

NES provides quality metadata services to world-class database publishers and institutes of higher education. We deliver unmatched value through a combination of process excellence, quality control and services such as information & content development, metadata modeling, controlled vocabulary services, data preparation, software development, product quality control & assurance, specialized email and technical phone support apart from assisting with product development.

NES is renowned for its mastery in the de-duplication of bibliographic and metadata records through the creation of composite records.

www.nes.co.in

NESTM

Core **ompetencies**

Make the best use of NES' expertise in:

METADATA

DEVELOPMENT & MANAGEMENT

CONVERSION & DE-DUPLICATION

OF BIBLIOGRAPHIC DATA

PRODUCT QUALITY

ASSURANCE & CONTROL

CONTENT DEVELOPMENT

THROUGH WEB RESEARCH

XML/HTML

ENCODING

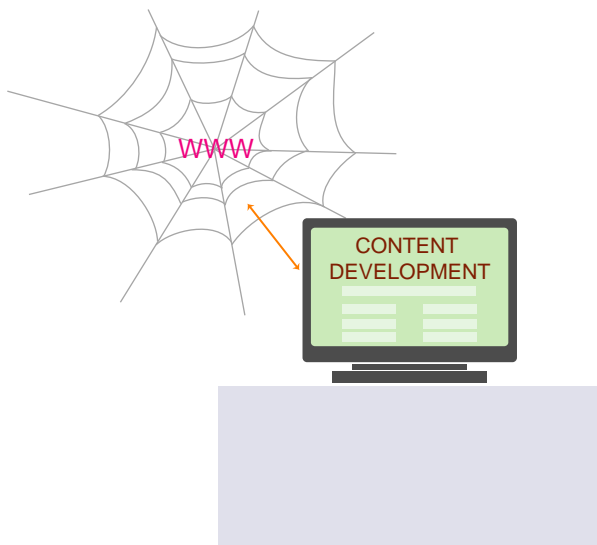
TECHNICAL

SUPPORT

CONTENT DEVELOPMENT

THROUGH WEB RESEARCH

NES



- We help customers publish e-journals on schedule.
- NES has experienced staff that can perform web-research and use resources on the web to construct accurate Author/Institution/Company profiles as desired by the customers.
- We identify and gather information on freely available electronic journals and keep the table of contents up-to-date on publishers' websites.
- We produce bibliographic records for secondary database publishers.
- We assign subject headings to database records.

PRODUCT QUALITY

ASSURANCE & CONTROL

You searched: Anquetin: sports femmes sterbeck

Author	Sterbeck, John
Title	Des sports et des femmes / John Sterbeck ; trad. de l'anglais par Maurice-Edgar Condreau ; dossier réalisé par Régis Wiener
Place / dates	(Paris) : Gallimard, 2008
Material	114 p.
Genre	Ficte. ; SF. Classification X06a siècle
Note	Traduit de: English and more
ISBN / Price	227030647X
FRISQING	Bulle Bibliothèque publique et scolaire (Mise à grutier)
REBO no	RD4326226
Other author	Condreau, Maurice-Edgar
Other author	Wiener, Régis
Series	Folio, Gallimard. Plus



- We execute several Quality Checks on the customers' publications.
- We are experienced in performing Quality Assurance tests on databases/publications using specifications, support files and other utilities specially developed.
- With nearly ten years of experience in the field, we have developed scores of program utilities to quality check and generate reports on erroneous or misplaced content in databases and publications.

Beale, C.L. and K.M. Johnson, 1998. "The Identification of Recreational Counties in Nonmetropolitan Areas of the USA," *Population Research and Policy Review* 17(1), 37-53.



Authors= Beale, C.L. | Johnson, K.M.
 Title= The Identification of Recreational Counties in Nonmetropolitan Areas of the USA
 JR= Population Research and Policy Review
 Volume= 17
 Issue= 1
 Start Page= 37
 End Page= 53
 Page Count= 17
 Publication Year= 1998
 Publication Type= JOURNAL ARTICLE

Cho, K., Yung, K. & Chang, S. (1982). *Preservation of Cultivated Mushrooms, Biological Nature and Cultivation Methods in the Tropical Mushrooms*. pp. 63-86. Hong Kong: The Chinese University Press.



Authors= Cho, K. | Yung, K. | Chang, S.
 Book Title= Preservation of Cultivated Mushrooms, Biological Nature and Cultivation Methods in the Tropical Mushrooms
 Place= Hong Kong
 Publisher= The Chinese University Press
 Start Page= 63
 End Page= 86
 Page Count= 24
 Publication Year= 1982
 Publication Type= BOOK

- Parse, proof-read, validate and create database records from article references.
- Quality Check & correct distorted OCR text.
- Substitute UTF-8 character set if improperly handled by the OCR software.
- Standardize text with look up tables wherever required.

DATA PREPARATION

duplicate
 duplicate
 duplicate
 duplicate
 duplicate
 duplicate
 duplicate
 duplicate
 duplicate

- At the data preparation stage, we convert any format of data into a native format, run utilities to clean up the data, de-dupe and load the data into any other desired format for delivery.
- NES specializes in the de-duplication of database records. Hundreds of utilities are employed to clean up, organize and standardize data.
- Formats handled include in-house proprietary formats, MARC, Dublin Core, XML of any schema, MS Access, SQL or any format that is in a field de-limited database format.

3

Standardize your metadata with look-ups and de-duplication.

Types of Problems in Bibliographic Databases

Databases filled with bibliographic records accompanied by links to fulltext provide an un-paralleled service to students, scientists and researchers. These databases, however, often are plagued by errors—both common and rare—that hamper effective search and retrieval of relevant items or accurate linking. It is imperative to have clean and consistently presented data. From a retrieval standpoint, an erroneous or wrongly entered record is a lost entity. Enriched with years of expertise, NES is equipped to tackle these problems through:

1. Bibliographic database analysis; and
2. Rectification of errors and normalization of variations

A database can arrive in any proprietary or other format such as MARC, CDF (comma de-limited format), TDF (tab de-limited format), TAGGED, DUBLIN Core, MS Access, HTML, and XML. NES reviews the data and prepares a specification for cleanup. Problems in databases usually are two-fold:

- a) Format errors
- b) Data errors

a) Format errors

One of the major errors in MARC format is the improper/wrong positioning of the data elements, which results in disarray of fields. CDF & TDF could miss some fields or position irrelevant data in fields or skip field de-limiters. Tagged format data at times misses the record de-limiters and could storm one field into another. Commonly, all these formats could have empty/null records. With decades of experience, NES is equipped to handle such situations. NES' utilities report the instances mentioned above and suggest the corrective measures to be taken.

b) Data errors

These types of errors constitute a major portion that affects the actual content of the bibliographic records. Some of the errors are listed below:

i) *Missing fields*

Essential fields go missing resulting in instances such as

- Journal article records without journal title/title/volume/issue information
- Book records without title/publisher
- Records without an individual/institutional author
- Records without collation/imprint
- Dissertations without the institution names
- Conference records without conference details

ii) *Wrong representation/misplaced data in various fields*

- Just alphabets in year field/Year data beginning with anything other than 1 or 2
- Numerals only in the author field
- ISBN/ISSN with lesser/more characters than permitted
- Records without title
- Repetition of fields such as author repeated in title or vice-versa
- Repetition of terms in index terms/keywords

iii) *Messed up fields*

- Instances of large strings of stars, question marks, slash, etc.
- Too many space/null characters within the data
- Field/Record de-limiters occurring within the running text of an abstract or full-text

iv) *Bad records*

- Records that are less than 50K in size could have missing fields/data
- Records with phrases such as Test record, check & testing, etc.

v) *Sundry checks performed*

- Check all entities, foreign-language character sets & non-ASCII characters used in the dataset.
- Examine if large text data fields (abstracts, notes, full-text, etc.) containing tables, paragraphs, indented text are properly formatted.
- Evaluate journal titles/publisher names/places for consistency and construct look-up tables to standardize them
- Run utilities to report non-functioning URLs in a database

vi) *To analyze and rectify XML files, NES can*

- Assist to design, create and modify XSD schema
- Provide solutions to counter discrepancies in the XML data
- Validate XML files using different schemas

De-duping bibliographic records

A common and un-avoidable feature in database files is the occurrence of duplicate records. These records look exactly similar with or without slight variations.

Why do they occur?

- Unknown duplication done at the same division due to accidental copying of records in the input tool.
- Improper bibliographic control—when two divisions of the same organization are distributed in different geographic locations.
- When two or more bibliographic databases are integrated

NES' COMPARE & COMPOSITE TECHNOLOGY

is a dynamic semi-automated process that unifies SIMILAR bibliographic records irrespective of the source of arrival. A multitude of rules have been laid down for the software to compare bibliographic elements pertaining to various publication types.

At times, minor variations exist between records from different databases because of typing or other errors. NES goes an extra mile in handling these subtle variations while de-duping them. Some examples follow:

Author Variations

Hayes, Lynn O'Roarke
Hayes, Lynn O'Rourke

Multhauf, R P
Multhaupt, Robert P.

Lang, Cheng-Te
Lang, Chang-Te

Pettegrew, Loyd S.
Pettegrew, Loyde S.

Title Variations

1989 annual forecast: colleges and universities
1989 Annual Forecast: Colleges & Universities

Australian tourism training: now and in the future
Australian Tourism Training--Now And In Future

Greenbrier's New Chef To Take The 'Light' Road
Greenbrier's New Chef To The Take 'Light' Road

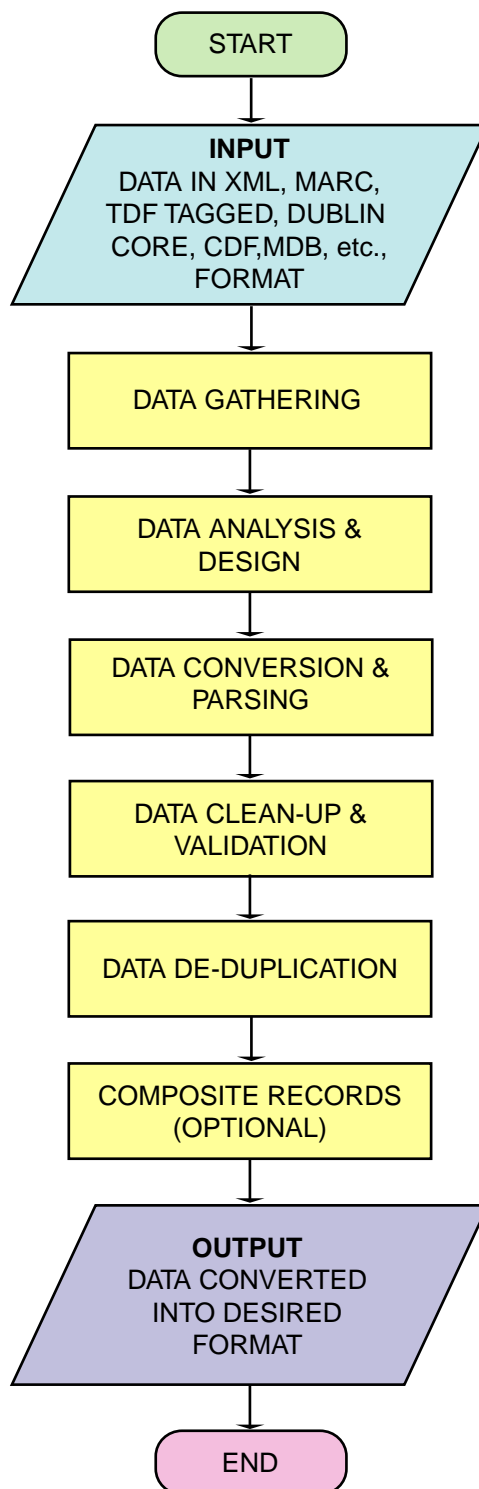
Anesthesia mishap: who's in charge.
Anesthesia mishaps: Who's in charge?

We also take care **NOT** to unify records that are apparent duplicates as follows. The titles are same but the **Authors are different**.

Hokfelt, T. / Tatemoto, K. **Galanin - 25 years with a multitaled neuropeptide.**
Cell Mol Life Sci. 65(12). p. 1791 - 1795

Mitsukawa, K. / Lu, X. / Bartfai, T. **Galanin - 25years with a multitaled neuropeptide.**
Cell Mol Life Sci. 65(12). p. 1796-1805

DATA PREPARATION FLOWCHART

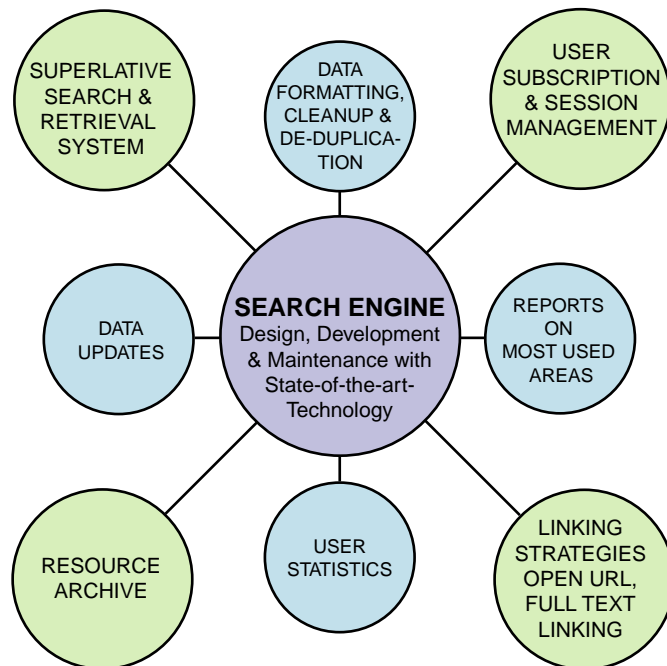


SEARCH ENGINE

NES

DEVELOPMENT

- NES has extensive experience in search engine design for international organizations.
- NES offers extensive and secure web hosting and mirror site services on state-of-the-art hardware.
- NES is best equipped to offer search and retrieval services.
- Several of our customers' databases are updated and hosted on NES' servers providing un-interrupted quality service. Customers often ask that we maintain their software and add regular improvements as requested.



TECHNICAL

SUPPORT



- Our Technical Support team has vast hands-on experience in handling customer queries from across the globe over phone and e-mail regarding database publishing services. International clients call or write to NES and we ensure that they always have a pleasant experience receiving our support services.
- Questions resolved pertain to various technical and non-technical issues ranging from interface access, customization, functionality, usage, content-related enquiries and many more....

NES: NISC EXPORT SERVICES PRIVATE LIMITED

Plot 95/1, Phase 2
IDA Cherlapally 500 051, HYDERABAD
Andhra Pradesh, India

Internet www.nes.co.in Email nes@nes.co.in
Phone +91 (040) 3911 4200, +91 (040) 2726 0801
Fax +91 (040) 3911 4219, +91 (040) 2726 0825

NES — your one-stop for database publishing services, world-class always