Package: orisma (via r-universe)

May 19, 2026

Type Package

Title Occupational Risk Integrated Systematic Mapping and Analysis

Version 0.1.0

Description A complete pipeline for systematic bibliometric mapping of occupational health and safety (OHS) evidence. Starting from reference files exported from major bibliographic databases such as Web of Science, Scopus, PubMed, Dimensions, EBSCO, and others, 'orisma' automates ingestion, deduplication, relevance filtering, occupational risk category extraction, bibliometric analysis, and report generation. The package is related to bibliometric science mapping and evidence synthesis workflows described by Aria and Cuccurullo (2017) <doi:10.1016/j.joi.2017.08.007>, Westgate (2019) <doi:10.1002/jrsm.1374>, and Lajeunesse (2016) <doi:10.1111/2041-210X.12472>, but adds a domain-specific occupational safety and health layer. The package implements three original bibliometric indicators: (1) the Worker-Risk Disconnection Index (WRDI), measuring the proportion of studies that characterise an occupational risk without including direct worker exposure data; (2) the Risk Category Saturation Index (RCS), measuring the relative over- or under-representation of each risk category relative to a uniform baseline; and (3) the Material-Gap Profile (MGP), measuring the ratio between a material's known hazard potential and its coverage in the occupational health literature. Two additional preventive intelligence indicators are provided: (4) the Abstract Sufficiency Score (ASS, 0-5), a cumulative hierarchical index of the preventively useful information contained in an abstract; and (5) the Bridge Article Score (0-5), identifying studies that simultaneously address technology, hazardous agent, worker population, exposure measurement, and preventive recommendations. Risk categories are extracted using a built-in occupational risk dictionary of 58 categories anchored in ISO 45001:2018, INSST, NIOSH, and EU-OSHA frameworks, organised in six blocks: Safety, Industrial Hygiene, Ergonomics,

Psychosociology, Biological Hazards, and Emerging Technologies. The dictionary is user-extensible. Outputs include bilingual HTML reports, occupational risk sheets, priority reading rankings, guided extraction matrices for systematic review, and reproducibility certificates with MD5 hashes.

License MIT + file LICENSE

Encoding UTF-8

Language en-GB

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

URL https://github.com/Aguilar-Elena/orisma

BugReports https://github.com/Aguilar-Elena/orisma/issues

Depends R (>= 4.1.0)

Imports cli (>= 3.4.0), digest (>= 0.6.29), dplyr (>= 1.1.0), ggplot2 (>= 3.4.0), ggrepel (>= 0.9.0), glue (>= 1.6.0), jsonlite (>= 1.8.0), magrittr (>= 2.0.0), pheatmap (>= 1.0.12), readr (>= 2.1.0), stringdist (>= 0.9.8), stringr (>= 1.5.0), synthesisr (>= 0.3.0), tidyr (>= 1.3.0), tools, utils

Suggests knitr, rmarkdown, rsvg, testthat (>= 3.0.0)

VignetteBuilder knitr

LazyData true

Config/pak/sysreqs libicu-dev libx11-dev

Repository https://aguilar-elena.r-universe.dev

Date/Publication 2026-05-07 15:16:56 UTC

RemoteUrl https://github.com/aguilar-elena/orisma

RemoteRef HEAD

RemoteSha 2aaabef7310dab7d895f33cf3465ec30720f2238

Contents

`orisma-package` *orisma: Occupational Risk Integrated Systematic Mapping and Analysis*

Description

`orisma` is an R package for systematic bibliometric mapping of occupational risk evidence. It is designed to help researchers and occupational safety and health practitioners analyse whether the scientific literature on a given topic is connected to workers, workplaces, exposure conditions and preventive decision-making.

Details

ORISMA provides a complete workflow for occupational risk evidence mapping:

- multi-source bibliographic ingestion;
- deduplication;
- relevance filtering through `orm_relevance_guard()` and `orm_run_guarded()`;
- risk category extraction using a 58-category occupational risk dictionary;

- preventive bibliometric indicators such as WRDI, RCS and MGP;
- Abstract Sufficiency Score (ASS);
- Bridge Article Score;
- bilingual academic reports and practitioner-oriented risk sheets.

Typical workflow:

```
library(orisma)

refs <- orm_load("my_references/")

result <- orm_run_guarded(
  refs,
  topic = "Collaborative robotics and occupational health and safety",
  mode = "conservative"
)

orm_report(result)
orm_risk_sheet(result)
```

Original indicators

- WRDI: Worker-Risk Disconnection Index.
- RCS: Risk Category Saturation Index.
- MGP: Material-Gap Profile.
- ASS: Abstract Sufficiency Score.
- Bridge Article Score.

Citation

Aguilar-Elena, R., & Delgado-Garcia, A. (2026). `orisma`: Occupational Risk Integrated Systematic Mapping and Analysis. R package version 0.1.0. Universidad Internacional de Valencia (VIU) & Universidad de Salamanca (USAL). https://github.com/Aguilar-Elena/orisma

Author(s)

Raul Aguilar-Elena raguilar@universidadviu.com
Occupational Risk Prevention and Occupational Health Research Group (GPRL), Universidad Internacional de Valencia (VIU), Valencia, Spain.

Ana Delgado-Garcia a.delgado@usal.es
Universidad de Salamanca (USAL), Salamanca, Spain.

Author(s)

Maintainer: Raúl Aguilar Elena <raguilar@universidadviu.com> (ORCID)

Authors:

- Ana Delgado-Garcia <a.delgado@usal.es>

See Also

Useful links:

- https://github.com/Aguilar-Elena/orisma
- Report bugs at https://github.com/Aguilar-Elena/orisma/issues

---

`orisma_dict` *Built-in risk dictionaries for ORISMA*

Description

ORISMA ships with a comprehensive normative risk dictionary anchored in four internationally recognised taxonomies:

- **INSST** Instituto Nacional de Seguridad y Salud en el Trabajo (Spain)
- **ISO 45001:2018** International OHS management system standard
- **NIOSH** National Institute for Occupational Safety and Health (USA)
- **EU-OSHA** European Agency for Safety and Health at Work

The dictionary covers 58 risk categories organised in 6 blocks: A) Safety at work (18), B) Industrial hygiene (8), C) Ergonomics (8), D) Psychosociology (11), E) Biological hazards (5), F) Emerging technologies (8).

Examples

```
# View available dictionaries
orm_dict_list()

# Load the default dictionary
dict <- orm_dict()

# View all categories
orm_dict_categories(dict)

# Add custom terms to a category
dict <- orm_dict_add_terms(dict, "nanomaterials",
                           c("metal powder", "powder bed"))
```

`orisma_indicators` *ORISMA bibliometric indicators*

Description

ORISMA implements five original bibliometric indicators designed specifically for occupational health and safety (OHS) evidence mapping. Three are corpus-level indicators (WRDI, RCS, MGP) and two are record-level indicators (ASS, Bridge Score).

1. Worker-Risk Disconnection Index (WRDI):

Definition:
The WRDI measures the proportion of studies in a corpus that characterise an occupational risk **without** including direct worker exposure data. A study is considered to have worker exposure data if its abstract contains terms indicating real measurement of exposure in actual workers under real working conditions (e.g. "worker exposure", "occupational exposure", "breathing zone", "personal sampling", "field study", "workplace measurement").

Formula:
For a given risk category c:

$$WRDI_c = 1 - \frac{N_{workers,c}}{N_{total,c}}$$

where $N_{workers,c}$ is the number of studies in category c that include worker exposure data, and $N_{total,c}$ is the total number of studies in that category.
The **global WRDI** is computed across all records:

$$WRDI_{global} = 1 - \frac{N_{workers}}{N_{total}}$$

Interpretation:

- **WRDI = 0**: All studies include direct worker exposure data. The body of evidence is fully connected to real workplace conditions.
- **WRDI = 1**: No study includes worker exposure data. The entire literature characterises the risk technically (e.g. in simulated environments, chambers, or in vitro) without measuring real exposure in workers.
- **WRDI >= 0.8**: Critical alert. The evidence has very low direct preventive transferability. On-site risk assessment is essential.
- **WRDI 0.5-0.8**: Attention required. More than half the evidence lacks worker data.
- **WRDI < 0.3**: Reasonable coverage. Most studies include worker data.

Important limitation:
WRDI detection is based on abstract text, not full text. Studies that measured worker exposure but did not mention it in the abstract may be misclassified. Manual validation via `orm_validate()` is recommended.

2. Risk Category Saturation Index (RCS):

Definition:
The RCS measures the **relative dominance** of a risk category in the corpus compared to a hypothetical uniform distribution across all categories. It identifies which categories are over-represented (saturated) and which are under-represented (gaps) in the literature.

Formula:

$$RCS_c = \frac{pct_c}{pct_{uniform}}$$

where pct_c is the percentage of records assigned to category c, and $pct_{uniform} = 100/K$ is the percentage each category would have under a uniform distribution across all K categories.
Equivalently:

$$RCS_c = \frac{N_c \cdot K}{N_{total}}$$

where N_c is the number of records in category c, K is the total number of categories, and N_{total} is the total number of records.

Interpretation:

- **RCS > 1**: The category is over-represented relative to a uniform baseline. The literature has concentrated disproportionately on this risk type.
- **RCS = 1**: The category has exactly the representation expected under a uniform distribution.
- **RCS < 1**: The category is under-represented. This risk type has received less attention than a balanced literature would suggest.
- **RCS = 0**: No studies address this category. Complete evidence gap.

Note:
RCS is a relative measure. A category can have RCS > 1 with very few absolute studies if the corpus is small or highly specialised. Always interpret RCS together with the absolute number of records (N).

3. Material-Gap Profile (MGP):

Definition:
The MGP is a domain-specific indicator designed for corpora where the corpus can be stratified by material, substance, or agent. It measures the ratio between a material's known hazard potential and its coverage in the occupational health literature, identifying materials that are dangerous but understudied.

Formula:

$$MGP_m = \frac{hazard\_proxy_m}{coverage_m}$$

where $hazard\_proxy_m$ is an estimate of the material's hazard potential (based on the number of distinct risk categories detected in studies involving that material), and $coverage_m$ is the proportion of corpus records that address that material.

Interpretation:

- **High MGP**: The material is associated with multiple risk categories but appears in few studies. Priority material for future research and on-site risk assessment.
- **Low MGP**: The material is well-covered in the literature relative to its known hazard profile.
- **MGP requires a material column**: The `material_col` parameter in `orm_analyse()` must point to a column classifying each record by material or agent. If not available, MGP is not computed.

4. Abstract Sufficiency Score (ASS):

Definition:
The ASS is a **cumulative hierarchical index** (0-5) measuring how much preventively useful information an abstract contains for an occupational health practitioner. It is not a measure of study quality, but of abstract informativeness for preventive purposes.
The score is strictly cumulative: a record cannot reach level N without satisfying all previous levels.

Levels:

0 - Non-informative The abstract contains no hazard or risk terms relevant to OHS. No useful preventive information.

1 - Hazard without context The abstract mentions a hazard or risk agent (e.g. nanoparticles, noise, vibration) but provides no occupational or workplace context. Could be an environmental or laboratory study.

2 - Occupational context The abstract mentions workers, employees, operators, or workplace/occupational setting. The study is clearly situated in a work context.

3 - Exposure measurement The abstract reports quantitative exposure data: concentrations, levels, measurements, or monitoring results. Implies some form of exposure quantification.

4 - Worker exposure with result The abstract explicitly reports exposure in workers (not just in the environment) with a result (e.g. exceeded a limit, found significant association, detected at breathing zone).

5 - Complete preventive abstract The abstract addresses all four dimensions: worker population + exposure measurement + study method/design + preventive recommendation or control measure. This is the highest OHS informative level.

Computation:
Each level is detected via regular expression patterns applied to the abstract text. Detection is strictly cumulative: the algorithm tests each level in sequence and stops at the first level not satisfied.

Interpretation:

- **Mean ASS < 2**: The corpus is predominantly technical with very little preventive context. High priority for on-site investigation.
- **Mean ASS 2-3**: Mixed corpus. Some workplace context but limited quantitative exposure data.
- **Mean ASS > 3**: Good preventive evidence base. Substantial proportion of studies report actual worker exposure data.
- **ASS = 5 articles**: These are the most valuable abstracts for practitioners and should be read in full first.

5. Bridge Article Score:

Definition:
A bridge article is a study that **connects technical science with applied OHS prevention**. It simultaneously addresses five dimensions that are rarely all present in a single study:

Criterion 1 - Technology/process The study involves a specific technology, industrial process, or work task (e.g. additive manufacturing, welding, construction, healthcare).

Criterion 2 - Hazardous agent The study characterises a specific hazardous agent (chemical, physical, biological, or psychosocial).

Criterion 3 - Workers (MANDATORY) The study involves a real worker population in a real workplace setting. This criterion is mandatory for bridge classification.

Criterion 4 - Exposure measurement (MANDATORY) The study quantitatively measures exposure (air sampling, biological monitoring, dosimetry, etc.). This criterion is mandatory for bridge classification.

Criterion 5 - Prevention/recommendation The study includes preventive recommendations, control measures, or intervention results.

Classification:

Strong bridge (score 4-5) Meets criteria 3+4 (mandatory) plus 2 or 3 additional criteria. Highest priority for full-text reading. These articles have already done the translation from laboratory science to workplace prevention.

Partial bridge (score 3) Meets criteria 3+4 (mandatory) plus 1 additional criterion. Valuable but incomplete bridge.

Technical study (score 0-2, or missing C3/C4) Does not meet the mandatory criteria. Contributes technical knowledge but lacks direct preventive applicability.

Priority score:
The overall **priority reading score** used in `orm_ranking()` combines all record-level indicators:

$$Priority = (Bridge \times 2) + (ASS \times 1.5) + (N_{cats} \times 0.5)$$

where N_{cats} is the number of risk categories detected in the record. Bridge score is weighted highest because it reflects the most direct preventive relevance.

References:

The WRDI, RCS, and MGP indicators were first described in:

Aguilar-Elena, R. & Delgado-Garcia, A. (2025). *Mapping the Safety Landscape of Emerging Technologies: A Bibliometric Analysis of Occupational Risks in Metal Additive Manufacturing*. (Under review)

The ORISMA methodological framework is described in:

Aguilar-Elena, R. & Delgado-Garcia, A. (2025). *orisma: A Framework for Occupational Risk Integrated Systematic Mapping and Analysis*. R package version 0.1.0. Universidad Internacional de Valencia (VIU) & Universidad de Salamanca (USAL).

See Also

`orm_analyse()` to compute WRDI, RCS, and MGP. `orm_ass()` to compute the Abstract Sufficiency Score. `orm_bridge()` to detect bridge articles. `orm_ranking()` to generate a priority reading list. `orm_validate()` to validate automatic classification with Cohen's Kappa.

`orisma_sample` *Sample bibliographic records for ORISMA*

Description

Twenty bibliographic records on occupational health in metal additive manufacturing (2015-2026) from Web of Science and Scopus, pre-processed with ORISMA to illustrate the full pipeline.

Usage

```
data(orisma_sample)
```

Format

A data frame with 20 rows and 9 variables:

record_id Character. Unique record identifier.

title Character. Article title.

abstract Character. Abstract (max 800 characters).

year Integer. Publication year.

doi Character. Digital Object Identifier.

source_db Character. Source database.

bridge_type Character. Bridge classification.

bridge_score Integer. Bridge score (0-5).

ass_score Integer. Abstract Sufficiency Score (0-5).

Source

Web of Science and Scopus (2015-2026).

`orm_analyse` *Compute ORISMA bibliometric indicators and analyses*

Description

`orm_analyse()` takes an extraction matrix and computes:

- **WRDI** - Worker-Risk Disconnection Index: the proportion of studies that characterise a risk without measuring direct worker exposure. A WRDI of 1 means all studies are purely technical (no worker data); 0 means all studies include direct worker exposure measurement.
- **RCS** - Risk Category Saturation Index: relative dominance of each risk category compared to a uniform-distribution baseline. RCS > 1 means the category is over-represented; RCS < 1 means it is under-represented.
- **MGP** - Material-Gap Profile: ratio of a material's known hazard potential (from the literature consensus) to its proportional coverage in the corpus. Detects hazardous materials that are academically under-studied.

It also computes co-occurrence matrices, temporal trends, and author networks for visualisation.

Usage

```
orm_analyse(
  mx,
  material_col = NULL,
  year_col = "year",
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

mx	An orisma_matrix object returned by orm_extract().
material_col	Character. Name of the column containing material information. If NULL (default), MGP is skipped with a warning.
year_col	Character. Column name for publication year. Default "year".
lang	Character. "en" or "es".
verbose	Logical. Print progress?

Value

A list (class orisma_result) with all indicators and analysis objects ready for orm_report() and visualisation functions.

Examples

```
## Not run:
refs    <- orm_load("my_references/")
deduped <- orm_dedup(refs)
mx      <- orm_extract(deduped)
result  <- orm_analyse(mx)

# View the three core indicators
result$indicators

# View WRDI
result$WRDI

## End(Not run)
```

---

orm_ass *Abstract Sufficiency Score (ASS)*

Description

`orm_ass()` computes an **Abstract Sufficiency Score** (0-5) for each record, measuring how much preventively useful information the abstract contains for an occupational health practitioner.

The score is **cumulative and hierarchical** - a record cannot reach level N without satisfying all previous levels:

- **0** Non-informative abstract for OHS purposes
- **1** Mentions a hazard or risk, but no occupational context
- **2** Mentions occupational/workplace context
- **3** Mentions exposure measurement or quantification
- **4** Mentions exposure in workers with some result
- **5** Mentions exposure, worker population, method AND control/prevention

Usage

```
orm_ass(
  mx,
  text_col = "abstract",
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`mx`	An `orisma_matrix` object from `orm_extract()`.
`text_col`	Character. Text field to score. Default `"abstract"`, falls back to `"title"` if abstract is mostly empty.
`lang`	Character. `"en"` or `"es"`.
`verbose`	Logical.

Value

The `orisma_matrix` object with added columns: `ass_score` (0-5), `ass_label` (descriptive label), `ass_level_reached` (highest level passed).

`orm_ass_plot` *Plot ASS distribution*

Description

Generates a bar chart showing the distribution of Abstract Sufficiency Scores across the corpus.

Usage

```
orm_ass_plot(mx, out_dir = NULL, lang = getOption("orisma.lang", "en"))
```

Arguments

mx	An orisma_matrix object after running orm_ass().
out_dir	Character or NULL. Directory to save the plot.
lang	Character. "en" or "es".

Value

A ggplot2 object invisibly.

---

orm_autodim *Automatic dimension extraction and risk cross-matrix*

Description

orm_autodim() automatically discovers the most relevant contextual dimensions of a corpus using two complementary modes:

Mode 1: Dictionary blocks (default, method = "blocks") Uses the normative blocks of the ORISMA dictionary (A-Safety, B-Hygiene, C-Ergonomics, D-Psychosociology, E-Biological, F-Emerging) as dimensions. Computes a block x block co-occurrence matrix showing how many studies address combinations of risk blocks simultaneously. Works for any corpus without any configuration.

Mode 2: Free text (method = "text") Extracts discriminant terms from abstracts using TF-IDF-like filtering. Useful for discovering domain-specific dimensions not covered by the dictionary (e.g. specific materials, sectors, tasks).

Usage

```
orm_autodim(
  mx,
  method = "blocks",
  text_col = "abstract",
  n_dims = 12L,
  min_freq = 3L,
  max_doc_pct = 0.35,
  min_cooccur = 0.5,
  fuzzy_sim = 0.85,
  stopwords = NULL,
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

mx	An orisma_matrix object from orm_extract().
method	Character. "blocks" (default) or "text".
text_col	Character. Text field for method = "text". Default "abstract".

n_dims	Integer. Max dimensions for method = "text". Default 12.
min_freq	Integer. Min document frequency for method = "text". Default 3.
max_doc_pct	Numeric (0-1). Max document proportion for method = "text". Terms above this are too generic. Default 0.35.
min_cooccur	Numeric (0-1). Min co-occurrence with a risk. Default 0.5.
fuzzy_sim	Numeric (0-1). Fuzzy grouping threshold. Default 0.85.
stopwords	Character vector. Extra stopwords for method = "text".
lang	Character. "en" or "es".
verbose	Logical.

Value

A list (class orisma_dims) ready for orm_dim_matrix().

See Also

orm_dim_matrix()

---

orm_bridge *Bridge Article Detection and Priority Ranking*

Description

orm_bridge() identifies **bridge articles** - studies that connect technical science with real occupational prevention. These are the highest-value articles for an occupational health practitioner because they have already done the translation from laboratory to workplace.

A bridge article simultaneously mentions:

1. **Technology/process** (what was studied)
2. **Hazardous agent** (what risk was characterised)
3. **Workers** (real people in real workplaces)
4. **Exposure measurement** (quantitative data)
5. **Prevention/recommendation** (actionable output)

Articles meeting 4 or 5 criteria are classified as **strong bridges**. Articles meeting 3 criteria (must include workers + measurement) are **partial bridges**. Others are technical studies.

Usage

```
orm_bridge(
  mx,
  text_col = "abstract",
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

mx	An orisma_matrix object, ideally after running orm_ass().
text_col	Character. Text field to analyse. Default "abstract".
lang	Character. "en" or "es".
verbose	Logical.

Value

The orisma_matrix object with added columns: bridge_score (0-5), bridge_type (Strong/Partial/Technical), bridge_criteria (which criteria were met).

---

orm_dedup *Automatic deduplication of bibliographic records*

Description

orm_dedup() removes duplicate records using a **three-step progressive pipeline**:

1. **Exact DOI match** — most reliable signal; decisive for records with DOIs.
2. **Normalised title match** — removes punctuation, accents, case, and extra spaces before comparing; catches the same article listed with minor typographic differences across databases.
3. **Fuzzy match** — compares title + year + first author using Optimal String Alignment distance; catches near-identical records that escape exact matching (e.g. different journal abbreviations, truncated author lists).

Only records that remain ambiguous after all three steps are flagged for optional manual review. These are saved to dedup_log.csv.

Usage

```
orm_dedup(
  refs,
  fuzzy_threshold = 0.9,
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE),
  save_log = TRUE
)
```

Arguments

refs	An orisma_refs object returned by orm_load().
fuzzy_threshold	Numeric (0–1). Similarity threshold for fuzzy matching. Default 0.90 (90% similarity = duplicate). Increase for stricter matching, decrease for more aggressive deduplication.
lang	Character. "en" or "es". Overrides orisma.lang option.
verbose	Logical. Print progress? Default TRUE.
save_log	Logical. Save dedup_log.csv to working directory? Default TRUE.

Value

An `orisma_refs` tibble with duplicates removed. Attributes record deduplication statistics for inclusion in the PRISMA log.

Examples

```
## Not run:
refs    <- orm_load("my_references/")
deduped <- orm_dedup(refs)

# More aggressive fuzzy matching
deduped <- orm_dedup(refs, fuzzy_threshold = 0.85)

# Spanish messages, no log file
deduped <- orm_dedup(refs, lang = "es", save_log = FALSE)

## End(Not run)
```

`orm_dict` *Load a risk dictionary*

Description

Load a risk dictionary

Usage

```
orm_dict(name = "iso45001_insst")
```

Arguments

`name`	Character. Dictionary name. Default `"iso45001_insst"`.

Value

A named list (class `orisma_dict`).

`orm_dict_add_category` *Add a new risk category to a dictionary*

Description

Add a new risk category to a dictionary

Usage

```
orm_dict_add_category(
  dict,
  key,
  label_en,
  label_es,
  terms,
  worker_exposure_terms = character(0),
  taxonomy = "user",
  block = "G - Custom"
)
```

Arguments

`dict`	An `orisma_dict` object.
`key`	Character. Short identifier (no spaces).
`label_en`	Character. Category name in English.
`label_es`	Character. Category name in Spanish.
`terms`	Character vector. Search terms.
`worker_exposure_terms`	Character vector. Worker exposure indicators.
`taxonomy`	Character. Source taxonomy label.
`block`	Character. Block label (e.g. "A - Safety").

Value

Updated `orisma_dict`.

orm_dict_add_terms *Add terms to an existing dictionary category*

Description

Add terms to an existing dictionary category

Usage

```
orm_dict_add_terms(dict, category, terms)
```

Arguments

dict	An orisma_dict object.
category	Character. Category key.
terms	Character vector. New terms to add.

Value

Updated orisma_dict.

orm_dict_categories *List risk categories in a dictionary*

Description

List risk categories in a dictionary

Usage

```
orm_dict_categories(dict, lang = getOption("orisma.lang", "en"))
```

Arguments

dict	An orisma_dict object.
lang	Character. "en" or "es".

orm_dict_list *List available built-in dictionaries*

Description

List available built-in dictionaries

Usage

```
orm_dict_list()
```

Details

This function takes no arguments.

orm_dim_matrix *Build a risk category x dimension cross-matrix*

Description

Builds a risk category x dimension cross-matrix and saves a hierarchical clustered heatmap with dendrograms and numeric values in each cell.

When `dims` was built with `method = "blocks"`, the matrix shows risk categories x normative blocks (A-Safety, B-Hygiene, etc.). When `dims` was built with `method = "text"`, the matrix shows risk categories x discovered text dimensions.

Usage

```
orm_dim_matrix(
  result,
  dims,
  min_records = 2L,
  out_dir = NULL,
  filename = "risk_dimension_heatmap.png",
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`result`	An `orisma_result` object from `orm_analyse()` or `orm_run()`.
`dims`	An `orisma_dims` object from `orm_autodim()`.
`min_records`	Integer. Min records for a risk category row. Default `2`.
`out_dir`	Character or NULL. Directory to save the heatmap PNG.
`filename`	Character. Output filename. Default `"risk_dimension_heatmap.png"`.
`lang`	Character. `"en"` or `"es"`.
`verbose`	Logical.

Value

Invisibly returns the cross-matrix (risk categories x dimensions).

`orm_extract` *Extract risk categories from bibliographic records*

Description

`orm_extract()` scans the **title**, **abstract**, and **keywords** of each record against the active risk dictionary and builds a **binary presence matrix** (record x risk category). It also detects whether each study contains direct worker exposure data - the key signal for computing the **WRDI** indicator.

Matching is case-insensitive and uses whole-word boundary detection to avoid false positives (e.g. "laser" does not match "eyelaser").

Usage

```
orm_extract(
  refs,
  dict = orm_dict(),
  fields = c("title", "abstract", "keywords"),
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`refs`	An `orisma_refs` object (output of `orm_load()` or `orm_dedup()`).
`dict`	An `orisma_dict` object. Default: `orm_dict()` (ISO 45001 / INSST / NIOSH).
`fields`	Character vector. Which text fields to search. Default `c("title", "abstract", "keywords")`.
`lang`	Character. `"en"` or `"es"`.
`verbose`	Logical. Print progress?

Value

A list (class `orisma_matrix`) containing:

`refs` Original `orisma_refs` tibble with added columns: one binary column per risk category (`cat_*`), `n_categories` (total categories matched), and `has_worker_data` (logical).

`matrix` Pure binary matrix (records x categories) for downstream analysis.

`dict` The dictionary used.

`categories` Category metadata tibble.

Examples

```
## Not run:
refs   <- orm_load("my_references/")
deduped <- orm_dedup(refs)

# Use default dictionary
mx <- orm_extract(deduped)

# Use a customised dictionary
dict <- orm_dict()
dict <- orm_dict_add_terms(dict, "nanoparticles", c("nano-dust", "UFP"))
mx   <- orm_extract(deduped, dict = dict)

# Restrict to title + abstract only
mx <- orm_extract(deduped, fields = c("title", "abstract"))

## End(Not run)
```

`orm_extraction_matrix` *Generate a guided extraction matrix for manual review*

Description

`orm_extraction_matrix()` generates a structured extraction template pre-filled with automatically extracted information. The practitioner completes the remaining fields using the full PDF.

Articles are selected and ranked by combined bridge score + ASS score. The matrix contains auto-filled bibliographic data, ORISMA scores, detected technology and risk categories, and empty fields for manual completion with full-text PDFs.

Usage

```
orm_extraction_matrix(
  mx,
  result,
  top_n = 30L,
  min_bridge_score = 2L,
  out_dir = "orisma_output",
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`mx`	An `orisma_matrix` object after `orm_bridge()` and `orm_ass()`.
`result`	An `orisma_result` object from `orm_run()`.
`top_n`	Integer. Max articles to include. Default `30`.

`min_bridge_score`
: Integer. Min bridge score. Default 2.

`out_dir`
: Character. Output directory.

`lang`
: Character. `"en"` or `"es"`.

`verbose`
: Logical.

Value

Invisibly returns the path to the saved CSV.

`orm_load` *Load bibliographic references from one or multiple files / folders*

Description

`orm_load()` is the entry point of every ORISMA analysis. It reads bibliographic files in **RIS**, **BibTeX**, or **CSV** format from a folder (or a vector of individual file paths), detects the format of each file automatically, combines all records into a single tidy data frame, and records the source database for each record.

All major bibliographic databases export to at least one supported format:

Database	Recommended format	Notes
Web of Science	RIS / Plain text	Max 1 000 records per batch
Scopus	RIS or CSV	Max 2 000 records per batch
PubMed	RIS	No limit
Dimensions	CSV or RIS	Max 2 500 per batch
EBSCO (CINAHL, BSC)	RIS	Up to 25 000
ProQuest	RIS or BibTeX	Max 100 per batch
Cochrane Library	RIS	No limit
Ovid / MEDLINE	RIS	Max 1 000 per batch
ScienceDirect	RIS	No limit
The Lens (free)	RIS or CSV	No limit

Usage

```
orm_load(
  path,
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

path	Character. Path to a **folder** containing reference files, or a **character vector** of individual file paths.
lang	Character. Language for console messages: "en" (default) or "es". Overrides getOption("orisma.lang").
verbose	Logical. Print progress messages? Default TRUE.

Value

A tibble (class orisma_refs) with standardised columns:

record_id Internal unique identifier assigned by ORISMA

source_file Name of the original file

source_db Database inferred from file name or format

title Article title

authors Authors (semicolon-separated)

year Publication year

doi Digital Object Identifier (if available)

abstract Abstract text

keywords Author keywords

journal Journal name

volume, issue, pages Bibliographic location

document_type Article, review, conference paper, etc.

Examples

```
## Not run:
# Load all .ris and .bib files from a folder
refs <- orm_load("my_references/")

# Load specific files
refs <- orm_load(c("wos_results.ris", "scopus_results.csv"))

# Spanish messages
refs <- orm_load("mis_referencias/", lang = "es")

## End(Not run)
```

orm_normativa *Cross-reference detected risks with applicable European regulation*

Description

`orm_normativa()` crosses the risk categories detected by ORISMA with the main applicable European directives and ISO standards, providing the occupational health practitioner with a direct regulatory anchor for each identified risk.

The regulatory database is built into ORISMA and covers EU directives, Spanish INSST technical notes (NTP), and key ISO standards. It is updated with each major package release.

Usage

```
orm_normativa(result, min_records = 1L, lang = getOption("orisma.lang", "en"))
```

Arguments

`result`	An `orisma_result` object.
`min_records`	Integer. Min records for a category to be included. Default `1`.
`lang`	Character. `"en"` or `"es"`.

Value

A data frame with detected categories and their applicable regulations.

orm_priority *Compute risk priority scores and traffic light classification*

Description

`orm_priority()` assigns a priority level to each detected risk category using three criteria combined into a single priority score:

- **Frequency** (RCS): how saturated is this category in the literature
- **Disconnection** (WRDI): how far is the research from real worker data
- **Evidence volume**: number of records

Categories whose RCS exceeds `context_rcs_threshold` are flagged as **context categories** (the dominant topic of the corpus, not a risk per se) and are reported separately rather than mixed with risk categories.

Priority levels for non-context categories:

- **RED**: WRDI >= wrdi_high AND RCS >= 1. Over-studied technically but no worker data. Urgent preventive gap.
- **AMBER**: Moderate evidence OR partial worker data.
- **GREEN**: WRDI < wrdi_low. Good worker data connection.
- **GREY**: n_records < min_records. Insufficient evidence.

Usage

```
orm_priority(
  result,
  min_records = 2L,
  wrdi_high = 0.7,
  wrdi_low = 0.3,
  context_rcs_threshold = 15,
  lang = getOption("orisma.lang", "en")
)
```

Arguments

result	An orisma_result object.
min_records	Integer. Min records for evaluation. Default 2.
wrdi_high	Numeric. WRDI threshold for high disconnection. Default 0.7.
wrdi_low	Numeric. WRDI threshold for low disconnection. Default 0.3.
context_rcs_threshold	Numeric. RCS above which a category is considered a context category (dominant topic) rather than a risk. Default 15.
lang	Character. "en" or "es".

Value

A list with two data frames: $risks (priority-classified risk categories) and $context (dominant topic categories).

---

orm_ranking *Generate priority reading ranking*

Description

orm_ranking() produces a **priority reading list** for occupational health practitioners, ranking articles by their combined relevance score (bridge score + ASS score + number of risk categories detected).

Articles at the top of the list are those most likely to contain actionable preventive information and should be read first in full.

Usage

```
orm_ranking(
  mx,
  top_n = 20L,
  out_dir = NULL,
  lang = getOption("orisma.lang", "en")
)
```

Arguments

mx	An orisma_matrix object after running orm_bridge() and optionally orm_ass().
top_n	Integer. Number of top articles to return. Default 20.
out_dir	Character or NULL. Directory to save the ranking CSV.
lang	Character. "en" or "es".

Value

A data frame with the top_n priority articles.

---

orm_relevance_guard *Relevance guard for occupational risk evidence mapping*

Description

Adds a relevance-control layer before ORISMA analysis. The function identifies whether each record is relevant to the target topic, whether it contains an occupational context, whether it is likely to be biomedical or clinical noise, and whether it should be excluded from the main occupational analysis.

Usage

```
orm_relevance_guard(
  data,
  topic = NULL,
  topic_regex = NULL,
  occupational_regex = NULL,
  noise_regex = NULL,
  title_col = NULL,
  abstract_col = NULL,
  keywords_col = NULL,
  mode = c("conservative", "flag", "strict")
)
```

Arguments

data	A data frame of bibliographic records.
topic	Optional topic label used to derive a topic-specific regular expression.
topic_regex	Optional regular expression defining the target technology/topic.
occupational_regex	Optional regular expression defining occupational relevance.
noise_regex	Optional regular expression defining likely off-topic biomedical/clinical noise.
title_col	Optional title column name. If NULL, it is detected automatically.
abstract_col	Optional abstract column name. If NULL, it is detected automatically.

keywords_col	Optional keywords column name. If NULL, it is detected automatically.
mode	Relevance filtering mode. "flag" excludes only records outside the target topic and marks uncertain records for review. "conservative" excludes off-topic and likely non-occupational biomedical/clinical records. "strict" also excludes records with weak occupational context.

Value

The input data frame with additional relevance-control columns.

`orm_report` *Generate all ORISMA outputs and reports*

Description

`orm_report()` takes a completed `orisma_result` object and generates the full set of outputs including improved visualisations and a rich bilingual HTML executive report.

Usage

```
orm_report(
  result,
  topic = NULL,
  lang = getOption("orisma.lang", "en"),
  out_dir = getOption("orisma.out_dir", "orisma_output"),
  formats = c("html", "csv", "plots", "certificate"),
  min_records = 1L,
  top_n = 8L,
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

result	An `orisma_result` object from `orm_analyse()` or `orm_run()`.
topic	Character. Domain or technology being analysed. Used in plot subtitles and report headers. If NULL, neutral generic text is used.
lang	Character. "en" or "es". Report language.
out_dir	Character. Output directory. Created if it does not exist.
formats	Character vector. Which outputs to generate. Options: "html", "csv", "plots", "certificate". Default: all.
min_records	Integer. Minimum records for a category to appear in plots. Default 1.
top_n	Integer. Number of top categories to show in temporal plot. Default 8.
verbose	Logical. Print progress?

Value

Invisibly returns the output directory path.

`orm_risk_sheet` *Generate an occupational risk sheet*

Description

`orm_risk_sheet()` generates a structured, actionable risk sheet for occupational health practitioners. It synthesises ORISMA outputs into a single HTML document that can be used as supporting evidence in a workplace risk assessment.

The sheet is **regulation-neutral**: it does not include country-specific regulations or limit values, as these vary by jurisdiction. The practitioner applies the relevant national/regional regulation based on the risk categories identified.

Content:

- Context analysis (dominant topic of the corpus)
- Priority traffic light (RED/AMBER/GREEN/GREY) per risk category
- Evidence summary with confidence level
- Knowledge gap alerts
- WRDI interpretation with confidence score
- Methodological section (bases, deduplication, WRDI definition, limits)

Usage

```
orm_risk_sheet(
  result,
  topic = "Occupational risk analysis",
  search_strategy = NULL,
  inclusion_criteria = NULL,
  out_dir = "orisma_output",
  lang = getOption("orisma.lang", "en"),
  min_records = 1L,
  context_rcs_threshold = 15,
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`result`	An `orisma_result` object.
`topic`	Character. Technology or domain being assessed.
`search_strategy`	Character or NULL. Description of the search strategy used (databases, keywords, date range). If NULL, a placeholder is used.
`inclusion_criteria`	Character or NULL. Description of inclusion/ exclusion criteria applied. If NULL, ORISMA defaults are described.

out_dir	Character. Output directory.
lang	Character. "en" or "es".
min_records	Integer. Min records for a category to appear. Default 1.
context_rcs_threshold	Numeric. RCS threshold for context detection. Default 15.
verbose	Logical.

Value

Invisibly returns the path to the generated HTML risk sheet.

---

orm_run *Run the complete ORISMA pipeline in one call*

Description

orm_run() is the **single-function entry point** for a complete ORISMA analysis. It runs all pipeline steps automatically:

1. Deduplication (3-step: DOI + title + fuzzy)
2. Risk category extraction (dictionary-based)
3. Bibliometric analysis (WRDI, RCS, MGP indicators)
4. Automatic dimension detection (normative blocks)
5. Abstract Sufficiency Score (ASS, 0-5)
6. Bridge article detection and priority ranking

Minimal usage (3 lines):

```
library(orisma)
refs   <- orm_load("my_references/")
result <- orm_run(refs)
orm_report(result, lang = "es")
```

All intermediate objects are stored in the result for downstream use with orm_report(), orm_risk_sheet(), orm_ranking(), and orm_extraction_matrix().

Usage

```
orm_run(
  refs,
  dict = orm_dict(),
  topic = NULL,
  autodim_method = "blocks",
  material_col = NULL,
  year_col = "year",
  fuzzy_threshold = 0.9,
```

```
  fields = c("title", "abstract", "keywords"),
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE),
  save_report = FALSE,
  out_dir = getOption("orisma.out_dir", "orisma_output")
)
```

Arguments

refs	An orisma_refs object from orm_load().
dict	An orisma_dict object. Default: orm_dict().
topic	Character. Domain or technology being analysed (e.g. 'Noise in construction', 'Metal AM'). Used in plot subtitles and report headers. If NULL, neutral generic text is used.
autodim_method	Character. "blocks" (default) or "text".
material_col	Character or NULL. Column for MGP. Default NULL.
year_col	Character. Year column. Default "year".
fuzzy_threshold	Numeric. Deduplication threshold. Default 0.90.
fields	Character vector. Text fields for extraction. Default c("title", "abstract", "keywords").
lang	Character. "en" or "es".
verbose	Logical. Default TRUE.
save_report	Logical. Auto-call orm_report()? Default FALSE.
out_dir	Character. Output directory if save_report = TRUE.

Value

An orisma_result object containing all indicators, analyses, dimensions (result$dims), extraction matrix (result$mx), ASS scores and bridge classification (in resultmxrefs), and priority ranking (result$ranking).

---

orm_run_guarded *Run ORISMA with a relevance-control layer*

Description

Runs ORISMA after applying orm_relevance_guard(). This is useful for real-world bibliographic searches where broad database queries may retrieve technically related but non-occupational or off-topic records.

Usage

```
orm_run_guarded(
  refs,
  topic = NULL,
  exclude_non_relevant = TRUE,
  min_records = 50,
  topic_regex = NULL,
  occupational_regex = NULL,
  noise_regex = NULL,
  mode = c("conservative", "flag", "strict"),
  ...
)
```

Arguments

refs	A data frame of references, usually produced by orm_load().
topic	Topic label passed to orm_relevance_guard() and orm_run().
exclude_non_relevant	Logical. If TRUE, records flagged as non-relevant are excluded before running the main ORISMA pipeline.
min_records	Minimum number of records required after filtering. If the filter leaves fewer records, the function stops to avoid accidental over-filtering.
topic_regex	Optional topic regex.
occupational_regex	Optional occupational relevance regex.
noise_regex	Optional noise regex.
mode	Relevance filtering mode. "flag" excludes only records outside the target topic and marks uncertain records for review. "conservative" excludes off-topic and likely non-occupational biomedical/clinical records. "strict" also excludes records with weak occupational context.
...	Additional arguments passed to orm_run().

Value

An ORISMA result object with an added relevance_guard component.

---

orm_validate *Manual validation assistant with Cohen's Kappa*

Description

`orm_validate()` supports methodological validation of ORISMA's automatic risk extraction by presenting a random sample of classified records for manual review. It then computes **Cohen's Kappa** to measure agreement between automatic and manual classification.

This addresses a key peer-review concern: distinguishing between "category detected by dictionary" and "risk actually evaluated in study".

The function saves a CSV file pre-filled with automatic classifications that the researcher edits manually, then re-loads for Kappa computation.

Usage

```
orm_validate(
  mx,
  n_sample = 30L,
  out_dir = "orisma_validation",
  validation_file = NULL,
  seed = 42L,
  lang = getOption("orisma.lang", "en"),
  verbose = getOption("orisma.verbose", TRUE)
)
```

Arguments

`mx`	An `orisma_matrix` object from `orm_extract()`.
`n_sample`	Integer. Number of records to sample. Default `30`.
`out_dir`	Character. Directory to save validation files.
`validation_file`	Character or NULL. Path to a completed validation CSV (output of a previous `orm_validate()` call) for Kappa computation. If NULL, creates the file for manual review.
`seed`	Integer. Random seed for reproducibility. Default `42`.
`lang`	Character. `"en"` or `"es"`.
`verbose`	Logical.

Value

If `validation_file` is NULL: invisibly returns the path to the validation CSV. If `validation_file` is provided: returns a data frame with Kappa statistics per category.

print.orisma_dict *Print method for orisma_dict*

Description

Print method for orisma_dict

Usage

```
## S3 method for class 'orisma_dict'
print(x, ...)
```

Arguments

x	An orisma_dict object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_dims *Print method for orisma_dims*

Description

Print method for orisma_dims

Usage

```
## S3 method for class 'orisma_dims'
print(x, ...)
```

Arguments

x	An orisma_dims object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_kappa *Print method for orisma_kappa*

Description

Print method for orisma_kappa

Usage

```
## S3 method for class 'orisma_kappa'
print(x, ...)
```

Arguments

x	An orisma_kappa object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_matrix *Print method for orisma_matrix*

Description

Print method for orisma_matrix

Usage

```
## S3 method for class 'orisma_matrix'
print(x, ...)
```

Arguments

x	An object to print.
...	Further arguments passed to or from other methods.

Value

Invisibly returns the input orisma_matrix object. Called primarily for its console-printing side effect.

print.orisma_normativa *Print method for orisma_normativa*

Description

Print method for orisma_normativa

Usage

```
## S3 method for class 'orisma_normativa'
print(x, ...)
```

Arguments

x	An orisma_normativa object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_priority *Print method for orisma_priority*

Description

Print method for orisma_priority

Usage

```
## S3 method for class 'orisma_priority'
print(x, ...)
```

Arguments

x	An orisma_priority object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_ranking *Print method for orisma_ranking*

Description

Print method for orisma_ranking

Usage

```
## S3 method for class 'orisma_ranking'
print(x, ...)
```

Arguments

x	An orisma_ranking object.
...	Further arguments (ignored).

Value

Invisibly returns x.

print.orisma_result *Print method for orisma_result*

Description

Print method for orisma_result

Usage

```
## S3 method for class 'orisma_result'
print(x, ...)
```

Arguments

x	An object to print.
...	Further arguments passed to or from other methods.

Value

Invisibly returns the input orisma_result object. Called primarily for its console-printing side effect.

Index