



# 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



Funded by  
the European Union

## What problem(s) are you going to solve?

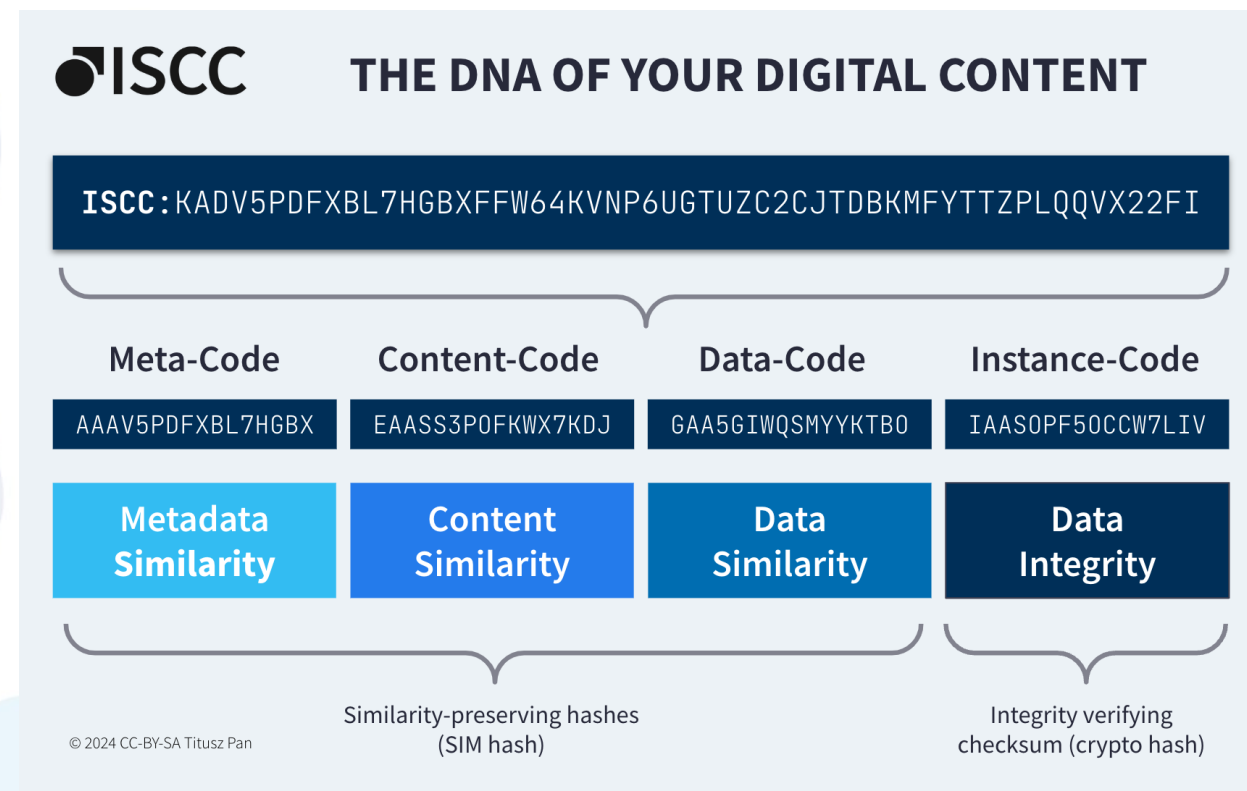
- Rapid **growth of bioimaging data** presents challenges in management, integrity, and reusability.
- **Lack of standardized identifiers** for bioimages leads to difficulties in tracking, verifying, and certifying data.
- Existing **FAIR principles** (Findability, Accessibility, Interoperability, Reusability) are not fully implemented in bioimaging workflows.
- Use of AI in bioimaging data analysis requires transparency in data provenance and authenticity.

## What are you planning to do to solve the problem?

- Implementing the International Standard Content Code (ISCC) for bioimaging data (e.g. OME-Zarr) and figure renderings.
- Developing a **proof-of-concept integration** of ISCC with existing bioimaging workflows.
- Automating 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.

## What is ISCC?

- ISCC (International Standard Content Code)
- ISO standard (ISO 24138) for content-derived identification of digital media (text, images, video, audio, mixed).
- A composite identifier that exhibits similarity-preserving properties
- Generated directly from the media asset – no central database or manual registration needed.



Source: [iscc.io](https://iscc.io)

## What is ISCC?

- ISCC (International Standard Content Code) is an ISO standard (ISO 24138) for content-derived identification of digital media, including text, images, video, and audio.
- ISCC generates a unique, reproducible identifier directly from the media asset – no central database or manual registration needed.
- ISCC combines cryptographic and similarity hashes, enabling:
  - Content integrity verification
  - Similarity detection
  - Transparent tracking of content provenance and versions

## What will be the results and how do you plan to make them available to the broader community?

- Standardized **content-based identifiers** for bioimaging data.
- Improved data **integrity, transparency, and reusability**.
- ISCC adoption in Euro-BioImaging and EOSC infrastructures.
- Enhanced **AI-readiness** of bioimaging datasets for trusted AI applications.
- Integration of ISCC within **vendor platforms** and research infrastructures.

## What risks could limit the success of the project, and how can they be mitigated?

- **Adoption Barrier:** Technical complexity may cause delay in adoption by bioimaging facilities and vendors → **Mitigation:** Engage stakeholders early, provide material for simple onboarding.
- **Technical Integration:** Development of ISCC for bioimaging data incl. proprietary image formats is challenging → **Mitigation:** Considering dedicated ISCC code unit for only OME-Zarr (and most common bioimaging formats).

## Who is doing it?

- **Leiden University:** Coordination, bioimaging data management.
- **ISCC Foundation:** ISCC technology implementation.
- Key Contributors:
  - **Sylvia Le Dévédec** (Leiden University) – Bioimaging data scientist.
  - **Katy Wolstencroft** (Amsterdam UMC) – FAIR data management expert.
  - **Titusz Pan, Martin Etzrodt** (ISCC Foundation) – ISCC lead developer.
  - **Josh Moore** (German BioImaging, OME team) – Advisor.
  - **Sebastian Posth** – Project manager.
  - Team from **Leiden University** (Leiden Cell Observatory, **Joost Willemse**) and **LUMC** (Light and Electron Microscopy Facility, **Maarten Paul, Lennard Voortman**)