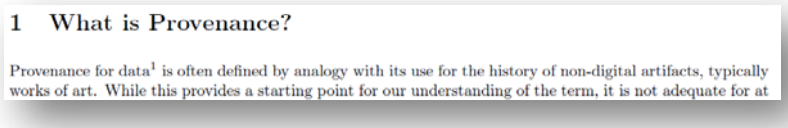
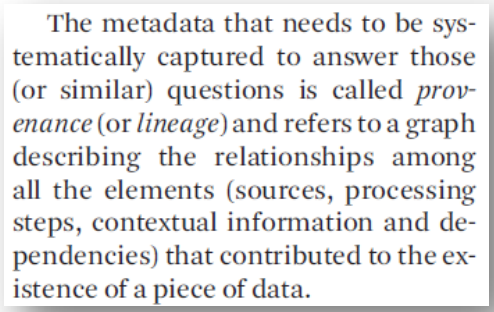
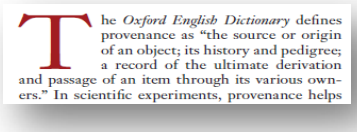
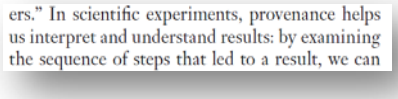
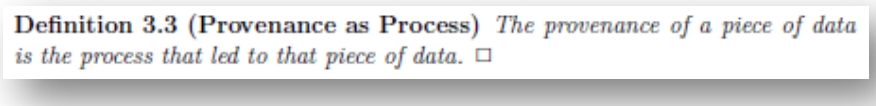
RCM of Provenance definitions

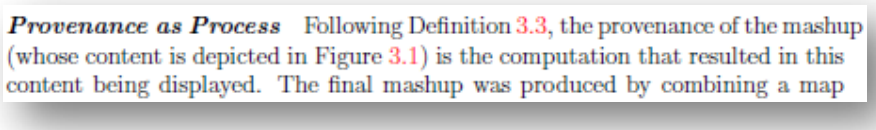
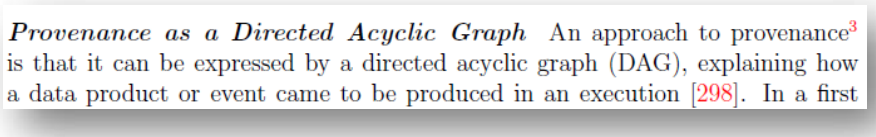
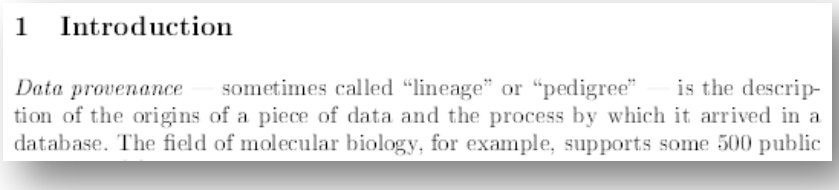
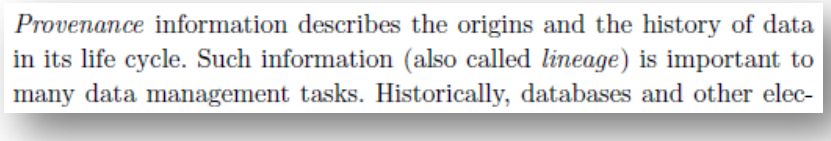
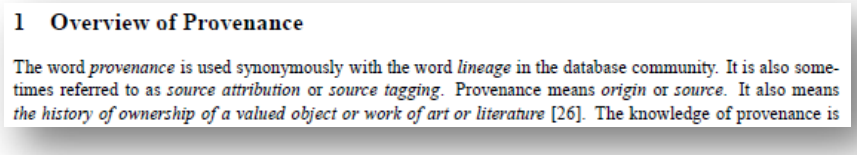
In this document, we provide the RCM created from the provenance definitions that we have extracted from the primary studies considered in the systematic review presented in [PeSR16].

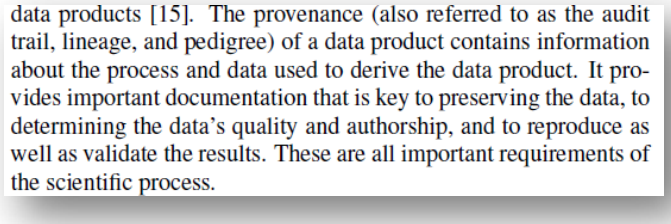
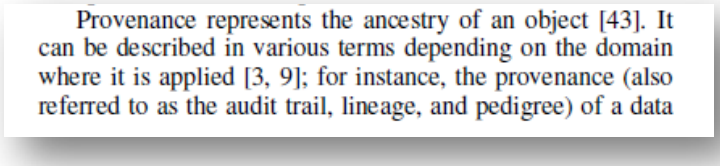
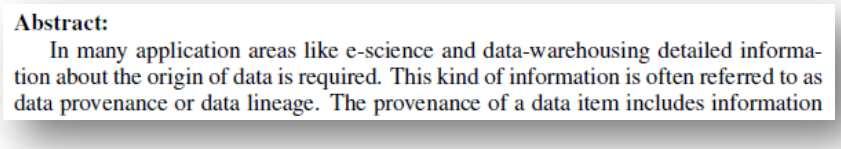
It is worth noting that for the sake of simplicity and readability, in this RCM we do not show the definitions that directly reference to other papers already considered to take part of this RCM. Thus, we have only taken into account *original* definitions.

Considering this, next we show the references from which we have obtained the provenance definitions, together with the extracts of such papers where the specific definition of *provenance* is provided. Later, we will present the RCM defined based on such definitions.

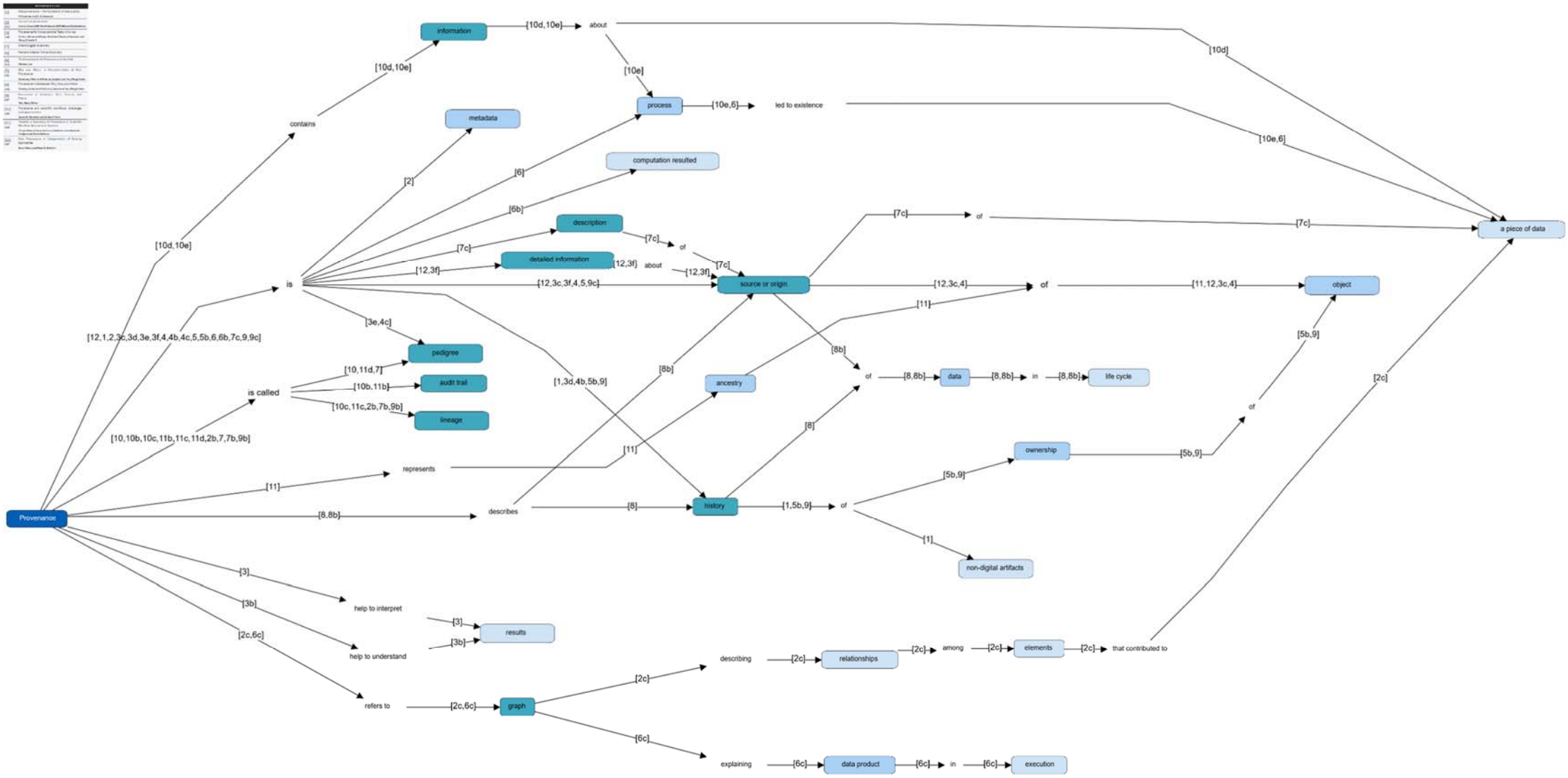
REFERENCES LIST	
[1]	Data provenance -- the foundation of data quality **P. Buneman and S. B. Davidson**
[2] 2014	A primer on provenance **Lucian Carata AND Sherif Akoush AND Nikilesh Balakrishnan**
[3] 2008	Provenance for Computational Tasks: A Survey **Freire, Juliana and Koop, David and Santos, Emanuele and Silva, Cl'audio T.**
[4]	Oxford English dictionary
[5]	Merriam-Webster Online Dictionary
[6] 2010	The Foundations for Provenance on the Web **Moreau, Luc**
[7] 2001	Why and Where: A Characterization of Data Provenance **Buneman, Peter and Khanna, Sanjeev and Tan, Wang Chiew**
[8] 2009	Provenance in Databases: Why, How, and Where **Cheney, James and Chiticariu, Laura and Tan, Wang-Chiew**
[9] 2007	Provenance in Databases: Past, Current, and Future **Tan, Wang Chiew**
[10] 2008	Provenance and scientific workflows: challenges and opportunities **Susan B. Davidson and Juliana Freire**
[11] 2009	Towards a Taxonomy of Provenance in Scientific Workflow Management Systems **S'ergio Manuel Serra da Cruz and Maria Luiza Machado Campos and Marta Mattoso**
[12] 2007	Data Provenance: A Categorization of Existing Approaches **Boris Glavic and Klaus R. Dittrich**

BUDA10 [1]	 1 What is Provenance? Provenance for data[1] is often defined by analogy with its use for the history of non-digital artifacts, typically works of art. While this provides a starting point for our understanding of the term, it is not adequate for at • Provenance is history of non-digital artifacts
CABB14_CARA14 [2]	 The metadata that needs to be systematically captured to answer those (or similar) questions is called *provenance* (or *lineage*) and refers to a graph describing the relationships among all the elements (sources, processing steps, contextual information and dependencies) that contributed to the existence of a piece of data. • Provenance is also called lineage • Provenance is the metadata • Provenance refers to a graph describing the relationships among all the elements that contributed to the existence of a piece of data.
FKSS08 [3]	 The *Oxford English Dictionary* defines provenance as "the source or origin of an object; its history and pedigree; a record of the ultimate derivation and passage of an item through its various owners." In scientific experiments, provenance helps  ers." In scientific experiments, provenance helps us interpret and understand results: by examining the sequence of steps that led to a result, we can • Provenance is source or origin of object • Provenance is pedigree • Provenance helps to interpret results • Provenance helps to understand results
Oxford English dictionary [4]	• Provenance is the source or origin of an object • Provenance is the history • Provenance is pedigree
Merriam-Webster Online Dictionary [5]	• Provenance is the origin or source • Provenance is the history of ownership of a object
More10 [6]	 **Definition 3.3 (Provenance as Process)** *The provenance of a piece of data is the process that led to that piece of data.* □ Provenance is process led to existence of a piece of data

	 ***Provenance as Process*** Following Definition 3.3, the provenance of the mashup (whose content is depicted in Figure 3.1) is the computation that resulted in this content being displayed. The final mashup was produced by combining a map • Provenance is computation resulted  ***Provenance as a Directed Acyclic Graph*** An approach to provenance[3] is that it can be expressed by a directed acyclic graph (DAG), explaining how a data product or event came to be produced in an execution [298]. In a first • Provenance refers to graph explaining data product in execution
BuKT01 [7]	 **1 Introduction** *Data provenance* — sometimes called "lineage" or "pedigree" — is the description of the origins of a piece of data and the process by which it arrived in a database. The field of molecular biology, for example, supports some 500 public • Provenance is also called lineage. • Provenance is also called pedigree. • Provenance is the description of the origins of a piece of data by which it arrived in a database.
ChCT09 [8]	 *Provenance* information describes the origins and the history of data in its life cycle. Such information (also called *lineage*) is important to many data management tasks. Historically, databases and other elec- • Provenance describes history of data in life cycle • Provenance describes source or origin of data in life cycle
Tan07 [9]	 **1 Overview of Provenance** The word *provenance* is used synonymously with the word *lineage* in the database community. It is also sometimes referred to as *source attribution* or *source tagging*. Provenance means *origin* or *source*. It also means *the history of ownership of a valued object or work of art or literature* [26]. The knowledge of provenance is • Provenance is history of ownership of object • Provenance is called lineage • Provenance is source or origin

DaFr08 [10]	 • Provenance is called pedigree • Provenance is called audit trail • Provenance is called lineage • Provenance contains information about a piece of data • Provenance contains information about process led to existence a piece of data
SeMM09 [11]	 • Provenance represents ancestry of object • Provenance is called audit trail • Provenance is called lineage • Provenance is called pedigree
GIDi07 [12]	 • Provenance is detailed information about source or origin of object

Taking these definitions into account, the RCM is depicted in the following figure (the original figure can be found here: http://www.unirioja.es/cu/beperev/rcm.png)


Provenance
is
is called
contains
represents
describes
refers to
information
metadata
process
led to existence
computation resulted
description
detailed information
source or origin
pedigree
audit trail
lineage
ancestry
history
data
life cycle
object
a piece of data
ownership
non-digital artifacts
help to interpret
help to understand
results
graph
describing
relationships
among
elements
that contributed to
explaining
data product
execution
about
of
in
[10d,10e]
[12,1,2,3c,3d,3e,3f,4,4b,4c,5,5b,6,6b,7c,9,9c]
[10,10b,10c,11b,11c,11d,2b,7,7b,9b]
[11]
[8,8b]
[3]
[3b]
[2c,6c]
[10d]
[10e]
[10e,6]
[2]
[6]
[6b]
[7c]
[12,3f]
[12,3c,3f,4,5,9c]
[3e,4c]
[10,11d,7]
[10b,11b]
[10c,11c,2b,7b,9b]
[1,3d,4b,5b,9]
[8]
[1,5b,9]
[12,3c,4]
[11,12,3c,4]
[5b,9]
[1]
[2c]
[6c]

References

[PeSR16] B. Pérez, C. Sáenz-Adán, J. Rubio, A systematic review of provenance systems, submitted for publication.