



OSCARS

Open Science Clusters' Action
for Research & Society

Funded Project

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

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

Implemented by



Universiteit
Leiden



German
Bioluminescence

Gesellschaft für Mikroskopie und Bildanalyse

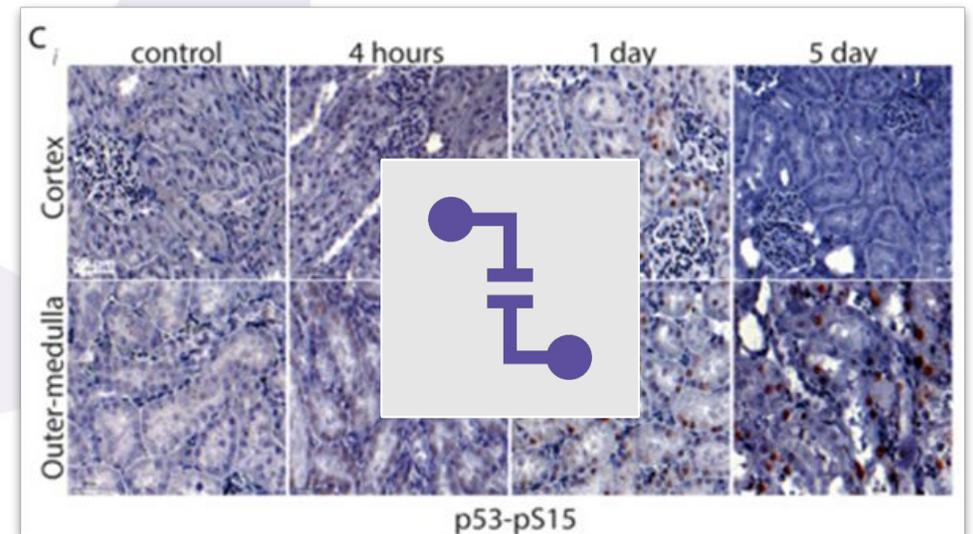
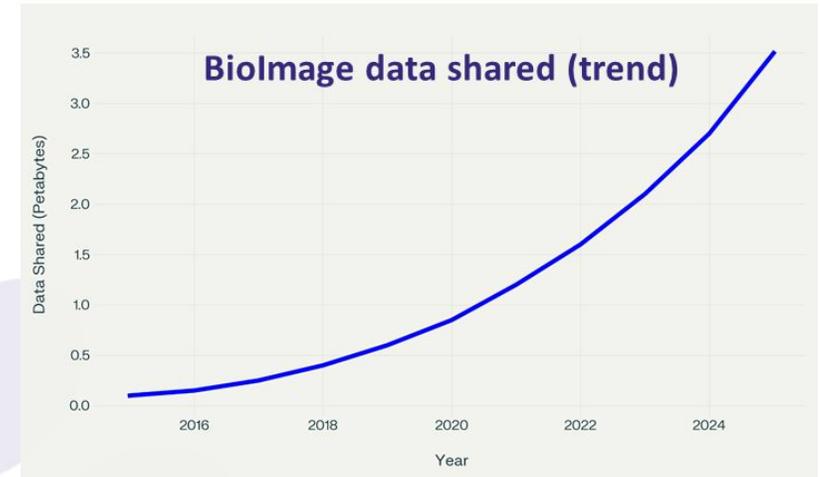


NL-BIOIMAGING AM



Funded by
the European Union

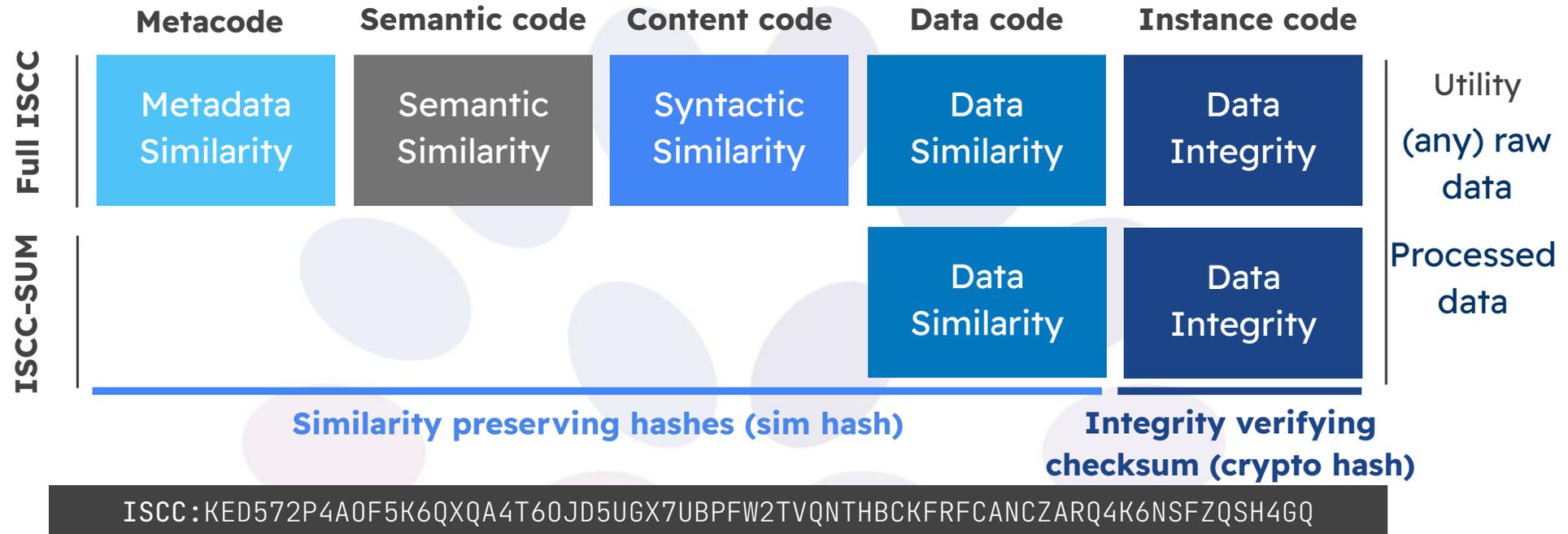
- **Growing volume of data:** Bio(imaging) data exist at different states — raw, repository, publications — with no shared identity across them.
- **No Audit Trail:** Challenges to verify data integrity or detect manipulation after the fact.
- **Lost Provenance:** Published figures are disconnected from raw data and the processing steps that produced them.



Lukas S Wijaya et al., *Cell Biol Toxicol*. 2025

- **International Standard Content Code (ISCC ISO 24138)**. Open and open-source. Interoperable content identification & fingerprinting system. Computed directly from the asset itself — can never be removed or decoupled from the data.
 - **Anyone can compute ISCCs from available data** — independently and without any central authority. Use the ISCC Generator at <https://iscc.io/resources/>
 - **Generating, signing, and timestamping ISCCs** creates persistent identifiers to securely reference and link data in repositories with images in papers.
-

The DNA of your digital content



- **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

OPEN SOURCE TOOLS

The BIOCODES toolkit

Three complementary tools covering the full bioimaging identification workflow, all Apache 2.0 licensed.

iscc-sum

Stable v0.1

High-performance ISCC Data-Code and Instance-Code generation. Single-pass processing with a Rust core and Python bindings — a drop-in replacement for md5sum and sha256sum in scientific pipelines. Faster than SHA-256 at any data size.

```
pip install iscc-sum
```

PLATFORMS: Linux macOS Windows

FORMATS: Zarr HDF5 OME-TIFF

NGFF

Docs Quickstart GitHub PyPI

Rust + Python

CLI

Apache 2.0

iscc-bio

Beta

ISCC processing for multi-dimensional bioimage data. Implements the IMAGEWALK specification — deterministic Z→C→T plane traversal for format-agnostic, reproducible content hashing of microscopy volumes.

PLATFORMS: Linux macOS Windows

FORMATS: OME-TIFF OME-Zarr CZI

ND2 LIF DICOM HDF5

GitHub PyPI

Python

OME-TIFF

OME-Zarr

CZI / ND2 / LIF

Apache 2.0

omero-iscc

Alpha

OMERO server integration plugin. Automatically generates and stores ISCC identifiers for images imported into OMERO, enabling facility-level deduplication and FAIR-compliant provenance tracking.

PLATFORMS: Linux macOS

GitHub

OMERO Plugin

Python

Server

Apache 2.0

<https://bio-codes.io/>

COMMUNITY & INTEGRATIONS

Built for the open bioimaging ecosystem

BIOCODES integrates with the tools researchers already use — no new infrastructure required.

OMERO

Available

Server plugin for automatic ISCC generation on image import. Enables facility-level deduplication and provenance tracking within OMERO's image management infrastructure.

- Automatic ISCC generation on import
- Facility-level deduplication
- FAIR-compliant metadata annotation

`omero-iscc`

Galaxy

Galaxy

Available

Galaxy tools for ISCC generation, near-duplicate detection, and content verification within reproducible Galaxy workflows. Part of the BMCV galaxy-image-analysis tool suite.

- `iscc_sum` — ISCC code generation for any file
- `iscc_similarity` — near-duplicate detection across datasets
- `iscc_verify` — content integrity verification

`galaxy-image-analysis / iscc-sum`

Napari

Planned

Plugin integration for interactive image analysis and ISCC annotation within the Napari viewer.

- Interactive ISCC annotation
- In-viewer provenance display

<https://bio-codes.io/>

TEAM



Titusz
Pan



Martin
Etzrodt



Kira
Lemke



Sebastian
Posth



Maarten
Paul



Sylvia
Le Dévédec



Josh
Moore



Universiteit
Leiden



NL-BIOIMAGING AM



German
Bioluminescence

Gesellschaft für Mikroskopie und Bildanalyse

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

Soon to come: ISCC / BIO-CODES community on [Image.sc](https://image.sc)