

Gustavo Polins Pedro and Andrew Kam

SIMSSA WORKSHOP 2018

SIMSSA DATABASE

May 28, 2018

OVERVIEW

- ▶ Current ELVIS Database
- ▶ New SIMSSA Database
- ▶ Objectives
- ▶ Current status
- ▶ Challenges

ELVIS DATABASE

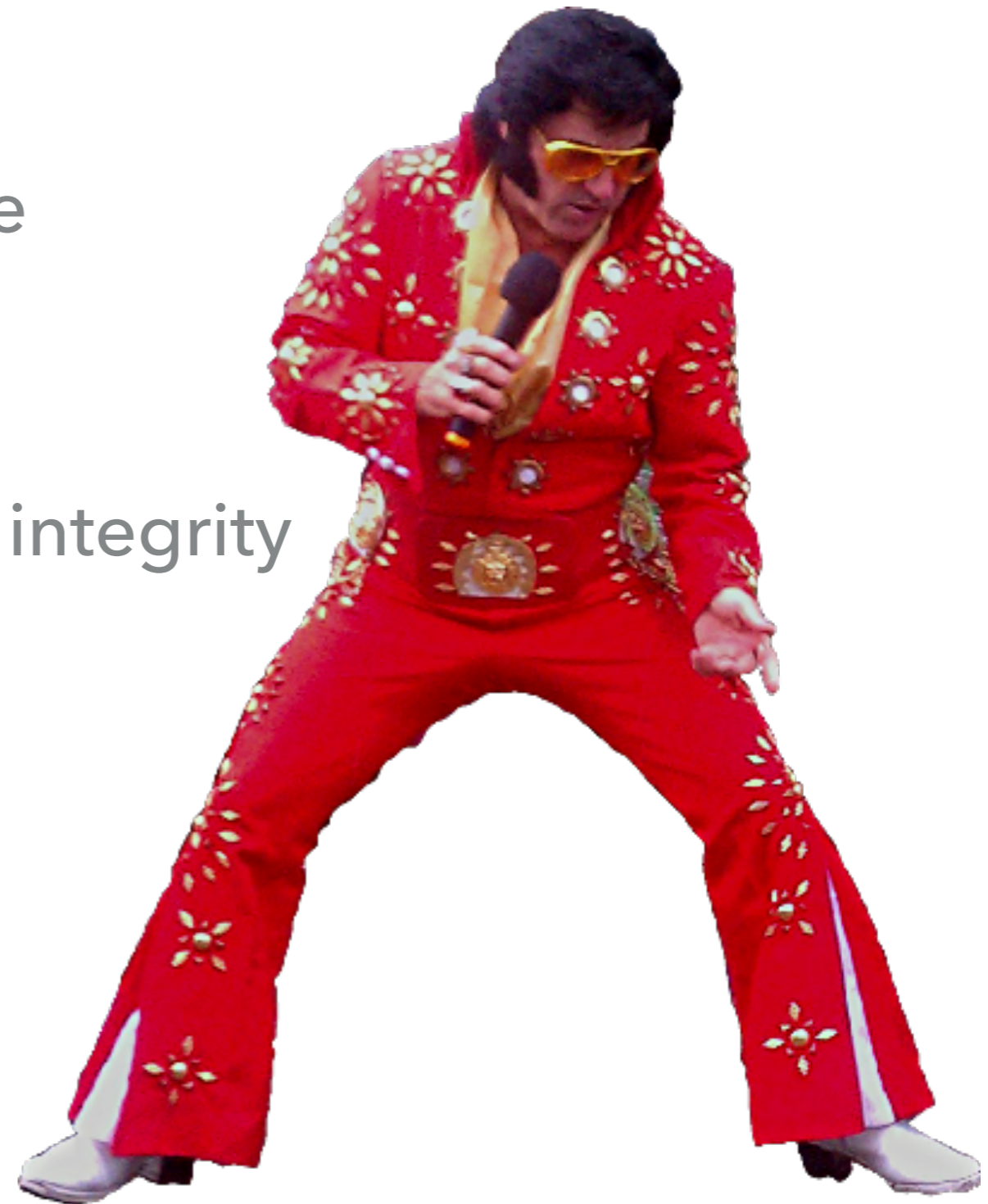
- ▶ Part of the Electronic Locator of Vertical Interval Structures (ELVIS) project started by Prof. Julie Cumming and colleagues in 2012
- ▶ Database of symbolic music files
- ▶ Computational music research tool
- ▶ Very useful tool, but there are challenges
 - ▶ Some of the data became inconsistent and redundant
 - ▶ Data model not as expressive as we want



Cumming, Julie E., Michael Scott Cuthbert, and Frauke Jürgensen. "Electronic Locator of Vertical Interval Successions (ELVIS): The first large data-driven research project on musical style." White paper report ID 105636. (2014)

NEW SIMSSA DATABASE

- ▶ Evolution of the ELVIS Database
- ▶ More expressive data model
- ▶ Focus on accessibility and data integrity



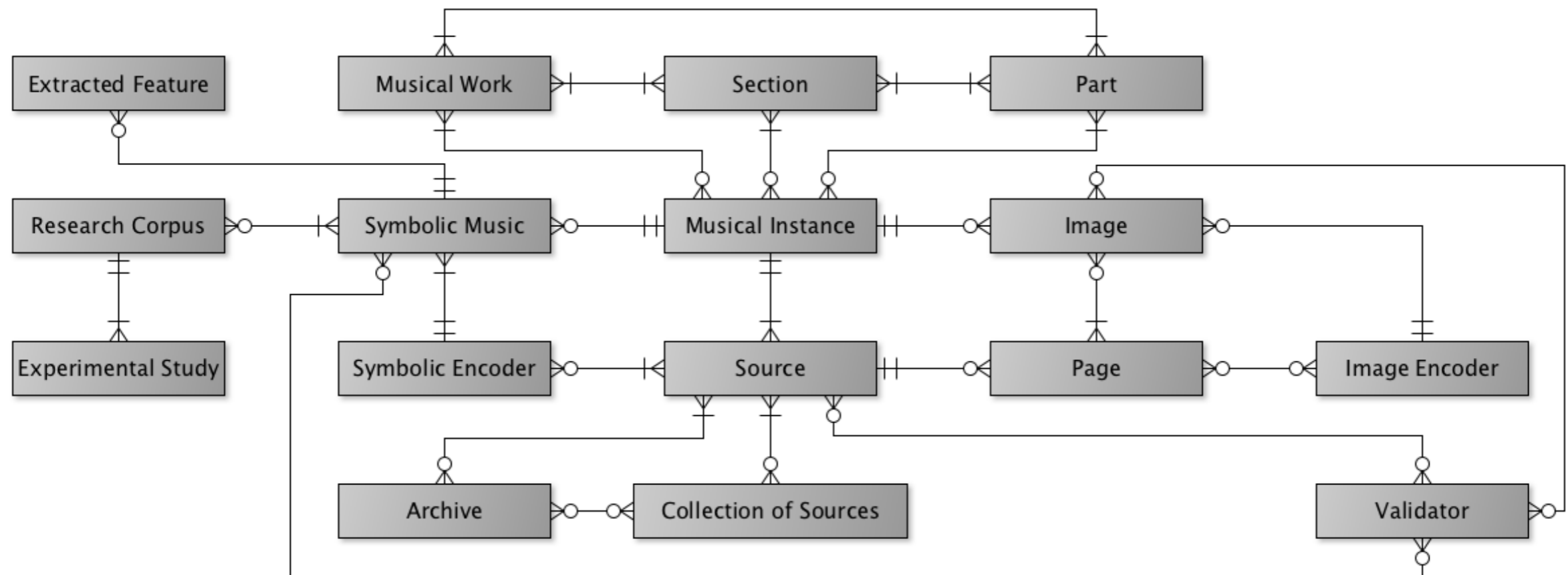
OBJECTIVES

- ▶ Researchers can access and contribute music and metadata
- ▶ Keep records of studies and experiments for replication purposes
- ▶ Search by metadata and content
 - ▶ i.e., features extracted with Musical Information Retrieval software tools
- ▶ Maintain provenance data
- ▶ Connect with other datasets using a Linked Open Data approach
- ▶ Future integration with OMR workflows

SAMPLE QUERIES AND USE CASES

- ▶ Find 4-voice pieces composed between 1480 and 1520 and sort them by final and key signature
- ▶ Return all pieces of Renaissance music that contain a vertical or horizontal tritone in one or more of their sections
- ▶ Find all the pieces composed between 1550 and 1600 where we have both MIDI and XML files
- ▶ Add some symbolic music files from research corpus Y to experimental study Z

NEW DATA MODEL



- ▶ Developed by Cory McKay, Andrew Hankinson, Julie Cumming, and Ichiro Fujinaga
- ▶ Highlights:
 - ▶ Musical Instance
 - ▶ Sources
 - ▶ Research Corpus, Experimental Study, Extracted Features

COMPETING PRIORITIES

- ▶ Accessibility
- ▶ Data Quality
- ▶ Cost
- ▶ Flexibility
- ▶ Integrity

DESIGN CHALLENGES

- ▶ Correct representation of disputed or unknown dates and locations
- ▶ Implementing the Linked Open Data approach
- ▶ Representing provenance

RELATIONAL DATABASE

- ▶ Pros
 - ▶ Mature and established technology
 - ▶ The query language is nearly universal
 - ▶ Easier to enforce model integrity
- ▶ Cons
 - ▶ Rigid data schema
 - ▶ Certain relationships are harder to model
 - ▶ Not ideal for sparse or heterogeneous datasets

	id	title	composer	genre
1	1	Missa Almanana	P. de la Rue	Mass
2	2	Missa Assumpta est Maria	P. de la Rue	Mass
3	3	Missa Ave Sanctissima Maria	P. de la Rue	Mass
4	4	Missa Conceptio Tua	P. de la Rue	Mass
5	5	Missa de Beata virgine	P. de la Rue	Mass
6	6	Missa de Feria	P. de la Rue	Mass
7	7	Missa de Sancta Anna	P. de la Rue	Mass
8	8	Missa de Sancto Antonio	P. de la Rue	Mass
9	9	Missa de Sancto Job	P. de la Rue	Mass
10	10	Missa de Septem doloribus	P. de la Rue	Mass
11	11	Missa de virginibus	P. de la Rue	Mass
12	12	Missa Incessament	P. de la Rue	Mass
13	13	Missa Inviolata	P. de la Rue	Mass

The result of a simple query in a relational database

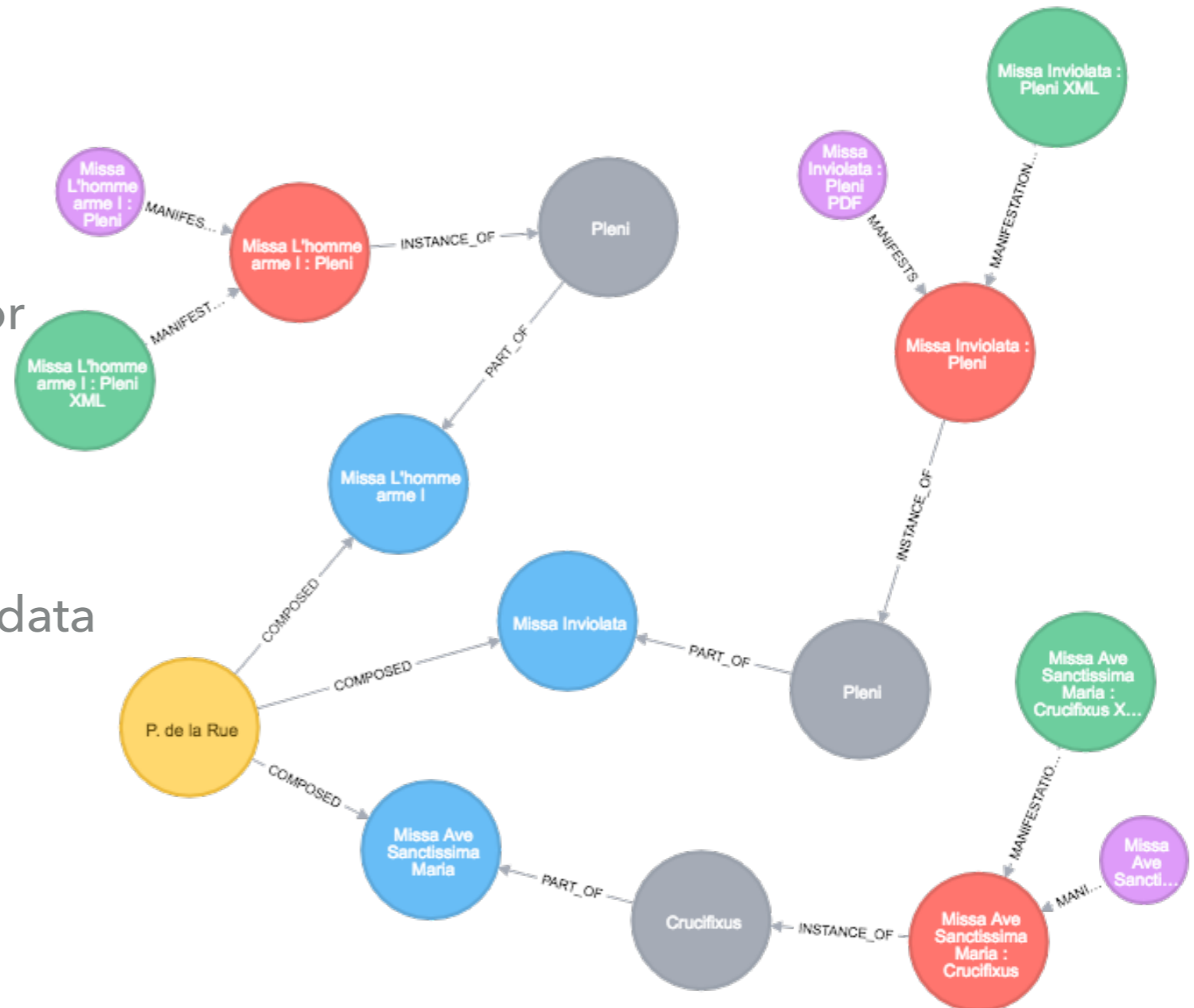
GRAPH DATABASE

► Pros

- Flexible
 - Works well with sparse or heterogeneous data
- Powerful queries
- Natural approach to linked data

► Cons

- New technology
- Harder to enforce integrity



CURRENT STATUS

- ▶ Experimenting and prototyping both kinds of databases
- ▶ Refining and expanding the data model
- ▶ Discussing the project with musicologists
 - ▶ Feedback and ideas
 - ▶ More sample queries and use cases
 - ▶ Soundness and correctness of the data model

THANK YOU!



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada 



McGill



Schulich School of Music
École de musique Schulich

DDMAL

DISTRIBUTED DIGITAL MUSIC
ARCHIVES & LIBRARIES LAB



Centre for Interdisciplinary Research
in Music Media and Technology

Fonds de recherche
Société et culture

Québec 