

European Integration Biographies reference database (EIBIO)

Daniele Guido

Centre virtuel de la Connaissance sur l'Europe
daniele.guido@cvce.eu

Marten Düring

Centre virtuel de la Connaissance sur l'Europe
marten.during@cvce.eu

Lars Wieneke

Centre virtuel de la Connaissance sur l'Europe
lars.wieneke@cvce.eu

Catherine Jones

Centre virtuel de la Connaissance sur l'Europe
catherine.jones@cvce.eu

The European Integration Biographies reference database (EIBIO) is a structured repository for information about people that is currently under development at the CVCE. EIBIO is inspired by profile pages on LinkedIn or IMDB which combine structured, machine-readable information with unstructured and free text. This combined approach provides a high level of flexibility in describing a person's historical significance and also makes it easy to automatically create links between the person and document collections. In the near future EIBIO will be used to:

1. Publish high quality biographical information on cvce.eu and provide a unique identifier for players in European integration with associated information accessible via an API;
2. Link biographical information with the CVCE's collection of multimedia documents;
3. Explore relations between people and institutions in the European integration process;
4. Aggregate information from third-party repositories and align cross-references with other repositories and identifiers (VIAF, Wikipedia, DODIS, etc.).

In contrast to existing persistent person identifiers such as VIAF or DBpedia, EIBIO does not have an inherent selection bias and can be used for historical and living persons independently of their activities or historical significance. This means that every user of EIBIO can define his/her own definition of who should be included.

As part of their work on ePublications, researchers at the CVCE have researched and published biographical information for 240 people involved in the European integration process. Using EIBIO, all this information will soon be available to others by means of an API, which will make it possible to automatically retrieve a wealth of information on their professional positions.

Alongside biographical information, EIBIO will also display all texts, images and videos in the CVCE's collection in which a given person appears. This ensures the seamless integration of the positions held by a given person and documents which tell of their activities.

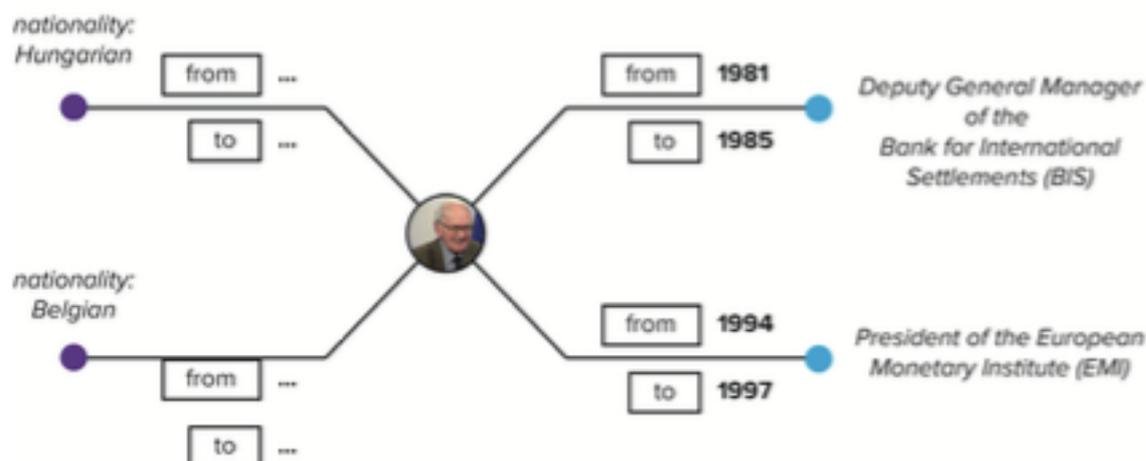


Figure 1: Schema: Representing biographical data in a Neo4j graph database.

EIBIO makes it very easy to run relation-based queries. One may, for example, want to find out which people connect the Bank for International Settlements and the Dutch Ministry of Finance, or which politicians have worked for the Bank for International Settlements in the past. Answers to these potentially insightful questions are hard to find using traditional research methods but are very easily obtained with EIBIO.

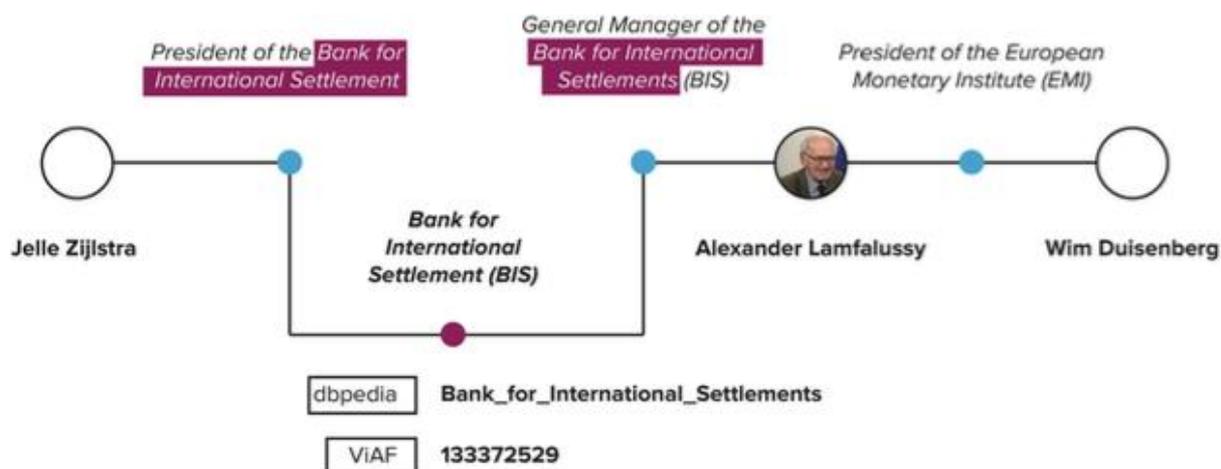


Figure 2: Schema: Constructing relations between biographies.

A person's biography can be enriched by using VIAF and DBpedia unique identifiers as a means of tapping into other repositories and aligning references between repositories. VIAF, for example,

stores publication records for individuals compiled from library catalogues around the world. In turn, these identifiers make it easy for others to retrieve data on a given person from EIBIO.

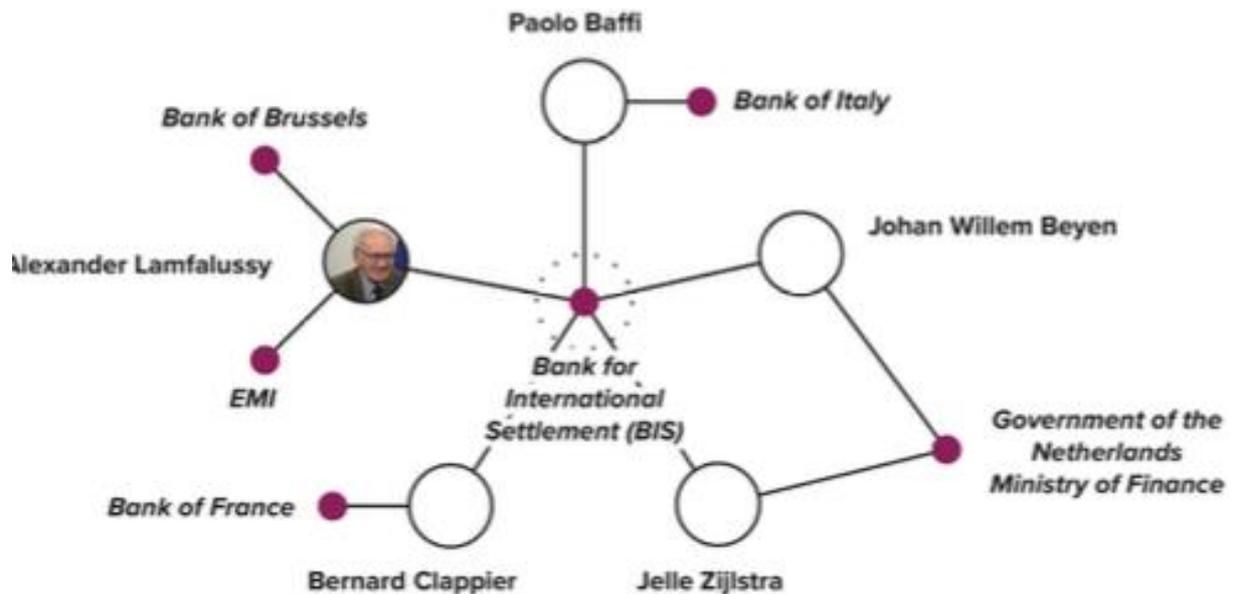


Figure 3: Schema: Retrieving relations between biographies.

All biographical data is stored as a set of relations and attributes in a Neo4j graph database. In contrast to traditional relational databases, graph databases are more flexible regarding changes to the underlying data model and perform better at graph-related queries. Consider the following example: A person node 'Alexandre Lamfalussy' has a temporary relation with the Bank for International Settlements which has the value 'Deputy General Manager'. We can easily add other relations with other institutions and different types of relation such as 'studied at' or attributes such as 'nationality'. Graph databases make it particularly easy to add new types of relation, something that is very hard to handle with relational databases. We used YAGO to automatically detect institution names in existing biographies and verified the output manually to ensure high quality and consistency.