



# Open Science Monitoring Initiative

WG4. Shared resources and infrastructure to analyze scholarly outputs

# OSMI WG4 - Subgroup Kick-Off

Cristina Huidiu  
Pragya Chaube

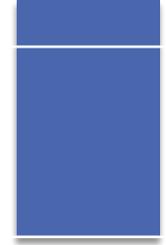


# Why subgroups?

- ▶ WG4 is broad.
- ▶ We need spaces that allow:
  - ▶ Ongoing exchange between technical & non-technical members
  - ▶ Visibility into existing projects
  - ▶ Identification of shared needs
  - ▶ Avoiding duplicated work
  - ▶ Faster informal collaboration

These are **not project teams**.  
They are **collaboration hubs**.

# What subgroups are (and **are NOT**)



## **Subgroups ARE:**

- ▶ Lightweight, self-organising spaces
- ▶ Places to bring existing work
- ▶ A way to surface needs and gaps
- ▶ Cross-pollination between tech & policy
- ▶ A matchmaking layer across initiatives

## **Subgroups are NOT:**

- ▶ Grant consortia
- ▶ Formal deliverable-driven teams
- ▶ Heavy governance structures
- ▶ Extra reporting layers

# How the model works

We invite members to:

- ▶ Bring existing projects
- ▶ Share technical challenges
- ▶ Share infrastructure needs
- ▶ Offer tools or expertise
- ▶ Identify collaboration overlaps

Think of it as:

**“What are you already building, and where could collaboration strengthen it?”**

# Subgroups 1 & 2 (Technical collaboration hubs)

## 1. Entity Extraction

## 2. Open Categorisation Algorithms

### These function primarily as:

- ▶ Technical exchange spaces
- ▶ Demonstration forums
- ▶ Needs-matching platforms

### What members can bring:

- ▶ Existing NLP pipelines
- ▶ Topic models
- ▶ Classification frameworks
- ▶ Annotation challenges
- ▶ Evaluation methods
- ▶ Infrastructure constraints

# The Bridge: Subgroup 3 (Economics of TDM)

## This subgroup:

- ▶ Draws from technical work in 1 & 2
- ▶ Bridges toward policy, licensing, and governance
- ▶ Translates infrastructure realities into economic discussions

## Key Role

- ▶ Connect:
  - ▶ Technical possibilities  
↕  
Cost models, sustainability, governance, licensing
- ▶ This is where:
  - ▶ Infrastructure meets policy
  - ▶ Compute costs meet access rights
  - ▶ Open science meets funding realities

# Subgroups 4 & 5 (Policy & system alignment)

## 4. TDM maturity & adoption mapping

## 5. Open infrastructure landscape analysis

These are more policy and systems-oriented.

Instead of defining detailed outputs upfront:

- ▶ Let's join Discord
- ▶ Share relevant initiatives
- ▶ Identify overlaps with Barcelona Declaration & others
- ▶ Clarify realistic next steps organically

### What members can bring:

- ▶ Existing NLP pipelines
- ▶ Topic models
- ▶ Classification frameworks
- ▶ Annotation challenges
- ▶ Evaluation methods
- ▶ Infrastructure constraints

# A dynamic hub model

We hope WG4 subgroups would become:

- ▶ A matchmaking layer
- ▶ A circulation space for tools & needs
- ▶ A bridge between technical builders and policy actors
- ▶ A coordination space across initiatives

**Not a hierarchy.**

**Not a silo.**

**A dynamic ecosystem node.**

# What participation can look like

Members can:

- ▶ Present an existing project
  - ▶ Share a challenge
  - ▶ Offer infrastructure or tools
  - ▶ Ask for collaborators
  - ▶ Connect another initiative
  - ▶ Flag policy blockers
  - ▶ Translate technical realities to governance concerns
- Both technical and non-technical expertise are needed.

# How we'll coordinate

- ▶ Full WG4 check-in every 6 weeks
- ▶ Discord for lightweight, ongoing exchange
- ▶ Subgroups define their own rhythm
- ▶ WG4 provides visibility & cross-linking

# Open questions

- ▶ Does this hub model resonate?
- ▶ Are scopes appropriate?
- ▶ Do we see missing bridges?
- ▶ Who wants to seed each subgroup with an existing project?
- ▶ How do we ensure cross-talk between tech & policy?

Thank you for your attention!  
<https://open-science-monitoring.org/>

