

Part 1: Works Single View

A Single View application aggregates and reconciles data from multiple sources to create a single view of an entity, in this case, a musical work.

A musical work consists of the musical notes and lyrics (if any) in a musical composition. A musical work may be fixed in any form, such as a piece of sheet music or a sound recording. It is usually represented by metadata like: title, contributors, roles, duration, etc.

Your task is to provide the code to create the Works Single View from the metadata in `works_metadata.csv` file. This implies parsing the input file, detecting duplicate musical works and storing reconciled works in a database.

Metadata from some sources may be incomplete, the aim of the Works Single View is to have, for each musical work, the most complete metadata without duplicates.

Instructions

- A Django project has to be created.
- Ingestion logic should be called with a Django command.
- To store reconciled works use a PostgreSQL database.
- Using docker is a plus.
- Instructions should be provided on how to execute your code.

`works_metadata.csv` columns:

- `title` : Work title
- `contributors` : Work contributors such as composers, lyricists, etc. we skip the role for simplicity. There can be multiple contributors, they are separated by `|`
- `iswc` : International Standard Musical Work Code, it's a musical work identifier

Musical work reconciliation tips

- ISWC could be enough to identify a work. If no ISWC is provided, title and at least one contributor should be the same.
- In case there are different sets of contributors from different inputs, you could take the union of them.

Questions

1. Describe briefly the matching and reconciling method chosen.
2. We constantly receive metadata from our providers, how would you automatize the process?

Part 2: Works Single View API

A customer has sent us a report with ISWC codes. He wants us to enrich it with contributors metadata.

Your task is to create a simple API to query the Works Single View by ISWC in order to get the metadata related to that work.

Instructions

- The API has to be written in Python using **Django Rest** framework.
- Response has to be in **JSON** format.
- Instructions should be provided on how to execute your code.

Questions

1. Imagine that the Single View has 20 million musical works, do you think your solution would have a similar response time?
2. If not, what would you do to improve it?