicación.
EID	Identificador electrónico. (<i>Electronic Identifier</i>). Tódolos documentos de Scopus teñen un EID.
SCP	Identificador Scopus. (<i>Scopus ID</i>). Identificador único. Tódolos documentos Scopus teñen un SCP.



Campo	Descrición
DOI	Digital Object Identifier. Cadena alfanumérica para identificar a propiedade intelectual nun entorno electrónico.
First author	Autor principal.
Complete author list	Lista completa de autores.
Affiliations	Institucións relacionadas co documento.
Abstracts	Resumos.

Os campos son devoltos pola API como unha cadea separada por comas.

Na documentación da API recollida na web de Scopus non está detallada a construción da cadea de búsqueda que permitiría filtra-la información da base de datos de Scopus.

Nesta lista de campos non aparecen nin as referencias citadas nin as palabras clave. As referencias citadas son imprescindibles si se queren facer estudos de cocitación empregando os datos de Scopus.

12.5.- Proceso de obtención do DOI (Digital Object Identifier)

O *Digital Object Identifier* (DOI) permite identificar contidos no entorno dixital.

Os nomes DOI asígnanse a entidades para usar en redes dixitais. Son usados para proporcionar información actualizada, incluído dónde pode atoparse na Internet.

A información sobre un obxecto dixital pode cambiar co tempo, incluso donde atopala, pero o nome DOI non cambia.

Se un documento dispón de DOI, este debe estar rexistrado no Sistema de Análise. Únicamente e nalgúns casos os documentos procedentes do ISI Web of Science conteñen o DOI asociado.

CrossRef, unha asociación independente fundada e dirixida por editores, posibilita a obtención do DOI a través de Web Services. O funcionamento é o seguinte:

- Debe enviarse un arquivo XML co formato axeitado e que inclúa un conxunto de documentos para os que se precise obter o DOI.



Por cada documento poden indicarse ata 10 campos, sendo obrigatorio un identificador de publicación (o ISSN e/ou título da publicación) e un autor ou primeira páxina que axude a identifica-lo artigo. Os demais campos son opcionais, pero é recomendable referilos para axudar a identificar correctamente o documento e obte-lo DOI asociado.

Os posibles campos a incluír na consulta son:

Campo	Descrición	Observacións
ISN	ISSN	ISSN de impresión e/ou electrónico.
TTL	Título da publicación	Título completo da publicación ou unha abreviatura (ou unha combinación dos dous)
NAM	Autor principal	Apelido do autor principal
VID	Id de volume	Número ou texto que especifica o volume (p.e. Outono, Q1)
IID	Id de número	Número ou texto que identifica o número da publicación.
PID	Páxina	Número ou texto que identifica a páxina. Tratarase coma un número (os zeros á esquerda non se terán en conta). Se hai caracteres diante do número (p.e. "pp") tamén se omitirán.
YNO	Ano	Ano en número. Os caracteres causarán un error. Dous díxitos interpretaranse como 19XX.
TYP	Tipo	full_text, abstract_only ou bibliographic_record
KEY	Clave única	Un identificador único que será enviado de volta cos resultados da consulta. É útil para identificar cada documento cos seus correspondentes resultados.
DOI	DOI	O DOI (sempre en branco na consulta).

- Como resposta, obtense outro arquivo XML co mesmo número de documentos para os que se enviou a solicitude indicando en cada caso si se resolveu ou non e os valores de cada campo.

Se se resolveu a obtención do DOI, indícase para cada campo cómo foi o resultado da búsqueda, se exacto ou aproximado.

Exemplo de identificación de documento no arquivo XML a enviar:

```
<query key="30043" enable-multiple-hits="false" forward-match="false"
xmlns="http://www.crossref.org/qschema/2.0">
```



```
<author>Fernandez</author>
<journal_title>Epicardial implantation of biventricular leads - Response</journal_title>
<issn>03008932</issn>
<year>2004</year>
<volume>57</volume>
<issue>12</issue>
<first_page>1254</first_page>
</query>
```

Exemplo de datos de documento no arquivo XML recibido:

```
<query key="43" status="resolved">
  <doi>10.1157/13069879</doi>
  <issn match="exact" type="print">03008932</issn>
  <issn match="exact" type="electronic">15792242</issn>
  <journal_title match="optional">Revista Española de Cardiología</journal_title>
  <author match="fuzzy">Fernández</author>
  <volume match="exact">57</volume>
  <issue match="exact">12</issue>
  <first_page match="exact">1254</first_page>
  <year match="exact">2004</year>
  <publication_type>full_text</publication_type>
</query>
```

Dentro do sistema debe habilitarse un proceso que permita seleccionar unha serie de documentos e xenerar o arquivo XML de solicitude de DOI.

Asimesmo, tamén hai que desenvolver outro proceso que, a partires do XML recibido, identifique aqueles documentos para os que se resolveu o DOI e actualice co seu correspondente valor o Sistema de Análise.



13.- Plataforma tecnolóxica

O Sistema de Análise de Producción Científica estará baseado na plataforma SQL Server 2005 Enterprise Edition e nas ferramentas que proporciona.

- **Xestor relacional:** Modelo de datos do Sistema de Análise e modelo de datos que permita xestionar-lo mantemento de autores e institucións.
- **Integration Services:** Codificación de procesos de carga.
- **Reporting Services:** Codificación e publicación de informes no portal web.

Para a codificación de procesos como a xestión de autores e institucións ou calquera outra tarefa que non sexa xestionada por SQL Server Integration Services empregárase preferiblemente tecnoloxía Microsoft .NET.



14.- Conclusións

Neste apartado recóllense as conclusións ás que se chegou despois do traballo realizado e o plantexamento que se considera máis axeitado á hora de implanta-lo Sistemad e Análise.

14.1.- Consideracións sobre o Sistema de Análise

Proceso de carga

Un apartado a ter en conta na carga de datos é a identificación correcta de parte da información que se procesa, especialmente no que atinxe aos autores e ás institucións.

No caso dos autores non existe un código que os identifique de xeito unívoco e únicamente se dispón do nome e apelido ou apelidos.

Por elo, faise imprescindible a intervención de usuarios do Sistema que determinen a identidade de aqueles autores que teñan o mesmo nome e sexan persoas diferentes ou ben cando se trate do mesmo autor que aparece rexistrado con nomes distintos.

No caso das institucións, tanto o emprego de abreviaturas como o xeito en que aparecen rexistrados os departamentos e diferentes institucións que colaboran na autoría dun documento dificultan a identificación axeitada de cada unha delas.

Modifica-los autores e institucións directamente sobre o repositorio de datos de produción científica é desaconsellable, tanto polo rendemento (gran volume de datos) como pola complexidade (moitas táboas de datos e relacións).

Unha alternativa é dispoñer dun modelo de datos propio para rexistrar e manter esta información. Os usuarios poderían, a través dun formulario web, modifica-los datos oportunos e, posteriormente, lanzar unha actualización sobre o Sistema de Análise.

Tamén, e de cara a facilita-la tarefa dos usuarios, o sistema debería axudar a identifica-los autores amosando as relacións que ten cada un deles coas institucións e as materias nas que ten produción científica.



Doutra banda, a obtención de datos de ISI Web of Science e Scopus debería facerse de xeito automatizado a través das ferramentas das que dispoñen as dúas compañías.

Tendo en conta que a documentación de ambas ferramentas non é o completa que sería desexable e que ademáis, sobre todo de cara á carga inicial de datos, o volumen de información a tratar é elevado, sería aconsellable verifica-la viabilidade e o rendemento dos procesos de obtención de datos de Scopus e ISI Web of Science.

Como xa quedou dito, os procesos de carga deben xestionar recargas de datos que inclúan modificacións e novos datos e a obtención e actualización do DOI (*Digital Object Identifier*).

Modelo de datos

O modelo de datos debe identifica-la procedencia da información, de xeito que poidan establecerse comparacións entre os indicadores en función da fonte de datos.

Empregar listas de valores comúns a tódalas fontes de datos, como institucións ou áreas temáticas, ampliaría as capacidades de análise permitindo establecer comparativas máis axustadas entre os resultados obtidos para cada fonte.

O deseño do modelo de datos debe busca-lo compromiso entre a complexidade das relacións existentes entre as entidades do Sistema (autores – institucións – documentos) e a facilidade de uso e bo rendemento que debe ofrecer á hora de obter informes. Non hai que esquecer que o obxectivo final do Sistema é realizar consultas de datos.

No momento de deseña-lo Sistema deben terse en conta as necesidades de información actuais e as que poideran presentarse no futuro, de xeito que o modelo dé resposta ás consultas que poideran plantexarse.

Informes e análise de datos

O obxectivo é buscar que a maioría das consultas que se precisen estén publicadas no portal de informes, de xeito que os usuarios accedan aos datos do Sistema sen necesidade de recurrir aos técnicos.

14.2.- Plantexamento do Sistema de Análise

14.3.- Configuración

Para o Sistema de Análise, propónse a seguinte configuración:

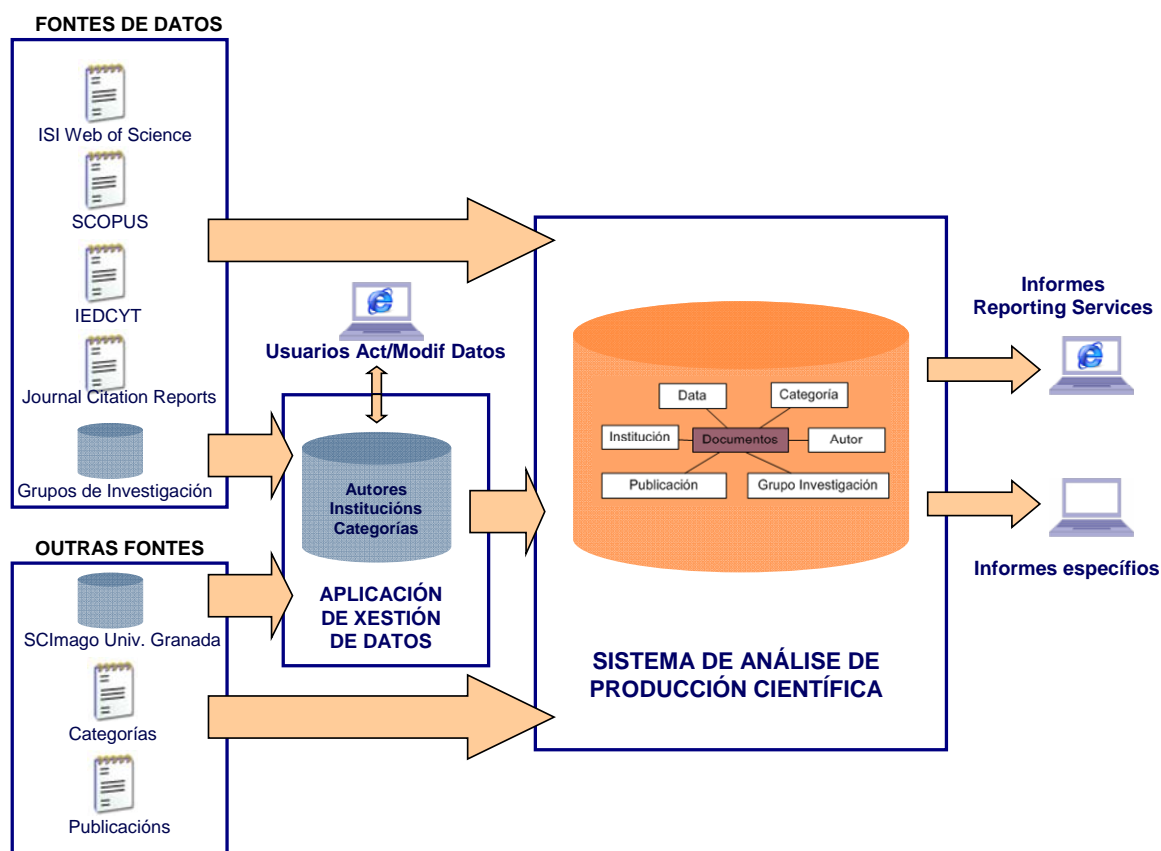


Figura 4: Configuración proposta para o Sistema de Análise

Para o Sistema de Análise propiamente dito recoméndase o emprego de técnicas e deseños propios dos sistemas datawarehouse, orientados á consulta de datos e a obtención de informes.

Considérase ademais necesario o desenvolvemento dunha aplicación para a xestión de autores e institucións, así como doutras listas de datos necesarias para a realización de informes como poden se-las categorías temáticas.



14.4.- Desenvolvemento

Debido á súa complexidade, non se considera aconsellable desenvolve-lo Sistema de Análise como un único proxecto. Pola contra, o plantexamento do Sistema en diferentes fases permitirá un maior control e redundará na consecución dos obxectivos plantexados.

Así, un posible plantexamento sería:

Fase 1

- **Proxecto 1:** Desenvolver a aplicación que xestiona as listas de autores, institucións e categorías, así como outras listas que sexan necesarias.
- **Proxecto 2:** Incorporar ao Sistema de Análise a información de Scopus, Grupos de Investigación e outros datos que sexan necesarios.
Tamén se deseñarían e publicarían os informes necesarios para consulta-los datos cargados.

Fase 2

- Incorporar ao Sistema de Análise os datos de ISI Web of Science, Journal Citation Reports e IEDCYT, así como completa-lo conxunto de informes do Sistema.



ANEXO. Sitios de interés na Internet:

A continuación recóllense algúns sitios na rede con información sobre o contido deste documento e información sobre a produción científica.

THOMSON SCIENTIFIC – ISI Web of Knowledge

- Páxina principal de Thomson Scientific:

<http://scientific.thomson.com/>

- Web of Science:

<http://scientific.thomson.com/products/wos/>

- Journal Citation Reports:

<http://scientific.thomson.com/products/jcr/>

- Essential Science indicators:

<http://scientific.thomson.com/support/recorded-training/esi/>

- Búsqueda de publicacións:

<http://www.isinet.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=D>

-Glosario de termos:

<http://scientific.thomsonreuters.com/support/patents/patinf/terms/>

- Web services para descargar información. Preguntas frecuentes:

<http://scientific.thomsonreuters.com/support/faq/webservices/>

SCOPUS

- Páxina principal de Scopus

<http://www.scopus.com/scopus/home.url>

- Documentación da API de Scopus

http://searchapidocs.scopus.com/scapi_vers_1.htm#JavaScript%20Example

IEDCYT

- Páxina principal de IEDCYT (antes CINDOC)

<http://www.cindoc.csic.es/>

- Bases de datos do CSIC

<http://www.cindoc.csic.es/servicios/dbinfo.htm>



VARIOS

- Clasificación de materias UNESCO (página da Universidade do País Vasco)

<http://www.et.bs.ehu.es/varios/unesco.htm>

- DOI (Digital Object Identifier)

<http://www.doi.org/>

- Páxina principal de Cross Ref.

<http://www.crossref.org/>

-CSA Illumina

<http://www.csa.com/>