ontotext

There is a common thread running through one of the main building blocks of the Semantic Web technology – the Uniform Resource Identifier (URI) – and the Greek myth about Odysseus who hides his name from a cyclops and manages to avoid being eaten by the one-eye giant. This common thread is the concept of naming.

Mr. Nobody, a Uniform Resource Identifier and the Challenge of Naming

The myth of Odysseus and the Cyclops has it that on his way back to Ithaca, the hero and his men ended up in the land of the Cyclopes – fierce, one-eyed giants. There, Odysseus was trapped in a cave of the Cyclops Polyphemus who ate two of the men the king of Ithaca was traveling with and was about to eat Odysseus himself. It is then, that Odysseus got Polyphemus drunk, told him that his name is "Nobody" and after that blinded the one-eyed giant.

Later, when the other Cyclopes asked Polyphemus: "Who blinded you?", he responded: "Nobody." With no proper way of naming the man who injured him, Polyphemus was unable to find him and seek revenge.

Before we embark on drawing our parallel about naming, let's briefly unpack the term URI. URI is the acronym for Uniform Resource Identifier and, as its name points, it is used for identifying a resource or a thing that needs to be described. The "uniform" part of the term touches the need for a common way of describing resources, which can be used across different contexts and systems, aiming for a global, unique naming scheme.

As we wrote in *The Truth about Triplestores* – a URI can be assigned to any resource be it physical or abstract:

> All data elements (objects, entities, concepts, relationships, attributes) are identified with URIs, allowing data from different sources to merge without collisions. All data is represented in triples (also referred to as "statements"), which are simple enough to allow for easy, correct transformation of data from any other representation without the loss of data.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

The idea of using URIs to identify ‘things’ (think for example authors, books, publishers, places, people, abstract terms, goods, articles, search queries, etc.) and the relationships between them is important for any data management endeavour aiming for a simple yet robust way to name data pieces in a consistent and interoperable way.

Now that we have a common definition of a URI, let’s put the myth about Odysseus and Polyphemus in the context of the Semantic Web technology and the increasingly important challenge enterprises face – that of data integration.

Imagine Cyclops was an algorithm (or a system) looking to talk to other systems (in our case his fellow Cyclopes). Given a false name, Polyphemus has no way of addressing the hero or, technically speaking, of properly identifying a resource. As a result, he is also unable to find him by “querying” other systems or, again in technical terms, by attempting to integrate information from them within his own “knowledge base”.

Without a unique, universal name in the memory of Polyphemus, no other operations (in our case finding and eating the legendary king of Ithaca) could be performed**.** Systems would not be able to talk to each other or they will be talking about one and the same thing, using different names

The above scenario is exactly what happens when systems use proprietary IDs when naming things. An idea worth revisiting, especially if at a certain point they will be part of a data integration plan.

Polyphemus’s failure to name Odysseus teaches us something seemingly simple: If you can name a thing, you can talk about it. And that simple idea is a sound basis for naming resources with a URI as part of the Linked Data paradigm. Without a proper (read universal and unique) naming scheme, managing identity is terribly hard. Integrating data across diverse systems, today’s world golden fleece, even more so.

The Nitty-Gritty of a Global Naming Scheme

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

> Practical systems need to access and mix objects which are part of different existing and proposed systems. Therefore, the concept of the universal set of all objects, and hence the universal set of names and addresses, in all namespaces, becomes important. This allows names in different spaces to be treated in a common way, even though names in different spaces have differing characteristics, as do the objects to which they refer.

Cit. The need for a universal syntax

Technically, a URI is a string of characters used to unambiguously identify a data object. As we saw, such data object can be any resource – from a column in a CVS table, to an event all the way to the location of a beacon on a store.

On the Semantic Web and within a Linked Data powered system (such as GraphDB), every resource has a URI. A URI can be the good old URL we all know or some other kind of a unique identifier. Unlike URLs (Web links), however, URIs do not necessarily enable access to

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

the resource they describe.

For example, the string *http://www.johndoesite.com/aboutme.htm*, if used as a URL, is expected to take us to a Web page of the site providing information about the site owner – the person John Doe. The same string however can simply identify that person on the Web irrespective of whether such a page exists or not. Thus URI schemes can be used not only for Web locations, but also for such diverse objects as telephone numbers, ISBN numbers and geographic locations.

There are different "schemes" such as FTP: URI, an ISBN: URI, an HTTP: URI. The HTTP URIs are the ones that tie data to the Web and are required for working with Linked Data. They use the hypertext transfer protocol and present a global agreement on how we refer to things on the Web, access, share and describe them and allows us to seamlessly link resources across systems. A benefit no enterprise can afford to overlook. [Dive deeper in The Benefits of URI Addressability!]

In an enterprise context, URIs (as a way of uniquely naming a thing in a standard manner so that a host of systems can refer to them and use and reuse them) can remove a lot of data integration hurdles. Share on X Using a set of universal conventions to refer to any resource – within the enterprise databases or outside them – is beneficial for:

- data exchange;
- resolving identities across disparate sources;
- data reuse;
- cost-effective data communication between differing systems.

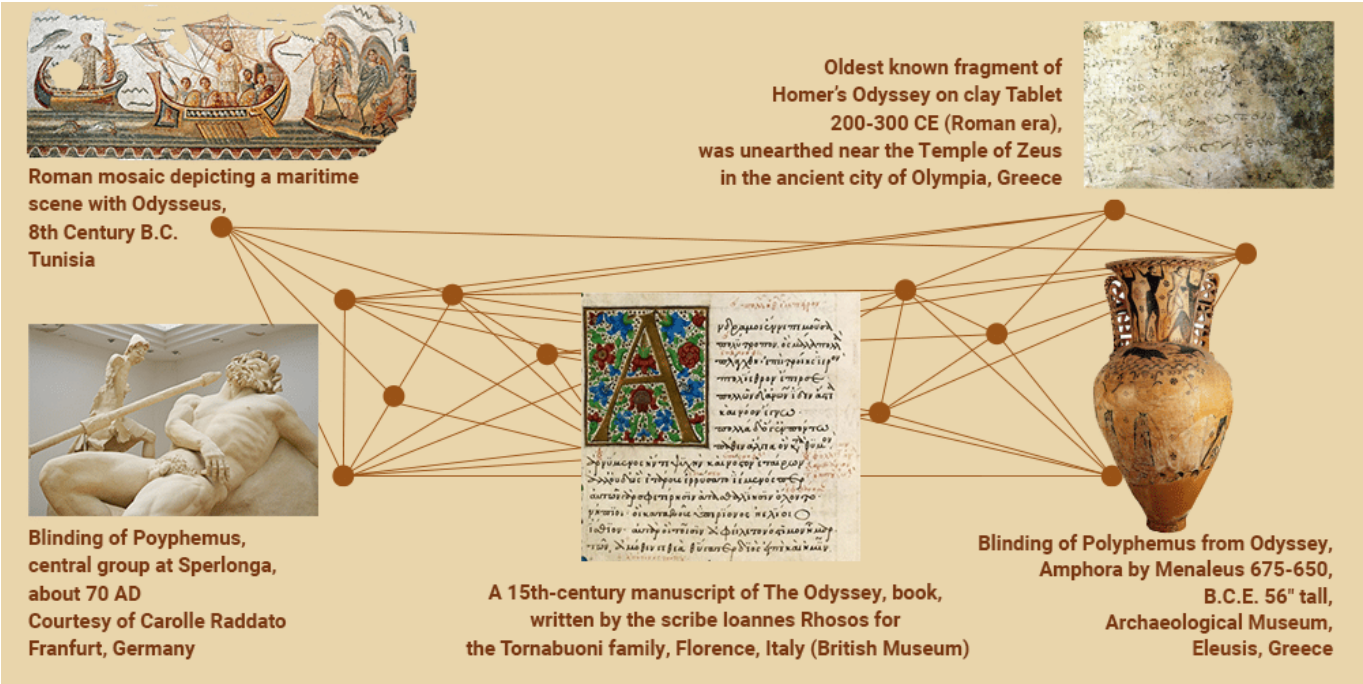
All the benefits of course come at a cost, which more or less comes down to several challenges:

- modeling your data in a universal way;
- conceptually defining the naming schemes you will be using;
- properly disambiguating resources.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

Epilogue: Don’t Store Mr. Nobody’s in Your Database



The reason we need a URI in the world of data-driven everything is one: data integration. And data integration, as complex as it might sometimes prove), starts with a simple concept: naming in a standard way, which allows machines to interoperate with data and its meaning across systems.

Arguably not all companies and applications need to bother with URIs. But those who are considerate about the the future of their knowledge and its interconnectedness, should be aware of the power a universal way of referring to data pieces brings to the well-integrated enterprise knowledge graph.

Interested in removing the Mr. Nobody entities in your database?

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com