**CMCA Gold Standard Codebook (v2 – Final, Excel-Aligned)**

**Purpose:** Defines the rules for coding the CMCA Gold Standard dataset (Cases 1–30)

**:** For consistent coding, adjudication, and audit comparison.

**Part A – Column Definitions (Excel Headings)**

* **Case ID** – Unique identifier (1–30).
* **CMCA Platform Used (Yes/No)** – Whether a CMCA platform was used in the study.
* **Authors (Family name, Initial. “and” before the final author)** – Author list, formatted consistently.
* **Title** – Publication title.
* **Journal** – Publication venue.
* **Platform(s)** – Simplified categories (SEM, TEM, SIMS, Raman, FIB, etc.).
* **CMCA Co-author (Yes/No)** – Whether a co-author is from CMCA.
* **CMCA Co-author Name(s)** – If yes, list exact names.
* **CMCA\_Acknowledged (Yes/No), including NIF, Microscopy Australia, Metabolomics etc.** – Whether CMCA or related facilities were acknowledged. Includes NCRIS/Microscopy Australia boilerplate.
* **CMCA Ack Exact Text** – Copy acknowledgement wording verbatim.
* **Pages** – CMCA Ack Exact Text
* **CMCA Affiliation in Text (Yes/No)** – Whether CMCA/UWA appears in the affiliations list.
* **Author on CMCA Staff List (Yes/No)** – Whether an author is a current or past CMCA staff member.
* **Instrument Terms (verbatim)** – Exact instrument wording from PDF (vendor, model).
  + Copy the description from the PDF as closely as possible (vendor + model if given). Do not worry about exact punctuation/formatting. If only a generic term is written (e.g., “scanning electron microscope”), copy it as is.
* **Instrument Code (if matched)** – Standardised code (e.g., SEM125, TEM185, OPCON201).
* **Label (Positive/Negative/Ambiguous)** – First-pass classification.
* **Notes** – Clarifications, mismatches, rationale.
* **Adjudicated Label** – Final group-reviewed classification.
* **Adjudicated Notes** – Team rationale for final decision.
* **Negative Type (Random/Hard)** – Only applies if Label = Negative.
* **DOI** – Digital Object Identifier.

**Part B – Coding Rules**

**1. Acknowledgement**

* **High-rank phrases (direct CMCA/UWA mention)**
  + Examples:
    - “Centre for Microscopy, Characterisation and Analysis, The University of Western Australia”
    - “CMCA, UWA” / “CMCA-UWA” / “CMCA/UWA”
  + → **CMCA\_Acknowledged = Yes**
  + → **Label = Positive**
* **Low-rank phrases (Microscopy Australia / NCRIS only)**
  + Examples:
    - “Microscopy Australia enabled by NCRIS”
    - “Facility support from Microscopy Australia”
  + → **CMCA\_Acknowledged = Unclear**
  + → **Label = Ambiguous** (unless CMCA also named → then Positive)
* **No acknowledgement text**
  + → **CMCA\_Acknowledged = No**
  + → **Label = Negative**

**2. Affiliation & Co-authors**

* **Affiliation = Yes** if CMCA/UWA appears in the affiliation line.
* **Co-author = Yes** if any author is CMCA staff.
* **Author on Staff List = Yes** if the person matches the CMCA staff directory.
* **Staff + Affiliation = strongest evidence.**

**3. Platform & Instruments**

* **Platform(s):** Map into simplified categories (SEM, TEM, SIMS, Raman, FIB, etc.).
* **Instrument Terms:** Always copy verbatim from PDF.
  + Instrument Terms = PDF first, Audit second (in brackets). Exact wording differences (e.g., JEOL JEM-2100 vs JEOL 2100F TEM) do not need to be harmonised. Consistency is handled in the **Instrument Code**.
* **Instrument Code:** Standardise using CMCA Instruments lookup matrix.

**3.1 Audit vs PDF rules:**

* If audit lists an extra platform not in PDF → include in red brackets [Audit:\_\_\_].
* If audit conflicts with PDF → record both (PDF first, Audit in brackets).
* Always note mismatches in **Notes**.

**4. Label Definitions (PDF-first, Audit-second)**

* **Positive**
  + PDF shows direct CMCA acknowledgement (any format: CMCA, UWA / CMCA-UWA / CMCA/UWA / full name), OR
  + CMCA/UWA appears in affiliation line, OR
  + A co-author is CMCA staff.
  + → **First-pass Label = Positive**
  + Audit evidence may confirm or mismatch but does not override Positive.
* **Ambiguous (C-type, People)**
  + PDF has no CMCA acknowledgement, but a CMCA/UWA affiliation or staff author appears.
  + → **First-pass Label = Ambiguous (C-type)**.
* **Ambiguous (F-type, Facilities)**
  + PDF shows a microscopy platform/instrument **or** only generic Microscopy Australia/NCRIS acknowledgement,
  + But no CMCA acknowledgement, no CMCA affiliation, no CMCA staff.
  + → **First-pass Label = Ambiguous (F-type)**.
  + If Audit adds extra platforms, record them in **Platform(s)** as [Audit: …] and explain in Notes, but **do not change the first-pass Label**.
* **Negative**
  + PDF has no CMCA acknowledgement, no CMCA staff, no CMCA affiliation, and no platform.
  + → **First-pass Label = Negative**.
  + Subtypes:
    - **Random Negative:** Microscopy present, but no CMCA/Microscopy Australia/NCRIS in PDF or Audit.
    - **Hard Negative:** Microscopy present but acknowledges Microscopy Australia/NCRIS at a **non-CMCA node** (ANU, UNSW, Sydney, etc.) → potentially misleading.

**Recording rule:**

* Always base **Label** on the **PDF first**.
* Record **Audit evidence** in Platform(s) with [Audit: …] and explain in **Notes**.
* Only adjudication can change a Label after PDF + Audit review.

**5. Decision Tree**

CMCA/UWA mentioned directly? → Positive

Else if Microscopy Australia/NCRIS mentioned? → Ambiguous (or Hard Negative if adjudicated Negative)

Else if CMCA staff or affiliation are present? → Ambiguous (C-type)

Else if microscopy platform only? → Ambiguous (F-type)

Else → Negative (Random)

**Part C – Workflow (PDF vs Audit)**

1. **Check PDF first** – for acknowledgement, affiliation, co-author, instruments.
2. **Assign first-pass Label** based only on PDF.
3. **Cross-check Audit**:
   * If audit agrees → note “Audit confirms.”
   * If audit disagrees → record both, mark mismatch in Notes.
4. **Adjudication** = team review for conflicts.

**Part D – Worked Examples**

* **Case 5 (Positive):** Ack present, Cryo-SEM listed, audit agrees.
* **Case 12 (Ambiguous F-type):** No ack/staff, but SEM listed → Ambiguous.
* **Case 13 (Ambiguous C-type):** Martin Saunders staff/affiliation, no ack → Ambiguous; adjudicated Positive.
* **Case 19 (Ambiguous F-type):** Only “funded by NCRIS.” No staff/affiliation→ Ambiguous; adjudicated Positive.
* **Case 21 (Random Negative):** SEM in Germany; no CMCA/MA.
* **Case 28 (Hard Negative / Ambiguous):** Acknowledges Microscopy Australia (ANU node), not CMCA.

**Part E – Notes Templates**

* **Positive:**  
  “Acknowledgement = Yes. Affiliation/Co-author present. Platform(s): SEM. Audit confirms. First-pass = Positive.”
* **Ambiguous (F-type):**  
  “Acknowledgement = No. Affiliation = No. Co-author = No. Platform present. First-pass = Ambiguous (F-type).”
* **Ambiguous (C-type):**  
  “Acknowledgement = No. Affiliation = Yes (CMCA listed). Staff author present. First-pass = Ambiguous (C-type).”
* **Random Negative:**  
  “Microscopy study. No CMCA or Microscopy Australia/NCRIS. Affiliations external.”
* **Hard Negative:**  
  “Microscopy present. Microscopy Australia acknowledged at non-CMCA node (ANU/UNSW). First-pass = Ambiguous. Adjudicated = Hard Negative.”