What is a datum and how it can be represented computationally

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Data Science (A.Y. 2022/2023)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

Any question about the previous lecture?

Two questions

What is a datum?

What is the difference between “datum” and “value”?

Value vs datum: an intuition

“raw” values

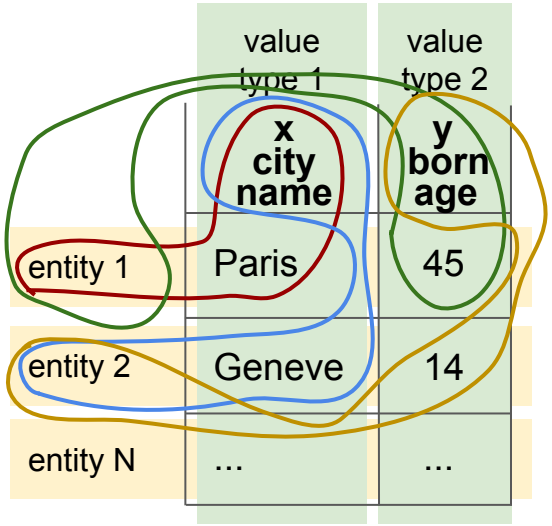
Paris

45

Geneve

14

data
(shown in tabular format)



Datum: a definition

A datum is a declarative statement subject-predicate-object

A datum either

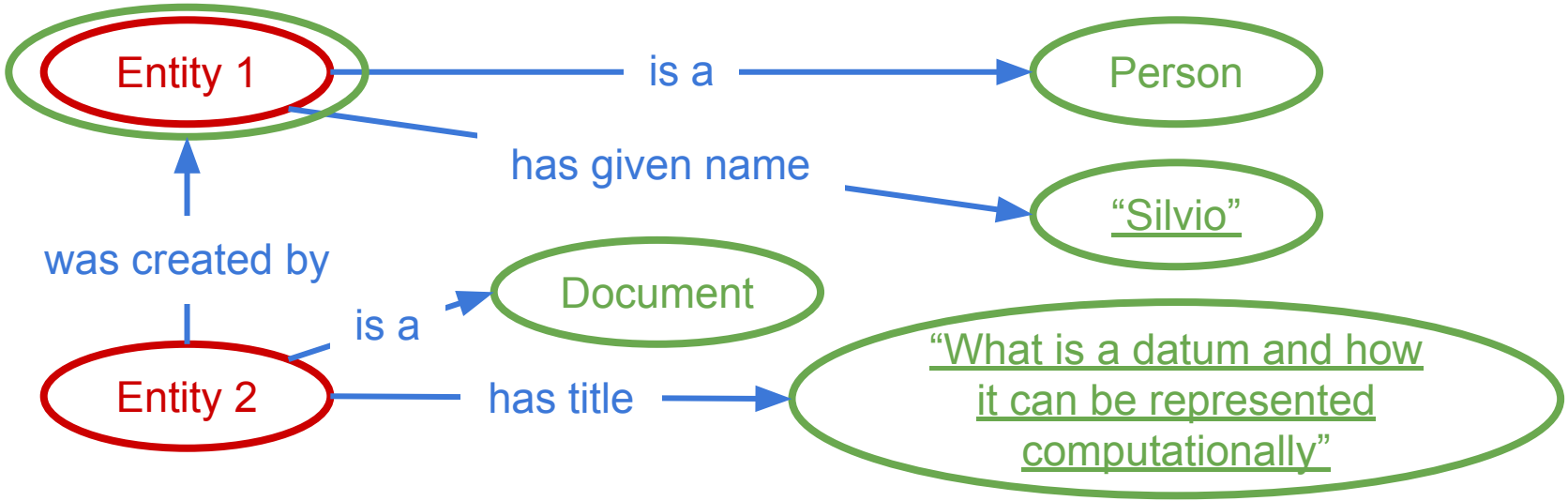
1. *attributes* (through the predicate) a value (the “object”) to an entity (the subject), or
2. *relates* (through the predicate) an entity (the subject) to another entity (the object)

Examples:

- Entity 1 is a Person
- Entity 1 has given name “Silvio”
- Entity 2 is a Document
- Entity 2 has title “What is a datum and how it can be represented computationally”
- Entity 2 was created by Entity 1

How to represent a datum

We can use a directed graph with labelled edges representing the predicate while the nodes identifies the subject and the object of the declarative statement



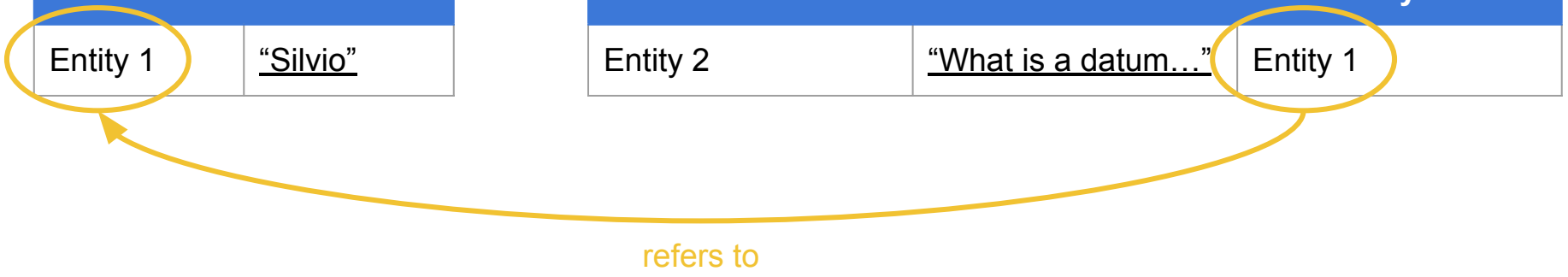
Same data, in tabular form

Person

Identifier	First name
Entity 1	“Silvio”

Document

Identifier	Title	Created by
Entity 2	“What is a datum…”	Entity 1



Attributes and relations

As anticipated, each datum can be of two different types depending on the kind of object:

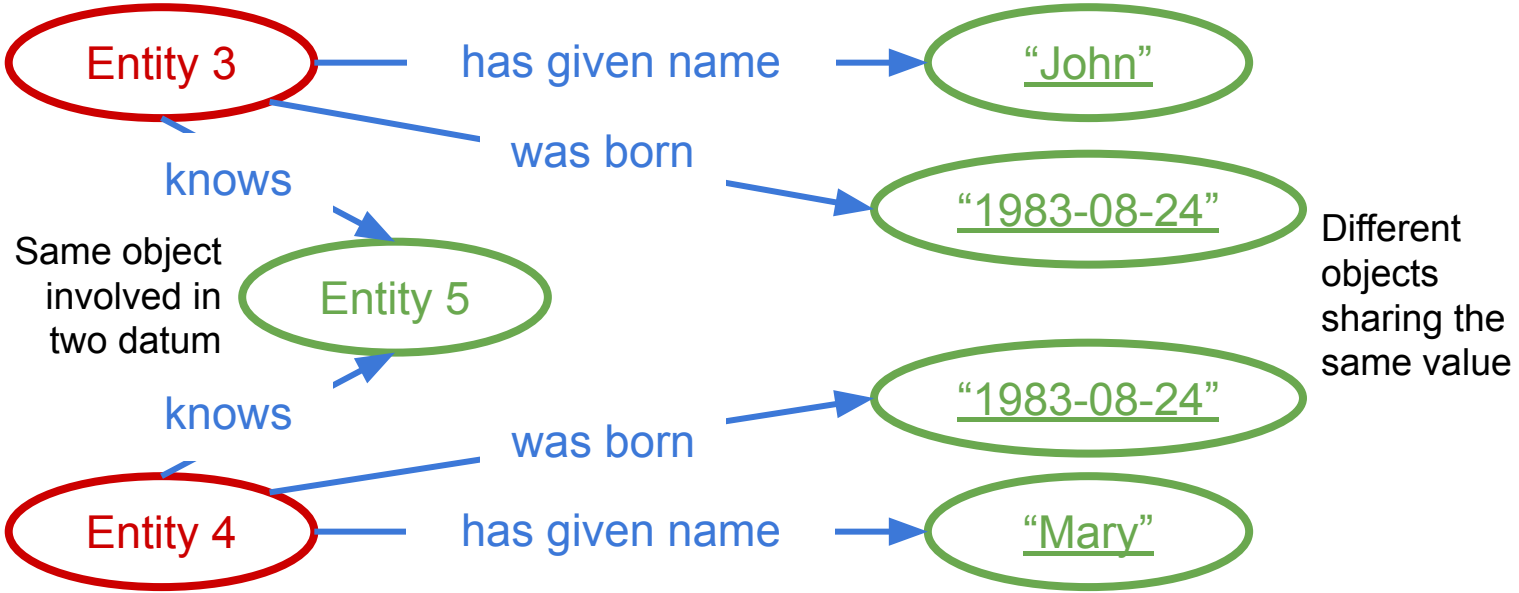
1. **Attribute**: associating a literal (a string, a number, etc., underlined) to a subject entity
   Entity 1 has given name “Silvio”

2. **Relation**: a labelled link between two entities
   Entity 2 was created by Entity 1

The object (i.e. the node in the graph) specified as an attribute of an entity cannot be reused in other data (i.e. involved by more than one edge)

However, it is possible to have two distinct objects of two attributes that share the same value

Attributes and relations: an example



Characteristics of attributes

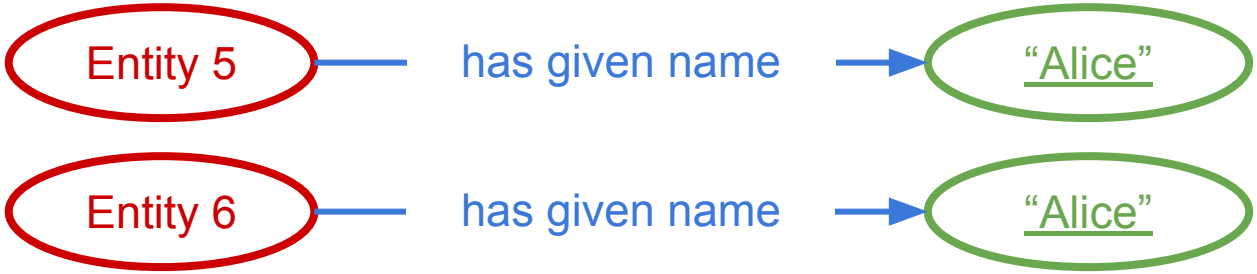
An attribute is intrinsically part of the entity to which it is associated

It is reasonable that the attribute of an entity can have the same value of an attribute of another entity, even when the same predicate is used (e.g. was born in the previous example)

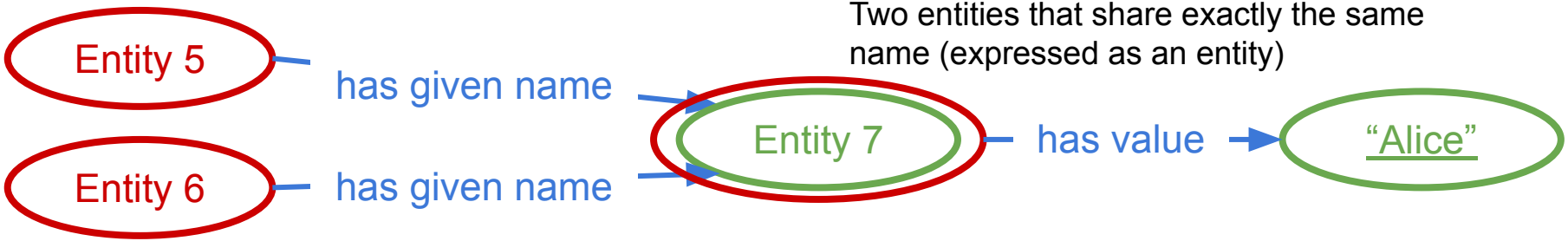
- There exists people born the same day
- There exists people having the same given name
- …

However, modifying the value of a certain attribute affects only the entity to which the attribute is associated

Alternatives to express data



Two entities that have the same name (i.e. same value)



Two entities that share exactly the same name (expressed as an entity)

It is important to express data correctly to avoid issues, but the *right* way depends on the context

Some hints

1. Each entity in a collection is uniquely identified by an identifier that can be explicitly defined (for instance, in databases, it is known as primary key) or implicit (as in the example in the table presented at the beginning of this lecture)
2. When an object of a statement refers to a thing that, potentially, can be involved in other data (as either a subject or an object), then it should be described as an entity and not as a literal
3. When defining a collection of data, it is necessary to use a predicate in the same way (i.e. either to associate attributes or to define relations)
4. Avoid to use two distinct predicates to convey the same semantics

Laboratory

Material:

- An A4 sheet of paper
- A pen

(alternatively, https://www.yworks.com/yed-live/ or similar)

Create a graph that describes the following scenario:

The article entitled “OpenCitations, an infrastructure organization for open scholarship” was published by the journal Quantitative Science Studies in 2020. The authors of this article, i.e. Silvio Peroni and David Shotton, also co-authored another conference paper entitled “The SPAR Ontologies”, that was published in the Proceedings of the Seventeenth International Semantic Web Conference, in 2018.

End

What is a datum and how it can be represented computationally

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Data Science (A.Y. 2022/2023)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA