# JACoW reference and DOI metadata consistency plan

Oct 28, 2026 · @Johan

## Summary

Fix the data once, at the source, and generate every output from one metadata model. Today the same paper is described differently in the proceedings, the DataCite DOI record, the reference exports and the style guides, and several of those descriptions are wrong.

The plan has four workstreams, in this order:

1. **Source data in Indico**: complete, structured author names, ORCID iDs and affiliations with ROR IDs.
2. **New proceedings from CAT**: MEOW generates DataCite records that carry the proceedings title, pages, venue, event dates and structured authors.
3. **Past DOIs**: re-send corrected metadata for every JACoW DOI since 2014, from SPMS for older conferences and from Indico for newer ones. The DOIs themselves don't change.
4. **Reference formats**: the exports and the new CSL style follow the JACoW Unique Style Guide (JUStG), which replaces the separate Word and LaTeX guides.

**Precursor: one style guide.** The 2026 team meeting decided to replace the Word and LaTeX style guides with a single JACoW Unique Style Guide (JUStG) that both Word and LaTeX must be able to produce, owned by someone outside either camp (Petr proposed, Adriana as fallback). The plan for it is the JUStG fixed by the end of 2026, final templates by the end of February 2027 and tested by the end of March 2027 for IPAC'27, with the full set of web instructions made consistent. Choices should simplify authors' and editors' work and avoid special packages or fonts. The reference rules in this plan come from the JUStG: the findings below are input to it, and the exports, templates and CSL style follow it once it is fixed. The existing list of template tests is also not exhaustive, since LaTeX bugs surfaced at IPAC'26.

Out of scope for now: the Reference Search Tool. It is updated later, once it can read clean source data. All code changes to Indico, PURR, MEOW and SPMS go through change requests to their maintainers.

## What was found

Every DataCite record checked is incomplete or malformed, in a different way for each era, so no DOI-based tool can produce a correct JACoW reference today. The first check covered eight IPAC DOIs from 2017, 2022 and 2025, the five reference export templates and the two JACoW style guides; a full inventory of all 26,164 JACoW DOIs and a 50-DOI sample, below, confirm it.

| Field | IPAC'17 (kernel-3) | IPAC'22 (kernel-4) | IPAC'25 (kernel-4) |
|---|---|---|---|
| Author names | given/family split | given/family split | one string, `Cousineau,Sarah`; no split, no name type, no affiliation |
| Proceedings title | only in free-text series | missing (ISBN only) | missing |

| Field | IPAC'17 (kernel-3) | IPAC'22 (kernel-4) | IPAC'25 (kernel-4) |
|---|---|---|---|
| | description | | |
| Venue | in series description; parsed as first page "Denmark" | geo-location (Bangkok, Thailand) | missing |
| Conference dates | missing | `2022-06-12/2022-06-17` | missing |
| Pages | missing | missing | only as a size, `585-587 pages` |
| Issue date | 2017 | 2022 | `05 November 2025`, not ISO; read as year 5 |
| ISBN / ISSN | ISBN | ISBN | missing |

As a result, the DOI Citation Formatter prints references such as "Kelliher,David, Oeftiger,Adrian, and Jolly,Carl, “The Vaccaro stability diagram of scaling FFAs.” JACoW Publishing, 5 AD." DataCite also converts the ConferencePaper type to an empty citation type, which is a converter issue on DataCite's side.

### Full inventory: five generations of records

None of the 26,164 JACoW DOIs at DataCite has page numbers in a structured field, only 40 carry an ORCID iD, and none a ROR ID. The records fall into five generations, each broken in its own way. Counts are paper DOIs per generation; the 50-DOI sample took 20, 12, 13, 2 and 3 papers from them, one per conference.

| Generation | Conferences | Paper DOIs | What the DOI tools print | Sample result |
|---|---|---|---|---|
| Proceedings title only in free text | 65, 2010–2022 | 18,378 | Correct names and proceedings title, but "vol. IPAC2017, p. Denmark": the venue country lands in the page field | 20 of 20 without pages or city |
| Venue and dates, no proceedings title | 13, 1991–2024 | 2,174 | Correct names; "JACoW Publishing, Geneva, Switzerland" in place of the proceedings | 12 of 12 without proceedings |
| Garbled names and dates | 13, 2022–2025 | 5,005 | "Spear,Bethany *et al.*, … JACoW Publishing, 1 AD" | 13 of 13 with broken names and year |
| No type or publisher | 2, 2022–2023 | 305 | Names, title and year only | 2 of 2 without proceedings |
| Current | 3, 2024–2025 | 216 | "Wiencek, Mateusz, … JACoW Publishing, Sept. 08, 2026": full | 3 of 3 without proceedings |

| Generation | Conferences | Paper DOIs | What the DOI tools print | Sample result |
|---|---|---|---|---|
| MEOW code | | | given names and the registration date as the year | or initials |

Not one of the 50 sampled DOIs produces a usable JACoW reference: 0 have pages, 20 name the proceedings, 34 print author names as initials and family name, and 37 print a sensible year. The full results are in `jacow_doi_inventory_summary.csv` and `jacow_doi_sample50_citations.csv` in the Reference Architecture Overhaul folder.

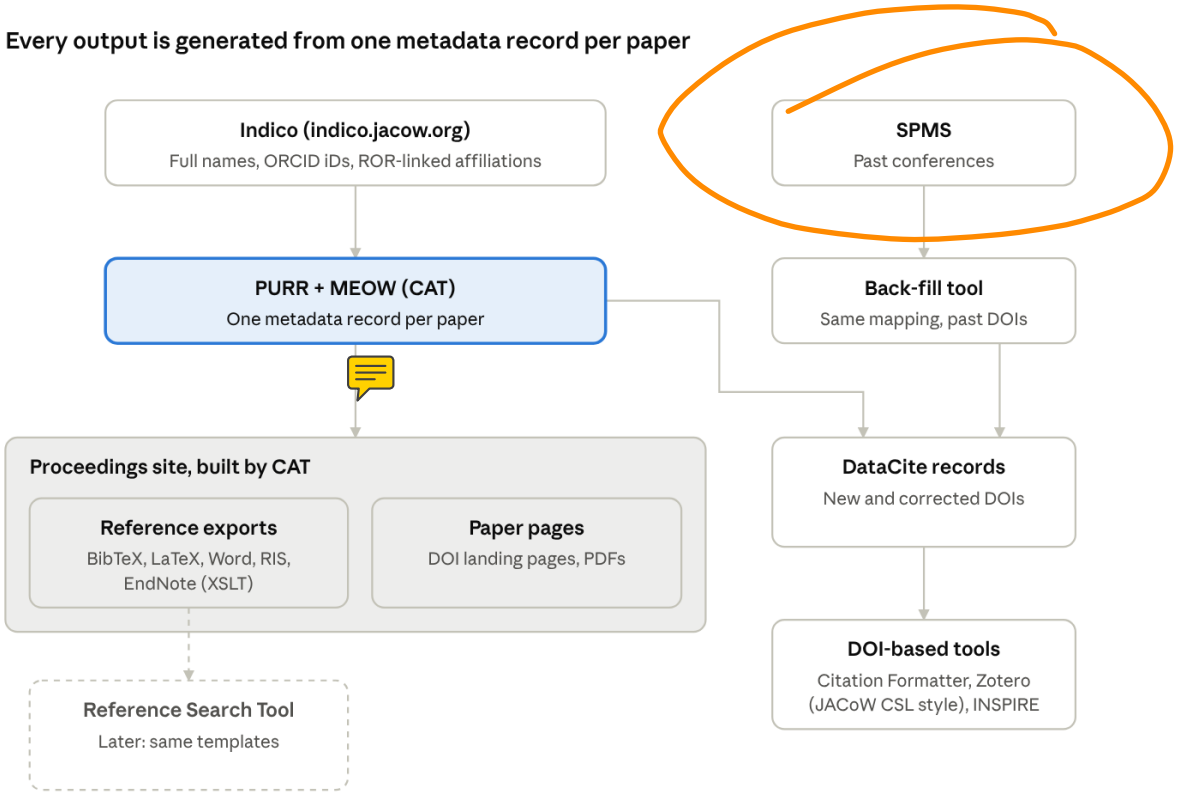
The reference exports (part of the proceedings sites) have their own problems:

- The Word and LaTeX templates used *et al.* above 3 authors; the guides say above 6.
- Papers from the current conference lose the paper code (LaTeX) or are called "unpublished".
- Only the first initial is kept, so "Jean-Pierre" becomes "J." instead of "J.-P.".
- RIS uses one space before the hyphen in tags (`TY  -  `) where the format needs two. EndNote has no reference type or proceedings title.
- The deployed templates differ between each released proceeding and are never updated.

The two style guides also disagree with each other and with themselves: the LaTeX guide asks for an en dash in page ranges but its examples use a hyphen, and example [10] in both guides prints "May."

## Target state

Indico is the single source for new conferences and ~~SPMS for past ones~~; everything downstream is generated, never typed again. One metadata record per paper, built by MEOW, feeds the proceedings site, including its reference exports, and DataCite, and the back-fill tool uses the same mapping for past DOIs.




target data flow · two sources, one record per paper, the proceedings site and DataCite as outputs

Fixing a field once in Indico, or once in the mapping, corrects every output that shows it.

## Workstream 1: source data in Indico

Every output is only as good as the author data in indico.jacow.org, so this workstream sets rules for that data and adds the two missing identifiers: ORCID for people and ROR for institutions. MEOW today sends each author's internal JACoW ID and free-text affiliation names to DataCite, with no ORCID or ROR.

1. **Agree the author data rules** and publish them in the author instructions:
    - Given and family names in separate fields, written in full, not initials.
    - Diacritics and hyphens kept (Jean-Pierre, Müller); name particles in the family name (Posthuma de Boer).
    - One affiliation record per institution, each linked to its ROR ID.
2. **Collect ORCID iDs** in Indico user profiles, ideally by signing in with ORCID or linking the account, so the iD is verified rather than typed.
3. **Link affiliations to ROR** by choosing institutions from a list instead of typing them.
4. **Check that contribution authors carry the data.** In Indico, a contribution's author list may hold its own copy of names and affiliations rather than a live link to the profile. If so, the copy has to pick up the ORCID and ROR ID too.

5. **Flag gaps before publication.** PURR's Papers Check, or a new check next to it, lists authors without full given names, ORCID or a ROR-linked affiliation, so editors can chase them before Final Proceedings is run.

Items 2, 3 and 5 are in progress in PR #32 on the Indico JACoW plugin: a JACoW profile with a verified ORCID iD and an ordered list of ROR-linked affiliations, every given name as an initial ("J.-P."), and event switches that require ORCID and ROR for newly added authors. The ORCID iD and ROR IDs don't reach CAT yet (change request 16), and existing contributions still need the pre-publication check. Item 4 remains a change request to the Indico and PURR maintainers.

## Workstream 2: DataCite records from CAT

MEOW already holds the proceedings title, venue, conference dates, ISBN and page range for every paper, but it doesn't send them to DataCite. The fix is mostly in one function: `_build_doi_attributes` in `meow/tasks/local/doi/models.py`.

| Field | What MEOW sends today | Change |
|---|---|---|
| Creators | `Last, First`, personal name type, JACoW ID, affiliation names | Add `givenName` and `familyName`; add ORCID as name identifier; add ROR ID to each affiliation |
| Proceedings | ISBN and ISSN as related identifiers only | Add a `relatedItems` entry: published in a conference proceeding, with title, ISBN, first page, last page and year |
| Pages | Only as a size, "585-587 pages" | First and last page in the related item |
| Venue | Not sent | Add as a geo-location place |
| Conference dates | Not sent | Add as a date of type Other, labelled as the event date |
| Publication year | The year the payload is built | The year of the Issued date |
| ISBN | Only added when the ISSN isn't `0000-0000` (the check tests the wrong field) | Test the ISBN itself |
| Publisher | JACoW Publishing | Decide whether to add Geneva, Switzerland, or a publisher identifier |

Two problems seen in the IPAC'25 records, the issue date "05 November 2025" and names like `Cousineau,Sarah`, don't match the current code, which sends `2025-11-05` and `Cousineau, Sarah`. The 216 records registered in 2026 confirm the fix: they have ISO dates and "Family, Given" names. Re-sending the IPAC'25 records with the current code would correct those two problems, but nothing else in the table.

Before each Publish, MEOW should check the payloads: required fields present, dates in ISO format, page range filled. A pilot conference is registered on DataCite's test system first, then its records are

read back through the DOI Citation Formatter and imported into Zotero to confirm they produce a correct JACoW reference.

The same review applies to the conference-level DOI and to the INSPIRE export in the same file, which also takes its year from the build date.

## Workstream 3: correcting past DOIs

All JACoW DOIs since IPAC'14 get corrected metadata, sent as updates to the existing records: the DOIs and their landing pages stay the same. DataCite allows metadata updates at any time through its API.

1. **Inventory.** List every 10.18429 DOI from the DataCite API, grouped by conference, with its schema version and which fields are present. Done in October 2026: 26,164 DOIs in five generations, summarised under What was found. Next is matching each generation to the tool that produced it.
2. **One mapping, two sources.** Write one converter from the metadata model to the DataCite payload. Feed it from an SPMS export for older conferences and from Indico, through MEOW, for CAT-era conferences.
3. **Never lose good data.** Compare each new payload with the live record first. Where the live record has something the source lacks, such as the IPAC'22 venue and event dates, keep it and log it.
4. **Pilot.** Correct one SPMS-era and one CAT-era conference on DataCite's test system, then in production. Check the result in the DOI Citation Formatter, Zotero and INSPIRE.
5. **Roll out conference by conference**, newest first, since recent papers are cited most. Keep a log of what changed per DOI.

The main risk is the SPMS data itself: if older records only hold initials or free-text affiliations, the corrected records can't be better than that. The inventory should measure this before the pilot. Updating published metadata in bulk also needs the publisher's approval.

## Workstream 4: reference formats

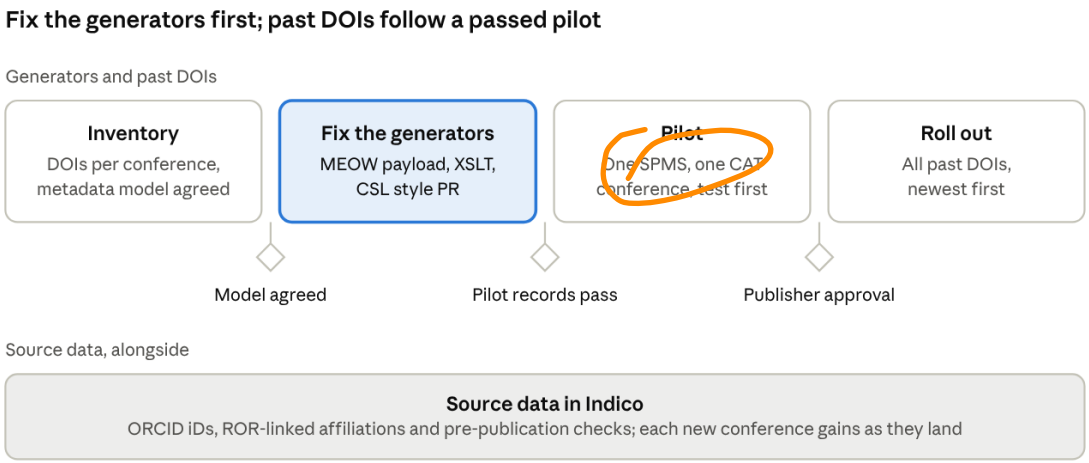
The reference exports on every proceedings site are made by MEOW from the five XSLT files in `xslt/`, so they are fixed there once. The Reference Search Tool's copies of `bibtex.xml`, `endnote.xml`, `latex.xml` and `ris.xml` are identical to MEOW's; the Reference Search Tool takes the corrected files later.

- **Word:** the corrected `word.xml` (HTML output, more than six authors for *et al.*, full initials, correct handling of each paper status) goes to MEOW as a pull request. The proceedings site must then show the reference as formatted text instead of a code block.
- **LaTeX:** the same author and status rules, `--` in page ranges, escaping of LaTeX special characters, and a citation key that matches BibTeX's.
- **BibTeX:** `month = jun` instead of `{06}`, the month in the note of unpublished papers, and a check of how `series` and `number` print.
- **RIS:** two spaces before the hyphen in every tag, `J2` as "Proc. IPAC'25", and no empty `ED`, `KW` or `T2` lines.
- **EndNote:** add the reference type and the proceedings title.

- **Tests:** render the JUStG examples through every template, export and the CSL style on each change, so a regression is caught before a conference goes out.
- **CSL style:** submit the prepared JACoW style to the CSL styles repository, so Zotero and Mendeley format JACoW references correctly; update it to the JUStG reference rules once they are fixed, before or after submission.
- **JUStG:** send the owner the reference questions this plan raised: *et al.* above six authors and the comma in "A. Alpha, B. Beta, *et al.*" for two primary authors; hyphen or en dash in page ranges; "May" without a period; whether DOIs keep a monospaced font such as Liberation Mono, given the aim of avoiding special fonts; and whether the hyphenated-initials and full-first-name options in PR #32 are allowed. Maurizio's BibTeX example entries can follow the JUStG reference order in the LaTeX template.

## Phasing

Fixing the generators comes first, because every conference published before that adds more DOIs to correct later. Past DOIs are only corrected after a pilot passes and the publisher approves. The Indico source-data work runs alongside and improves each new conference as it lands.




phases and gates · no dates set yet

Dates depend on how quickly the MEOW and Indico maintainers can take the change requests; the next conference due for final proceedings is the natural first target.

## Change requests

Sixteen changes are needed; the MEOW payload changes (rows 4 to 7), together with row 16 that feeds them ORCID and ROR, unlock the most and should be requested first. Owners and contacts are still to be added.

| # | Request | Where | Workstream | Status |
|---|---|---|---|---|
| 1 | ORCID iDs on user profiles, linked by signing in | Indico JACoW | 1 | In progress |

| # | Request | Where | Workstream | Status |
|---|---|---|---|---|
| | with ORCID | plugin, PR #32 | | |
| 2 | Affiliations chosen from a list linked to ROR IDs | Indico JACoW plugin, PR #32 | 1 | In progress |
| 3 | Check for missing full names, ORCID and ROR before Final Proceedings; PR #32 adds "Require ORCID" and "Require ROR" for new submissions | PURR | 1 | In progress |
| 4 | Given and family names, ORCID and ROR in DataCite creators | MEOW | 2 | Not started |
| 5 | Related item with proceedings title, ISBN and pages; venue; event dates; year from Issued | MEOW | 2 | Not started |
| 6 | Fix the ISBN condition that tests the ISSN | MEOW | 2 | Not started |
| 7 | Validate payloads before Publish | MEOW | 2 | Not started |
| 8 | Report the empty citation type for ConferencePaper records | DataCite support | 2 | Not started |
| 9 | Metadata update for already-published DOIs, with a diff log | MEOW or a separate tool | 3 | Not started |
| 10 | Export of authors, affiliations and pages per paper | SPMS | 3 | Not started |
| 11 | Corrected `word.xml` and fixes to the other four templates, with tests | MEOW `xslt/` | 4 | In progress |
| 12 | Show the Word reference as formatted text on paper pages | MEOW site templates | 4 | Not started |
| 13 | JACoW style pull request | CSL styles repository | 4 | In progress |
| 14 | Reference rules settled in the JUStG (page-range dash, "May", et al. comma, DOI font), and templates built to it by end of February 2027 | JACoW Unique Style Guide (owner: Petr proposed) | 4 | In progress |

| # | Request | Where | Workstream | Status |
|---|---|---|---|---|
| 15 | Approval for bulk metadata updates of past DOIs | JACoW publisher | 3 | Not started |
| 16 | Pass each author's ORCID iD and ROR-linked affiliations from Indico to CAT with the contribution data | Indico JACoW plugin, PURR | 2 | Not started |

## Open questions

- ☐ Which conferences were registered from SPMS and which from CAT? The inventory in workstream 3 answers this.
- ☐ Does SPMS hold full given names and structured affiliations for older papers, or only initials and free text?
- ☐ Indico 3.3 supports ROR-linked affiliations, and PR #32 adds ORCID through the JACoW plugin. Is indico.jacow.org on a version that can take the PR?
- ☐ Should the DataCite record lead with the conference month (Jun. 2025, as in the style guide) or the proceedings issue date (Nov. 2025)? DOI-based tools print the issue date.
- ☐ Publisher name: "JACoW Publishing" alone, with "Geneva, Switzerland", or with a publisher identifier?
- ☐ Who maintains JACoW-Indico plugin, PURR and the Indico server, and is a GitHub issue on JACoW-org the right way to request changes?
- ☐ Who at JACoW approves bulk updates to published DOI metadata?
- ☐ Do the editors accept the two name options in PR #32 that go beyond the style guides: hyphenating separate given names ("N.-J.") and writing the first name in full? If so, the reference templates and the CSL style need to follow them.

## Sources

- MEOW DataCite payload code and reference templates
- MEOW final proceedings tasks and DOI events
- PURR final proceedings settings
- DataCite records: MOOCA1 (IPAC'17), TUOXGD2 (IPAC'22), MOPS003 (IPAC'25)
- DOI Citation Formatter documentation
- JACoW MS Word and LaTeX style guides, Annex B / Appendix B