Sylvia Le Dévédec – Leiden University, NL-BioImaging
ORCID: 0000-0002-0615-9616

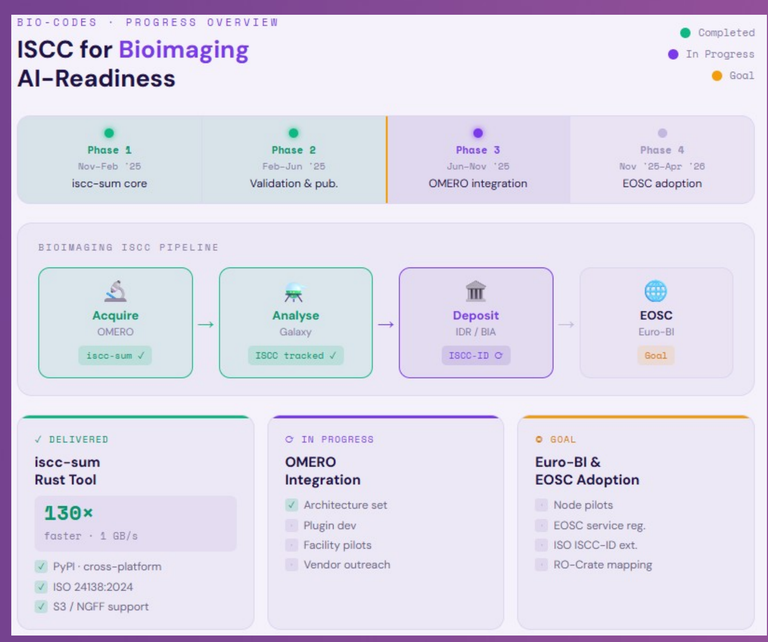
Enhancing AI-Readiness of Bioimaging Data with Content-Based Identifiers (BIO-CODES)

https://oscars-project.eu/projects/bio-codes-enhancing-ai-readiness-bioimaging-data-content-based-identifiers

Summary

ISCC: a new global, open standard to uniquely identify bioimages—whether human‑made or AI‑generated.

BIO‑CODES brings AI‑readiness to bioimaging by adding content‑based identifiers to complex datasets.

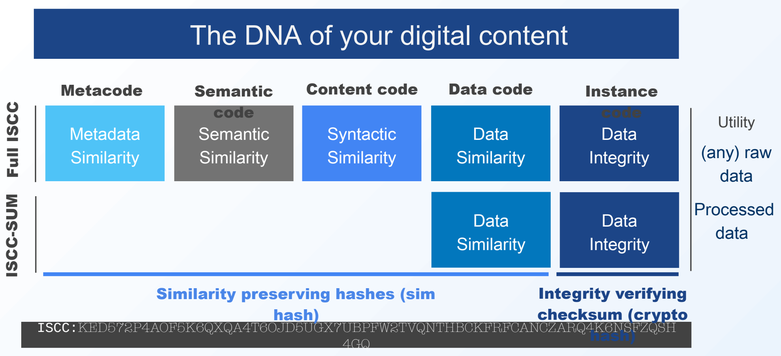


Challenge

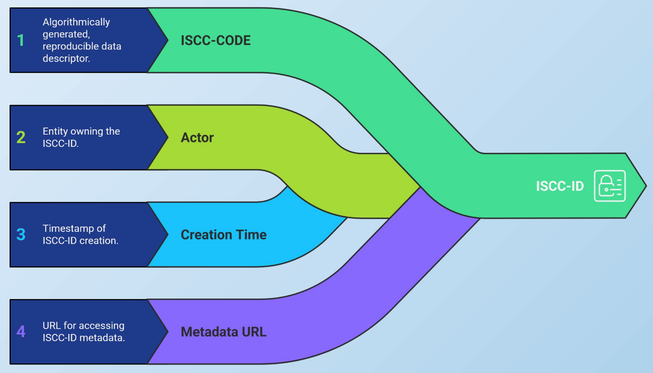
- **Growing volume of data:** Bio(imaging) data exist at different states (raw, repository, publications)
- **No Audit Trail:** Challenges to verify data integrity or detect manipulation
- **Lost Provenance:** Published figure disconnected from raw data and processing **steps**

Solution

- **International Standard Content Code**
  - ISCC-ISO 24138
  - Open and open-source
  - Interoperable content identification & fingerprinting system
- **Generating, signing and timestamping** ISCCs creates **persistent identifiers** to securely reference and link data in repositories with images in papers



Feature	Traditional Checksums	ISCC-SUM
Data Similarity Detection	❌ No	✅ Built-in
Container Level Checksums	❌ No	✅ Yes, storage agnostic
Standard Compliance	Various standards	ISO 24138:2024



ISCC *(International Standard Content Code)*

- **Standardized (ISO 24138) multi-component fingerprint** for various media types and files formats
- **Computed from the asset itself** (can never be removed)
- Can assess similarity at semantic, syntactic and data level

ISCC-SUM

- ISCC of bitstream input based combination of similarity and crypto hash (Data- & Instance-Code)
- Checks **data integrity** and **near-duplicates** in any large scientific datasets
- **High-performance, faster than sha256**
- Cross-platform (Linux, Mac, Windows)
- Handles container formats (e.g. **NGFF**)

sum.iscc.codes

ISCC-ID

- **Persistent identifier** from ISCC content codes
- Fingerprinting helps find metadata
- **Digital signing proves authenticity**
- **Timestaming** demonstrates when content was created
- **Secure** ISCC-ID linking enables **provenance verification**

Sustainability

- Automate ISCC code generation at the point of image acquisition for better data traceability.
- Exploring integration with existing bioimaging metadata standards like RO-Crate and platforms like OMERO
- Evolve ISCC-ISO 24138 to include ISCC-ID

Project homepage: https://bio-codes.io/

Soon to come: ISCC / BIO-CODES community on Image.sc

Funded by the European Union

Universiteit Leiden · ISCC · German BioImaging Gesellschaft für Mikroskopie und Bildanalyse · NL-BIOIMAGING AM