Helio

A Link Data Service to interact with heterogeneous data sources

cimmino@fi.upm.es

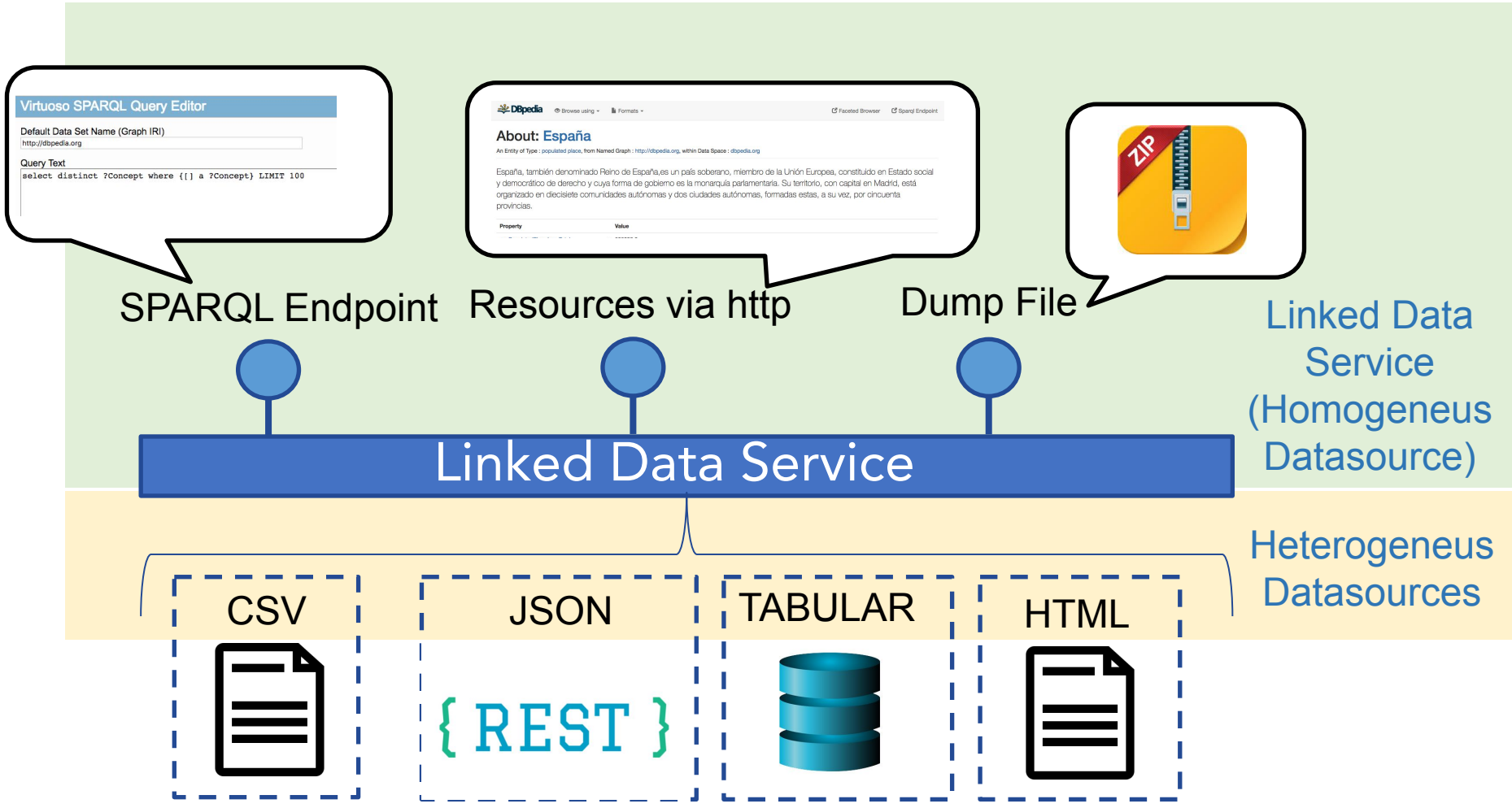
Andrea Cimmino
Ontology Engineering Group
Universidad Politécnica de Madrid, Spain

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 688467

European Commission | Horizon 2020 European Union funding for Research & Innovation

CC BY SA

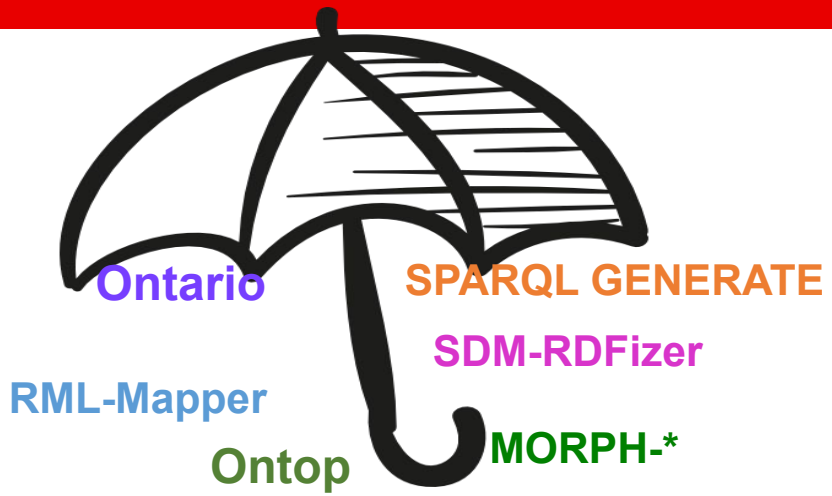
Helio overview



Common challenges

- Cope with the different data sources
  - Access methods, i.e., API, file, DB
  - Formats, i.e., JSON, CSV, Tabular, HTML
  - Security, e.g., APIs with Oath, files with passwords
- Clean data
  - Lowercase, missing values
- Link data
  - Interlink data from different sources
- Publish as an RDF view the data
  - Enable a SPARQL endpoint
  - Allow resource access
  - Dump generation
- Others
  - Real-time data
  - API restrictions in the number of calls per day

Helio unique challenges

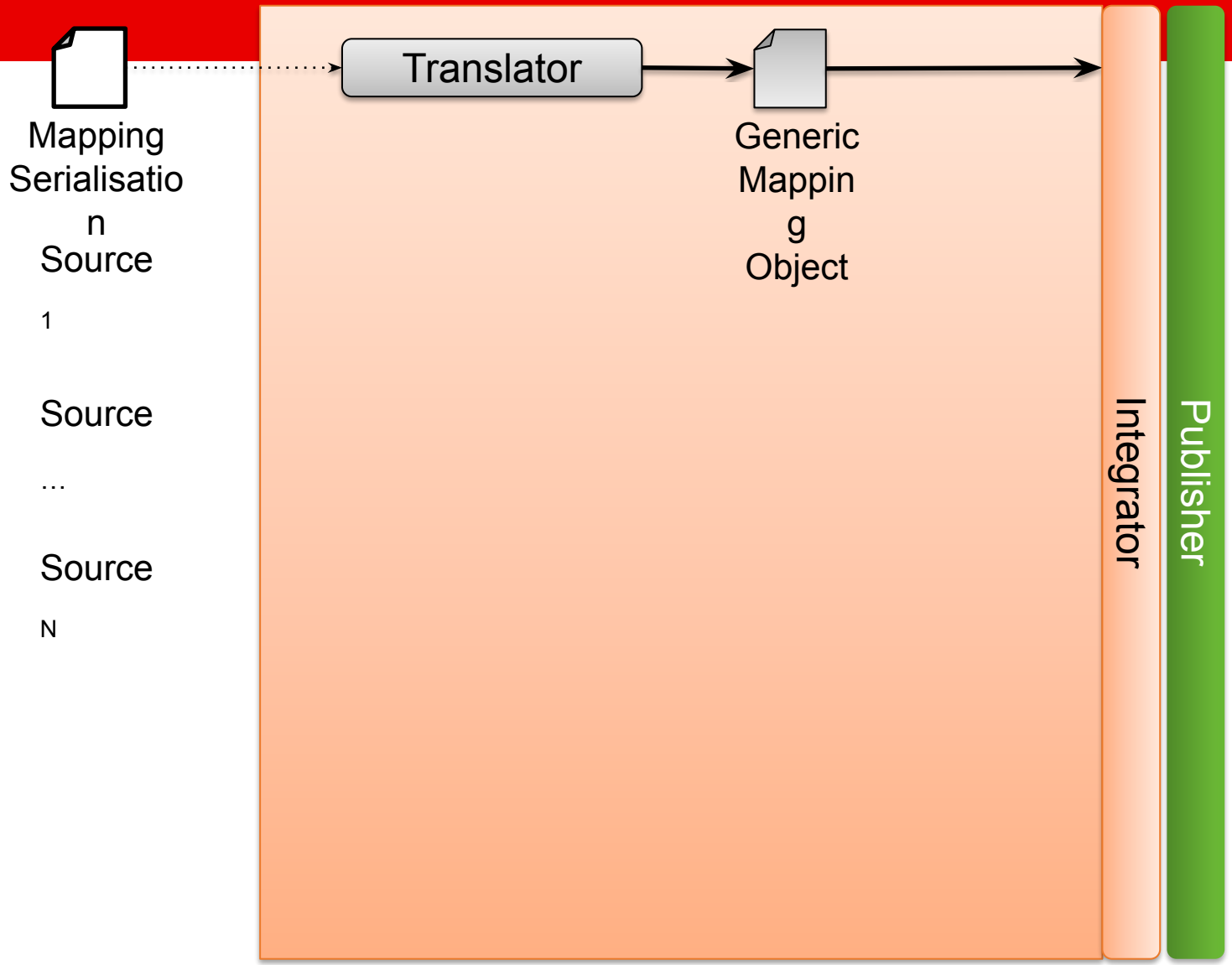


Integrating existing technologies

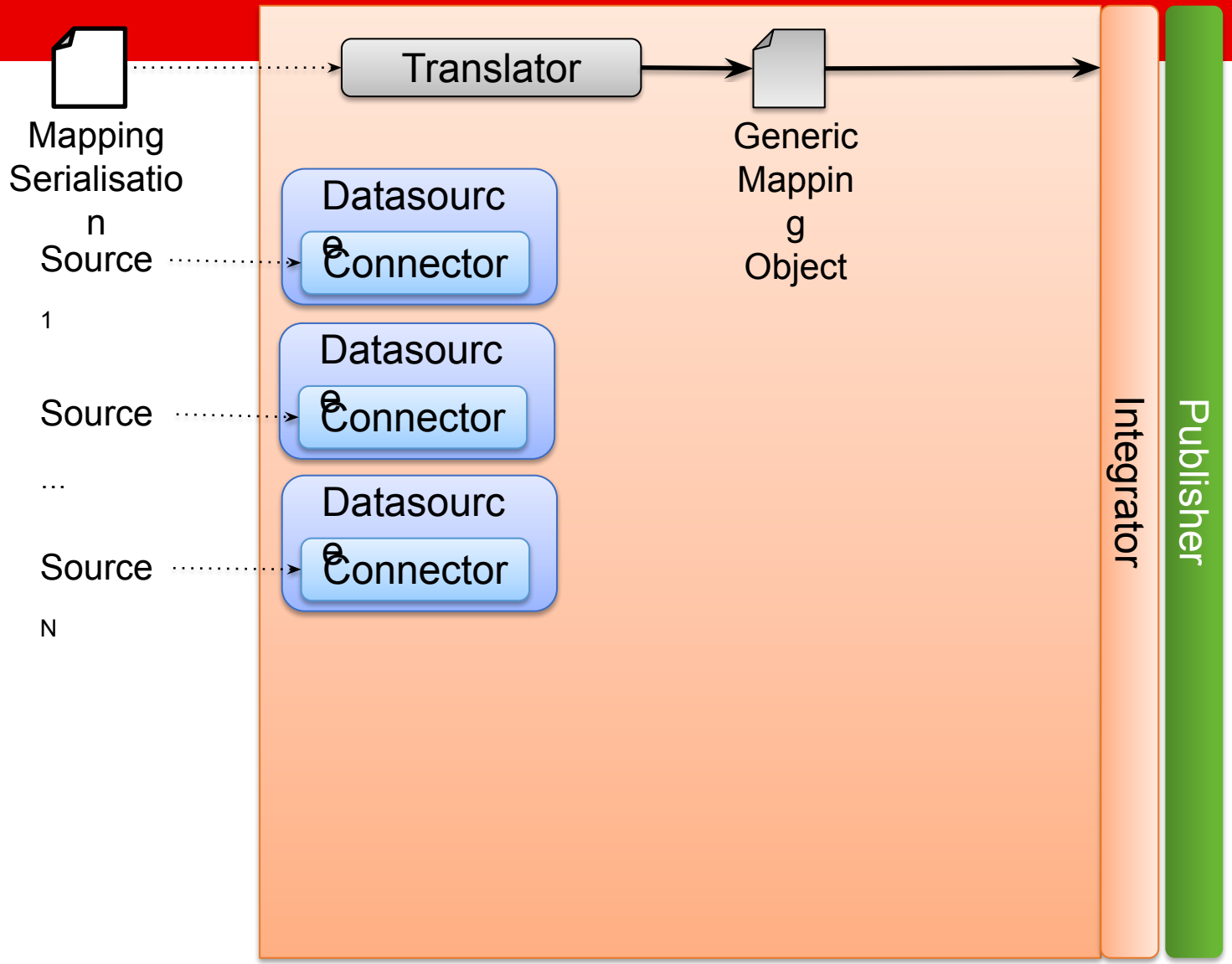
Interact with the sources by transforming from RDF the data to their heterogeneous format+model

Have control over the quality of the published data

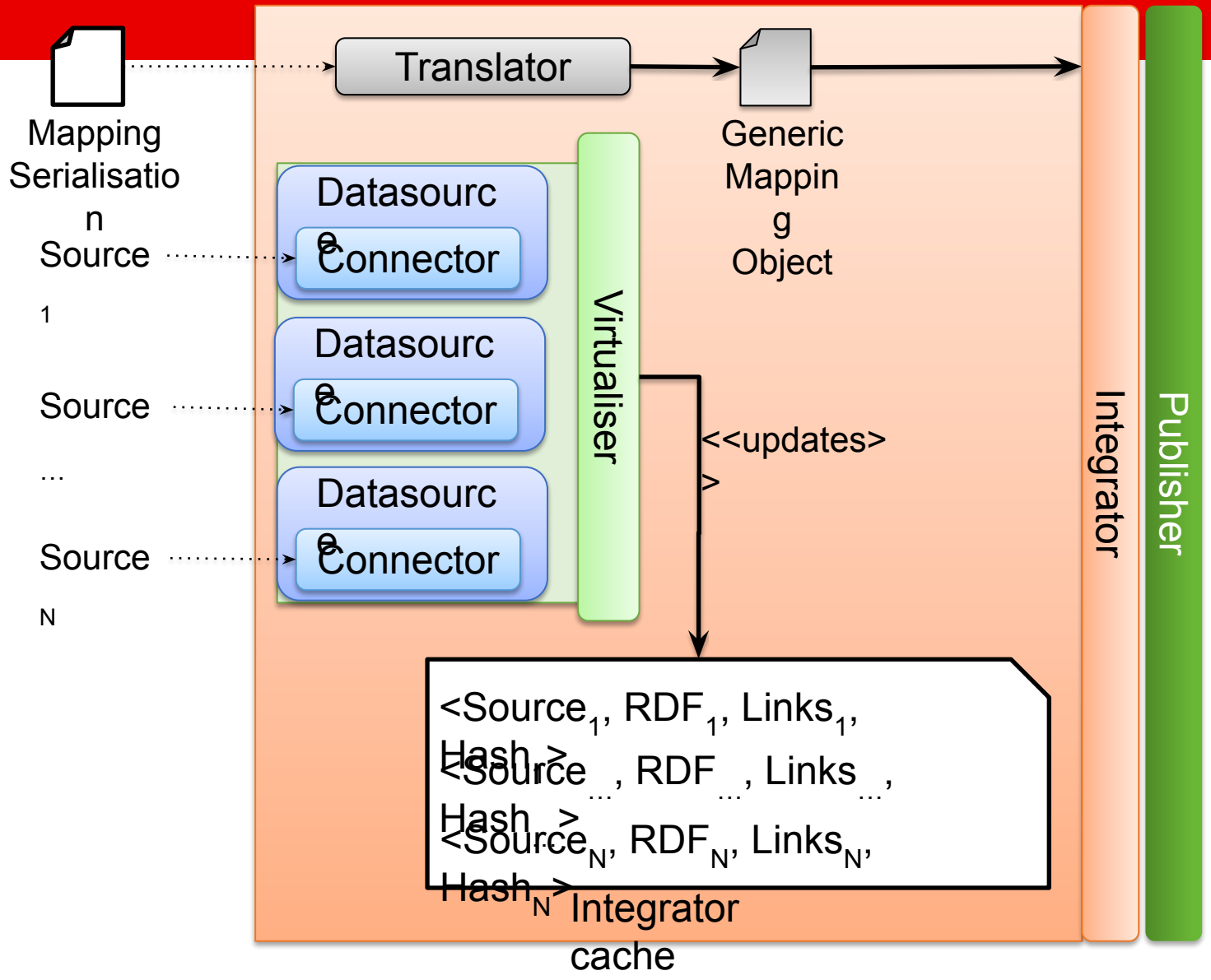
Publishing RDF data



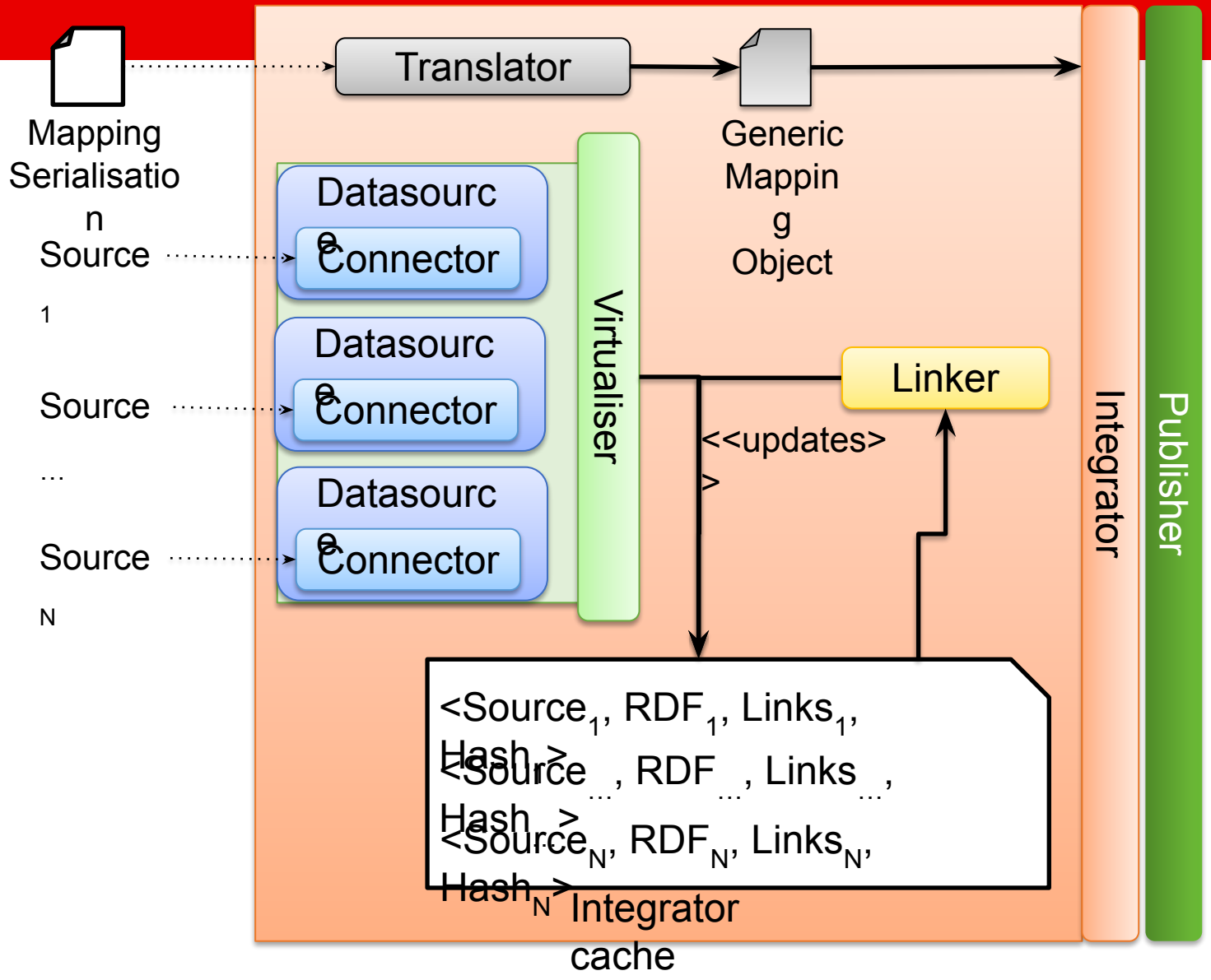
Publishing RDF data



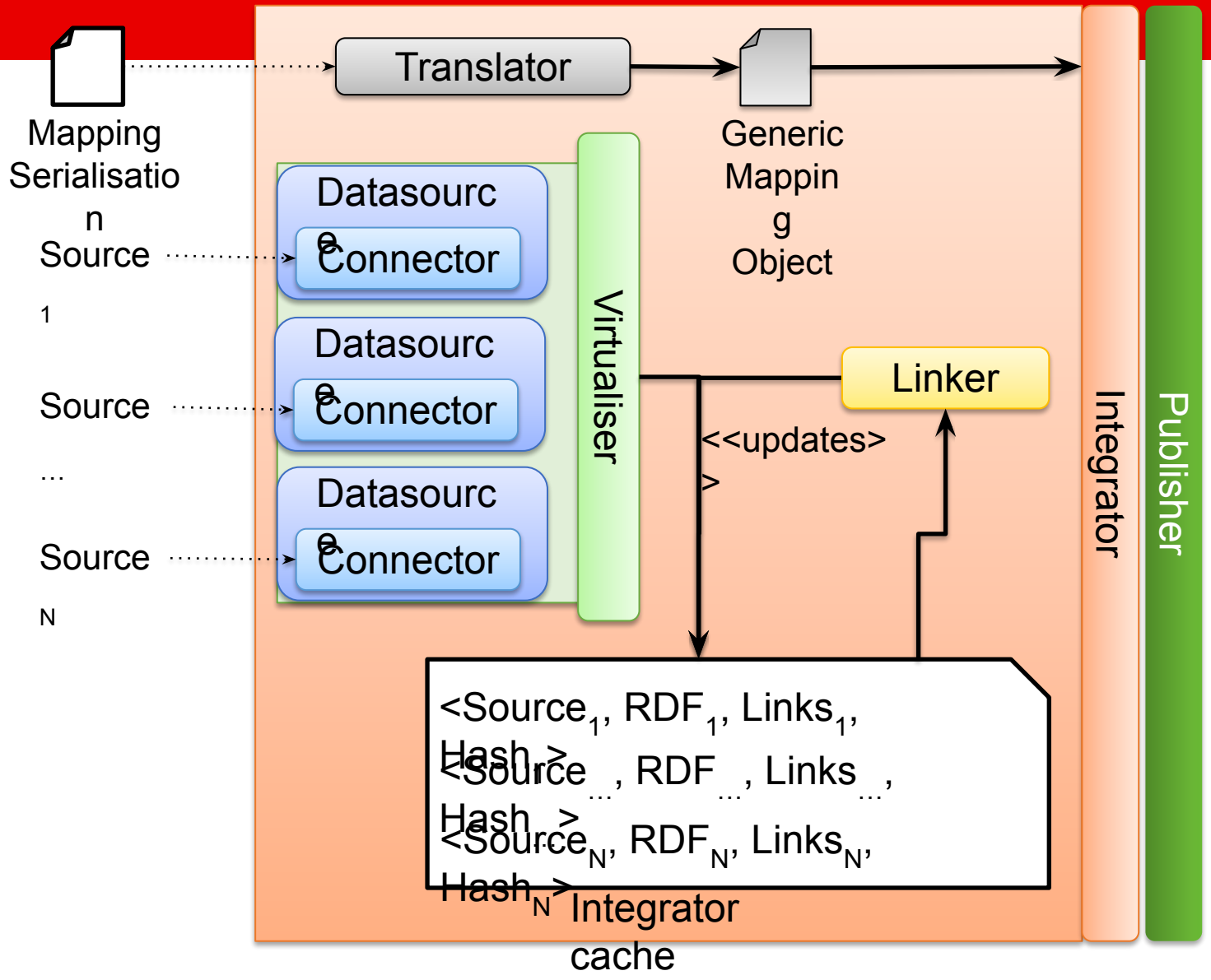
Publishing RDF data



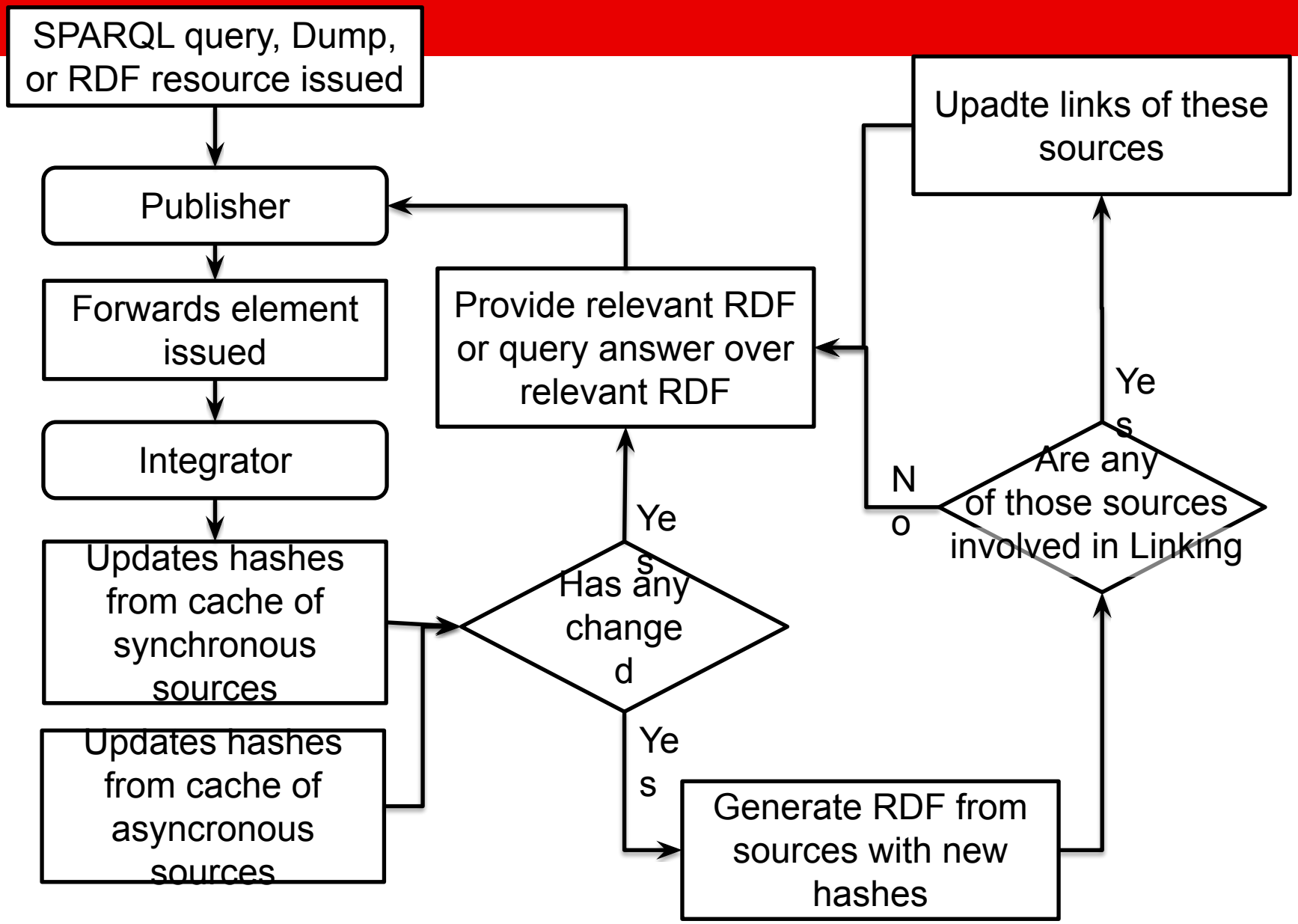
Publishing RDF data



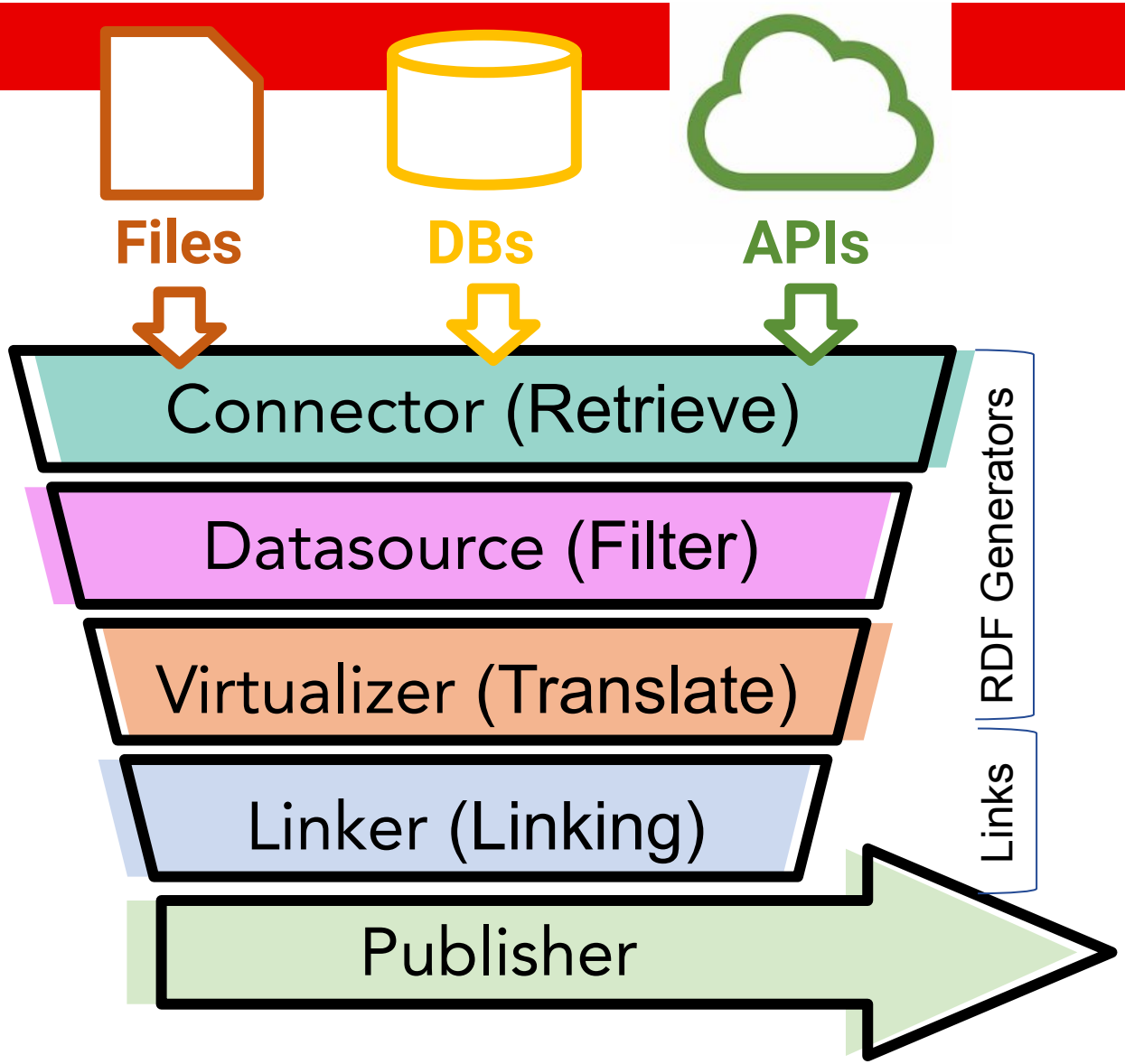
Publishing RDF data

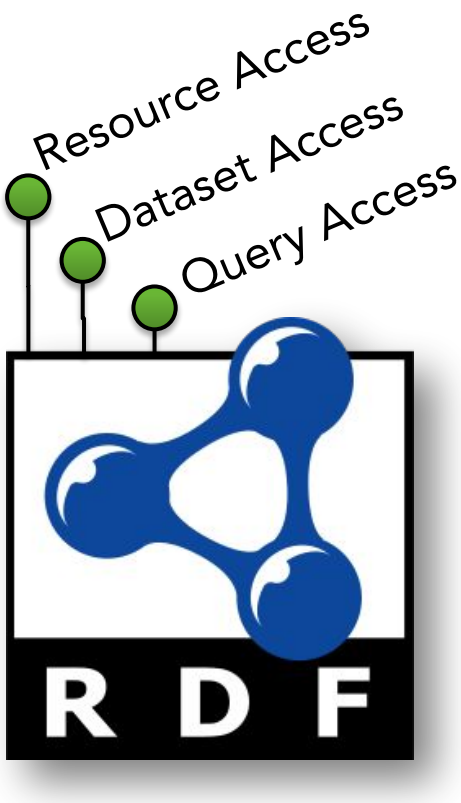


Workflow


SPARQL query, Dump, or RDF resource issued
Publisher
Forwards element issued
Integrator
Updates hashes from cache of synchronous sources
Updates hashes from cache of asyncronous sources
Has any change d
Yes
Provide relevant RDF or query answer over relevant RDF
Upadte links of these sources
Are any of those sources involved in Linking
Yes
No
Yes
Generate RDF from sources with new hashes

Helio Internal Pipeline





Helio core implementation status

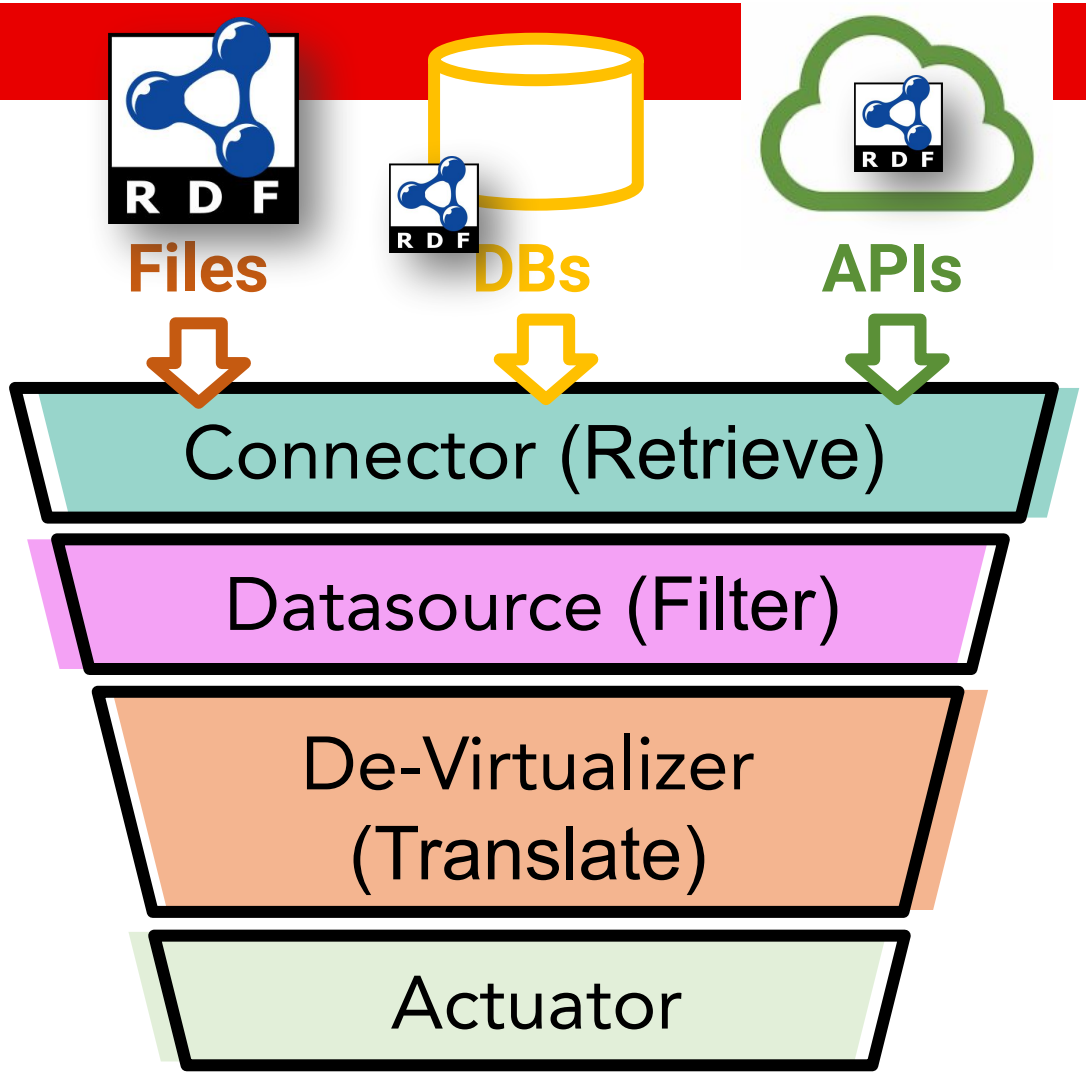
Specification	Elements	Implemented
Helio Specification	Connector	FederatedSparqlConnector, GetConnector, LocalFileConnector, TwitterConnector, URLConnector,
	Datasource	CsvDatasource, HtmlDatasource, JsonDatasource, RDFDatasource, TextDatasource, XmlDatasource,
	Translator	Helio Virtualizer
	Linking	Helio Linker
RML* (proof of concept)	Connector	LocalFileConnector
	Datasource	JsonDatasource
	Translator	Helio Virtualizer
	Linking	Helio Linker

Extension points

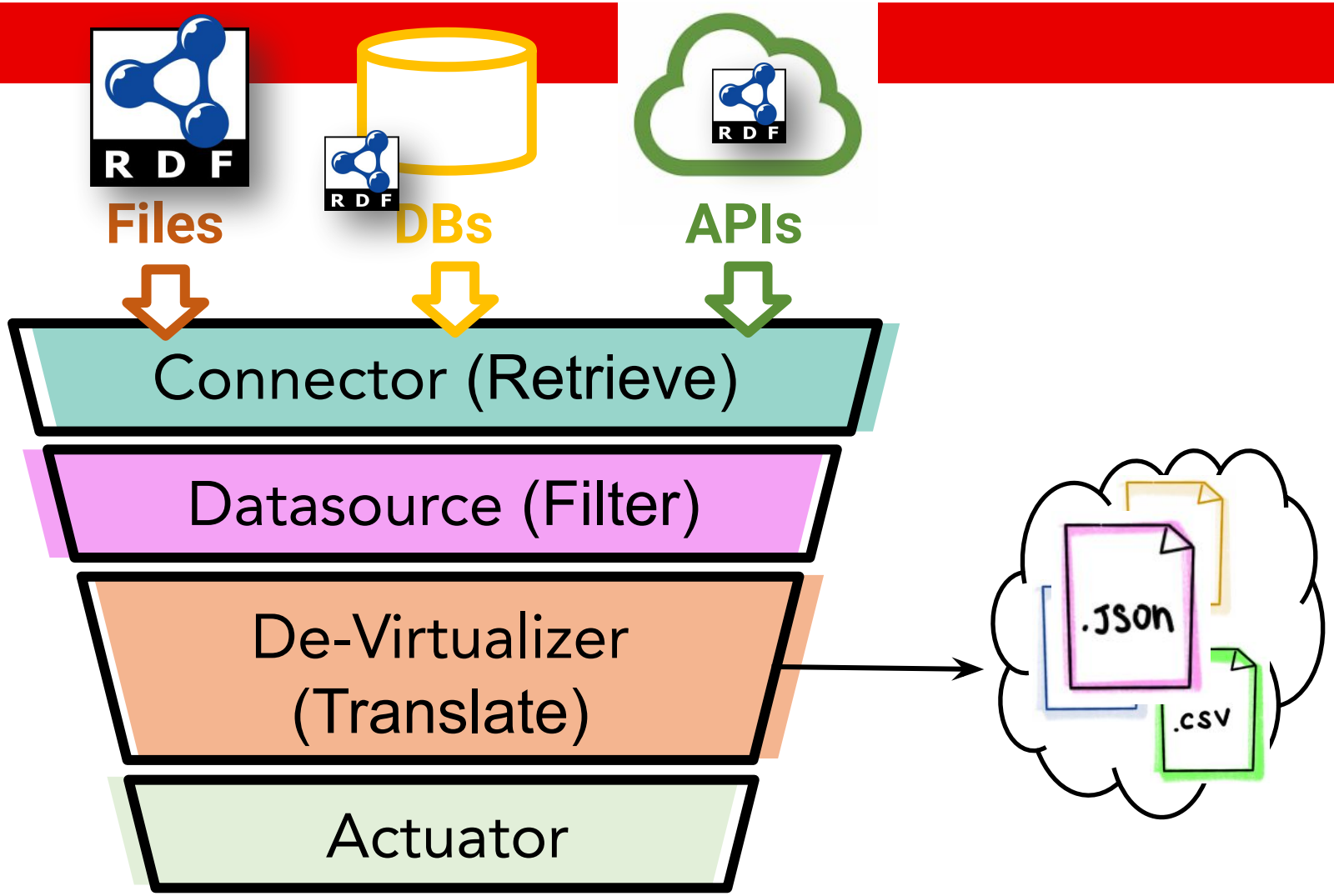
- Extend the original framework
  - Allows Helio to be setup using other mappings
    - WoT, RML, …
  - Extend formats supported by Helio (Datasource)
  - Extend sources supported by Helio (Connector)
  - Extend functions

- Plugin system
  - Implementing external code, and compiling a jar, Helio will automatically be able to use such plugin
  - Extend sources supported by Helio (Connector)
    - Third-party systems can be integrated with this mechanism
  - https://github.com/oeg-upm/helio-plugins

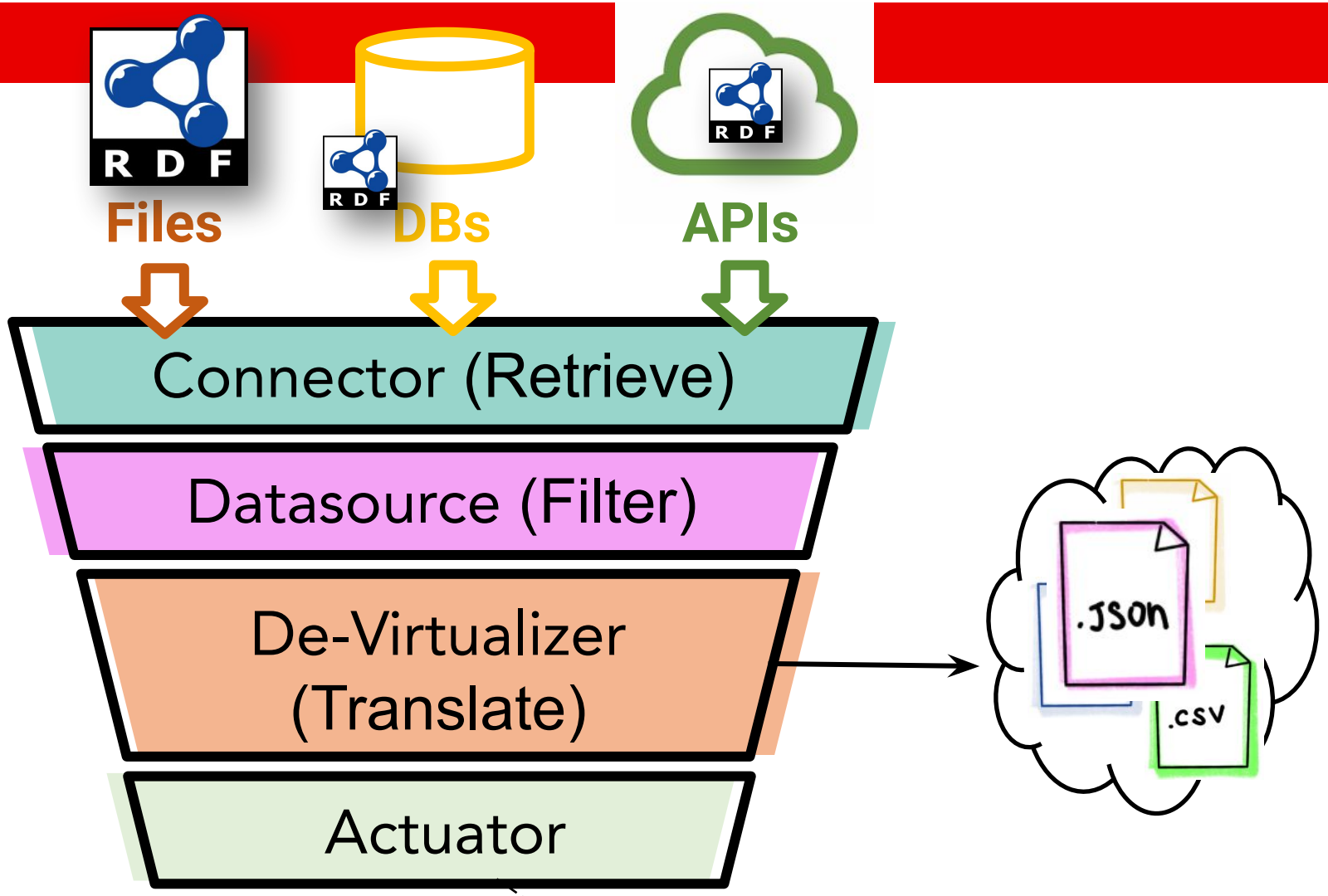
Writting from RDF to heterogeneus fotmats



Writting from RDF to heterogeneus fotmats



Writting from RDF to heterogeneus fotmats



- Store generated data in file
- Combine data with static data and publish
- Make a POST with data
- …

Writting from RDF to heterogeneus fotmats

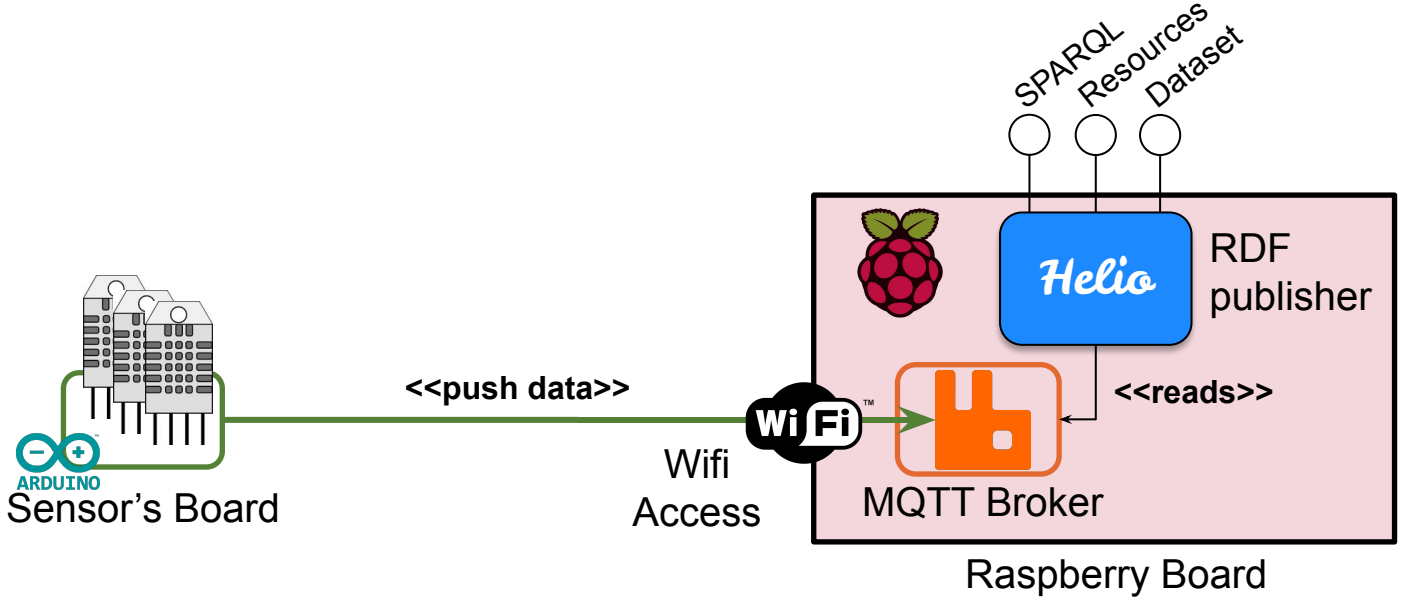
- Requires similar mappings of generation
- This functionality is triggered when:
  - POST, PUT,DELETE or UPDATE
  - SPARQL INSERT/DELTE/UPDATE queries

Applications

- VICINITY project
  - Gateway API distributed query client library (2018-2019)
    - https://github.com/vicinityh2020/vicinity-open-gateway-api-distributed-query-client
    - Cimmino et al. VICINITY: IoT Semantic Interoperability Based on the Web of Things (2019). In 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS) IEEE.
  - eWoT (2019-2020)
    - https://github.com/oeg-upm/eWoT
    - Cimmino, A., Poveda-Villalón, M., & García-Castro, R. (2020). eWoT: A Semantic Interoperability Approach for Heterogeneous IoT Ecosystems Based on the Web of Things. Sensors, 20(3), 822.
    - Helio plugin: https://github.com/oeg-upm/helio-plugins/tree/master/wot-translator
- DELTA project
  - Astrea (2019-2020)
    - https://astrea.helio.linkeddata.es
    - Cimmino, A., Fernández-Izquierdo, Alba., García-Castro, Raúl. Astrea: automatic generation of SHACL shapes from ontologies. ESWC 2020.
  - Common Information Model (2019-2020)
    - https://github.com/oeg-upm/DeltaCimApp
    - *Paper pending*

- Ongoing papers:
  - DELTA
    - Cimmino et al. Semantic interoperability for DR schemes employing the SGAM framework. SEST 2020 **(DELTA)**
  - Cimmino A. And García-Castro. HELIO: a Link Data Service for interacting with heterogeneus sources of information.
  - Cano-Benit,o Juan et al. Benchmarking the efficiency of RDF-based access for blockchain environments. SEKE 2020 **(DELTA)**

- TFM
  - Blockchain (2018-2019)
    - https://github.com/oeg-upm/helio-plugins/tree/master/ethereum-connector
- TFG
  - Taxons (2018-2019)
    - https://github.com/oeg-upm/helio-plugins/tree/master/rml-connector
  - Semantic Sensors deployment (2019-2020)
    - https://github.com/oeg-upm/helio-plugins/tree/master/mqtt-connector



Plugins: Integrating other technologies

RMLConnector

ValkyrConnector-IE

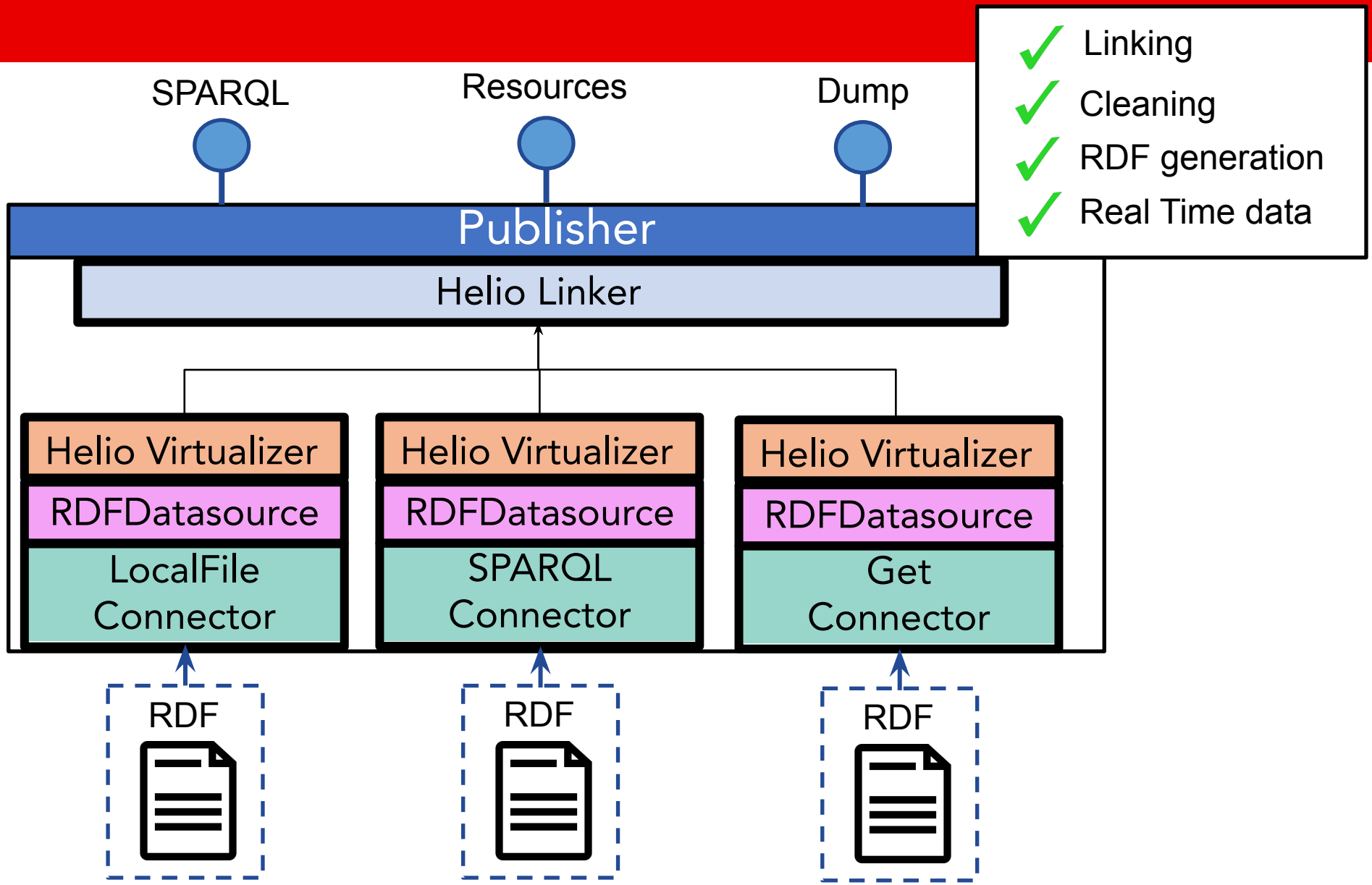
Dbpedia Spotlight

Conclusions

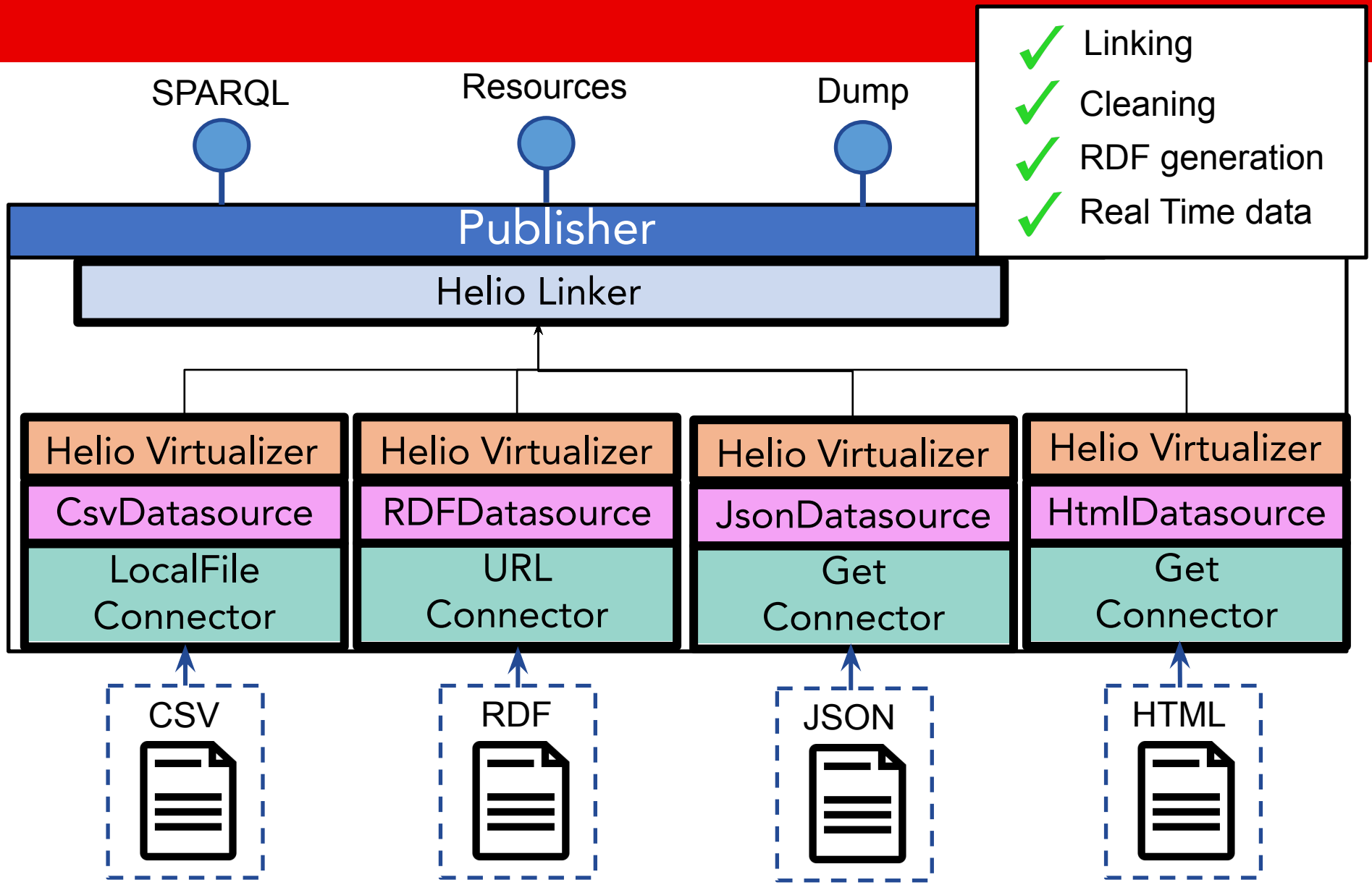
- Helio
  - Allows the generation of RDF form heterogeneus datas sources
  - Allows the iteraction with those sources by genrating heterogeneus data from RDF
  - Enables a Linked Data Service for third-party systems to interact with the heterogeneuos data sources
- Extensibility
  - By implementing an interface (Conector, Datasource, Translator or Function) from the core project
  - By implementing plugins
- Scope
  - Has been tested in two European projects (VICINITY & DELTA)
  - Has been used, and extended in 3 master thesis
  - Has received colaboration from external users

Thanks to all

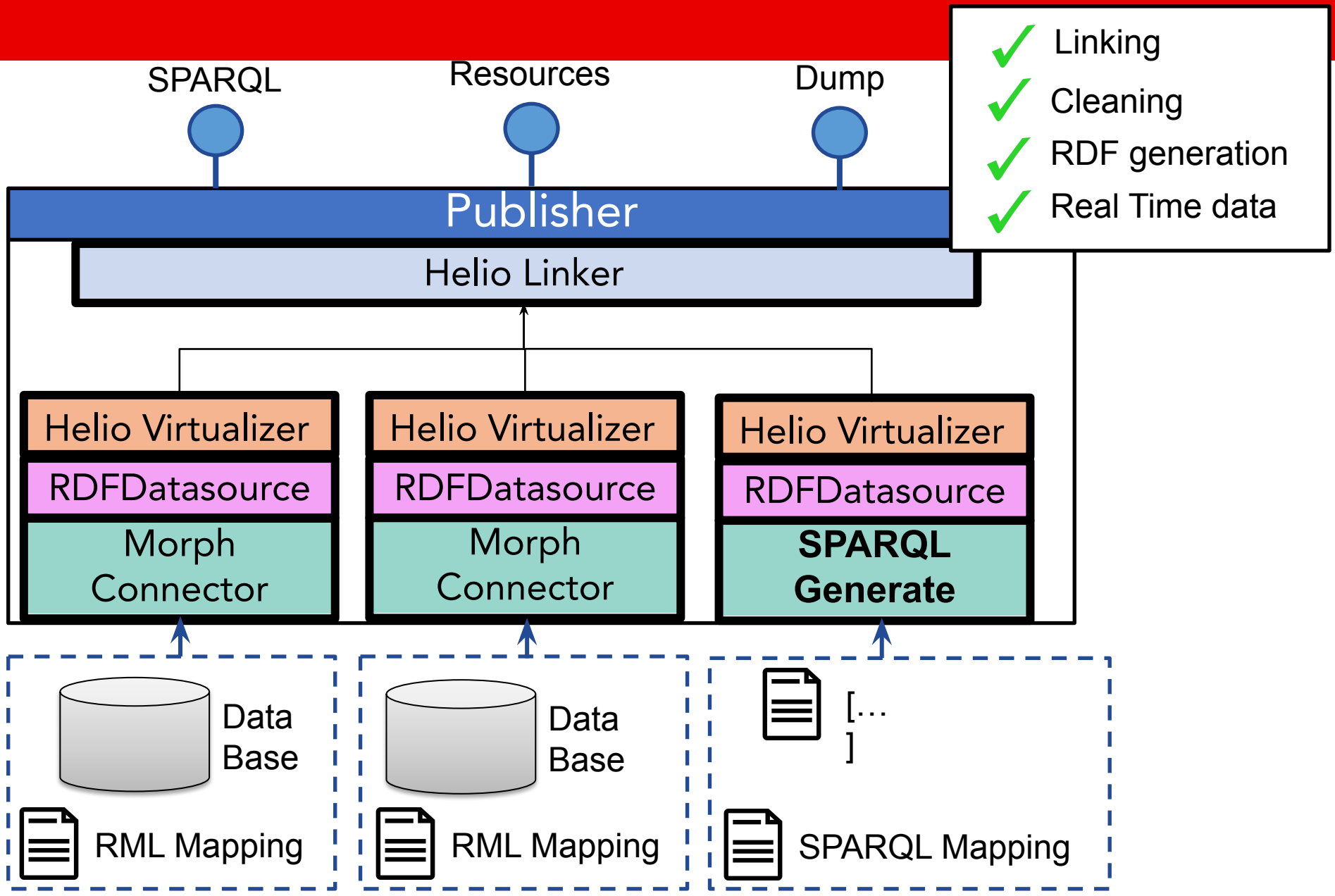
Helio Scenario 1: Refining RDF

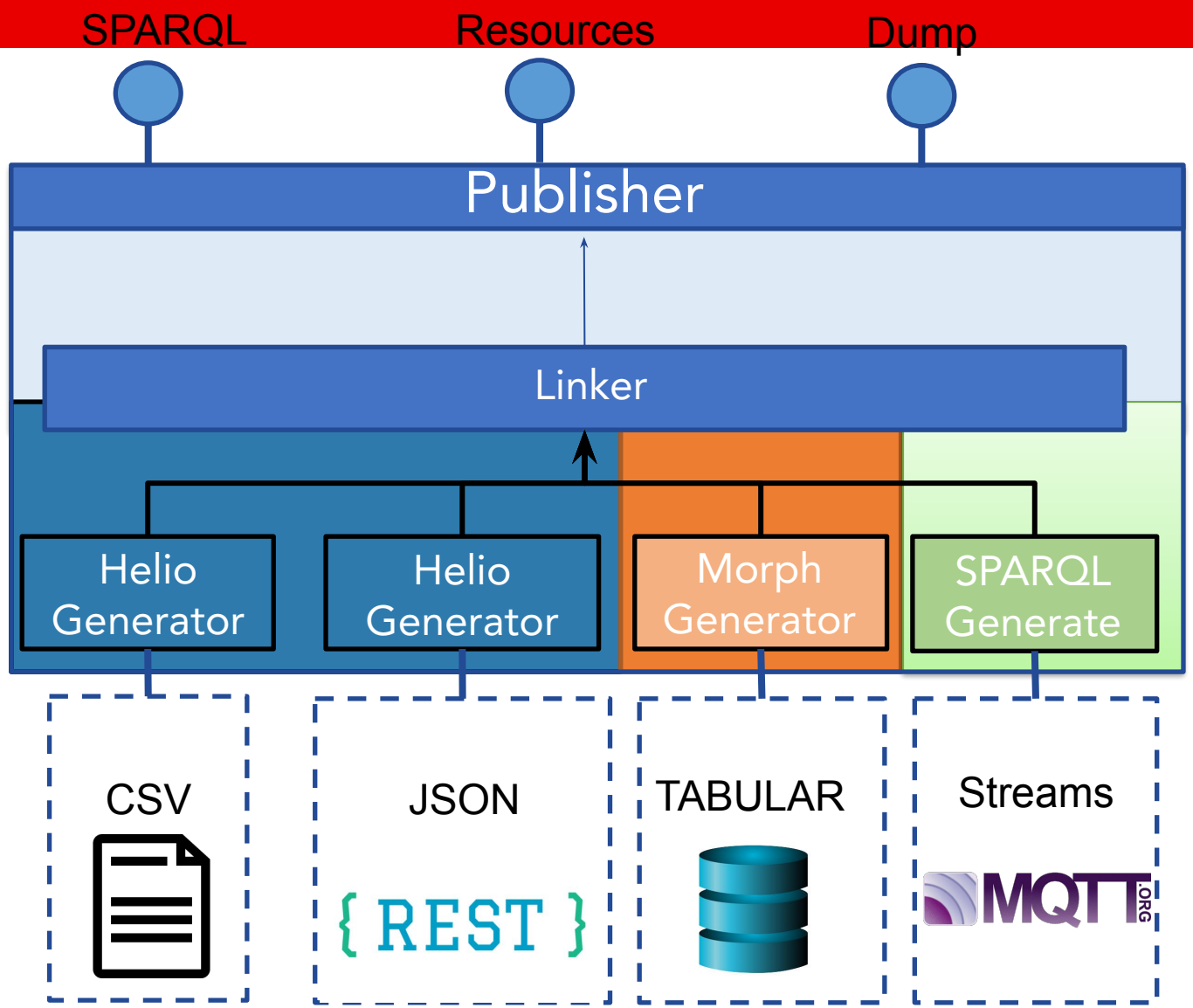


Helio Scenario 2: Data Integration



Helio scenario 3: Integrating third-part RDF generators



2
He
Helio

One Helio to gather them all