# Challenges in the Adoption of Pipelines in the LA Community

ALA-LAs Workshop (Madrid, March 2025)

David Martin, Tim Robertson, Vicente J. Ruiz Jurado, 2025/03/12

# Current Adoption Status

- **ALA**, **Flanders**, **France**, **Spain**, and **Sweden**: Using pipelines in production.
- **UK**: Still in the testing phase?
- **Kew**: Starting to test.
- **Other LA Portals**: Not using `pipelines` yet.

# Key Barrier: Hardware Requirements

- `pipelines` require significant hardware resources when dealing with big data resources.

- Memory-intensive steps (e.g., duplicate detection) impact performance.

- This has been a concern since initial testing.

# Challenges for Newcomers

Because of that:

- Newcomers are installing older ALA components that rely on the legacy `biocache-store` backend, instead of the newer `pipelines` system.
- Relying on quite outdated software:
    - Ubuntu, Java, database software, Tomcat, etc.

# Challenges for Newcomers (cont)

Causing:

- Compatibility issues between old and new components: Some legacy ALA components depend on `biocache-store`, making it harder to transition to `pipelines`. Mixing old and new versions leads to unexpected failures and technical debt.
- Delayed adoption of new functionality or **Docker-based** deployments.
- Results in continued support for legacy software.

# Additional Challenge: ALA's Outdated Branch

- ALA is currently using a branch of `gbif/pipelines` that is slightly outdated (issue #1084).
- This complicates patching and maintenance efforts.

# Possible Solutions

1. **Optimize pipelines code**
   - Reduce hardware demands.
   - Avoid the use of **spark** using a `java` only solution.
   - Allocate development resources (e.g., through **CESP**, or similar initiatives).
2. **Shared infrastructure between portals**
   - Technically feasible but administratively challenging.
3. **Cloud-based** `pipelines` **clusters**
   - Build/destroy on demand, similar to ALA's approach.

# The Approach of countries using `pipelines`

- **Investing in hardware** rather than development resources to improve `pipelines`.
- Aim: Minimize hardware costs by focusing on software optimization.

# Discussion & Next Steps

- How can we improve adoption among LA Portals?
- What are the main challenges preventing adoption in your country?
- Is there interest in collaborative development efforts?
- Open floor for discussion.