

Digital Music Lab

An AHRC Digital Transformations Project

A New System for Big Music Data Analysis

Daniel Wolff



Arts & Humanities
Research Council



CITY UNIVERSITY
LONDON



Queen Mary
University of London



LIBRARY
HSILIRB

The DML System Provides ...

- ▶ **Access:** Systematic exploration of heterogeneous and **large music libraries**
- ▶ **Control:** Interfacing with complex automatic **music analysis tools**
- ▶ **Analysis:** Gain **summarised** knowledge on large numbers of recordings
- ▶ **Sharing:** Experiments **reproducible** with same data, clear **provenance** of analysis results.



Arts & Humanities
Research Council

The Technical Perspective

- ▶ **Access to data**
 - ▶ Audio – access restricted by physical location
 - ▶ Metadata – unification of different formats
- ▶ **Control** via web interface to large-scale analysis
 - ▶ Interactive UI for overview and exploration
- ▶ Scalable **analysis** is available on collection-level and recording-level
- ▶ **Share** the well-defined and derived data
 - ▶ Re-use of existing software and published code for analysis



Arts & Humanities
Research Council

Software Ecosystem

- ▶ **Distributed** system

- ▶ Virtual machines (VirtualBox)
- ▶ Open Source OS (Ubuntu)



- ▶ **Parallelised** existing analysis tools

- ▶ Python (NumPy)
- ▶ **Vamp Plugins**
- ▶ Big-Data **map-reduce** (Spark)



- ▶ **Computation management**

- ▶ Built on **semantic** architecture

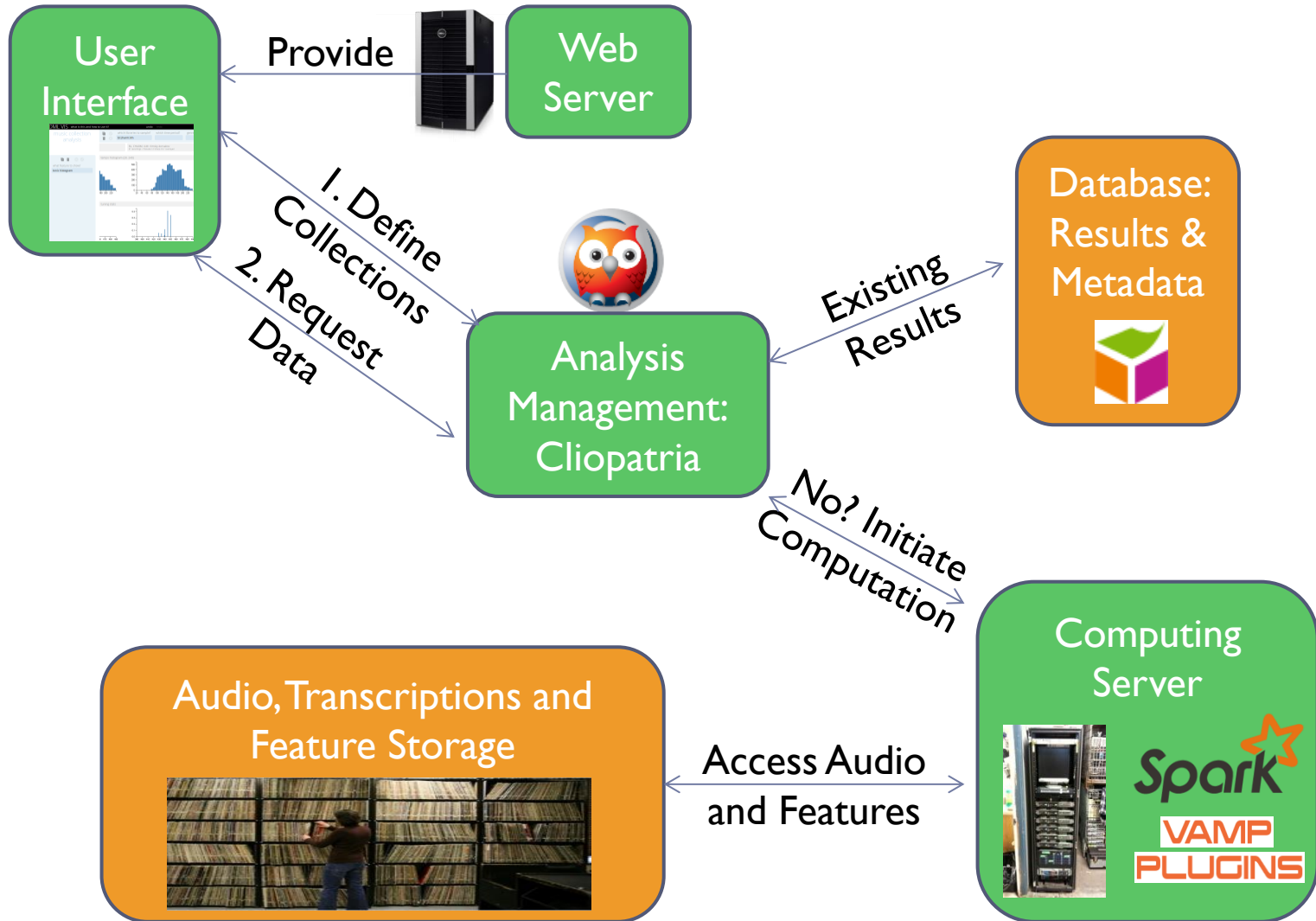


- ▶ Interactive user interface for **exploration** and **analysis**

- ▶ Built using state-of-the-art web technologies

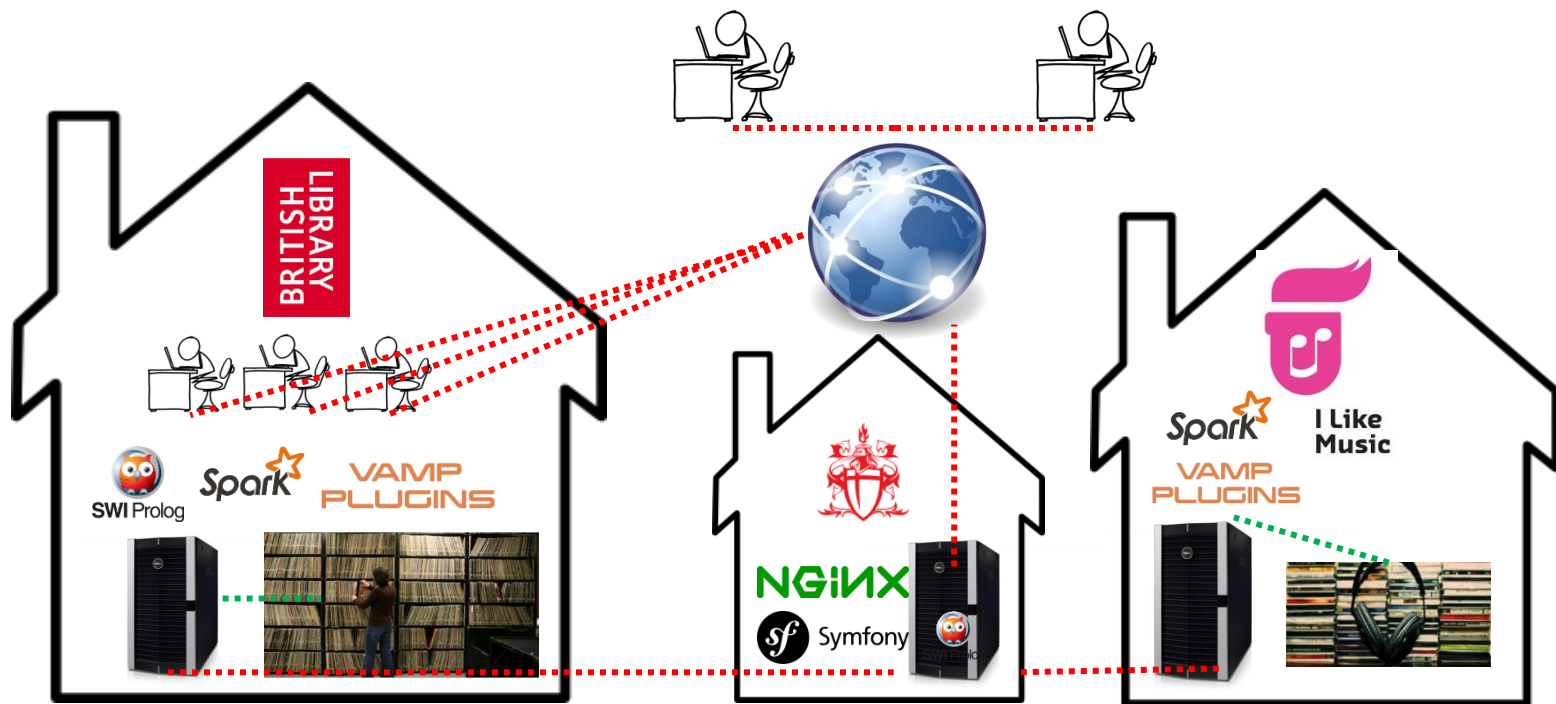


Data-Flow for Computational Analysis



Physical Locations Matter: Content Access

- ▶ Two computing servers, located at BL and ILM
 - ▶ Allow for **in-place access** to restricted data
- ▶ Dedicated server at City for web access



Sustainability

- ▶ Preference on **Open Source**

- ▶ Basic infrastructure (Ubuntu, Spark, Vamp ...)

- ▶ Soundsoftware repository for

- ▶ **Publishing versioned code** of newly developed software
- ▶ **Backup and sharing:** Open data / features / results

 soundsoftware.ac.uk
Sustainable Software for Audio and Music Research

- ▶ Open and reproducible method

- ▶ **Enables similar set-up** in further institutions



Arts & Humanities Research Council

Results Implemented in the DML System

- ▶ Conceptual framework (including implementation) for **collection-level analysis**
 - ▶ Collection in focus as object of analysis
- ▶ Data-flow allowing for **interactive retrieval of results**
 - ▶ Secure, responsive and redundant network structure
- ▶ **Distributed placement** of computation resources
- ▶ Open-source software ecosystem for large-scale music analysis
 - ▶ **Parallelised feature extraction** and **results management**
 - ▶ **Collection-level** analysis, interface and visualisation



Arts & Humanities
Research Council