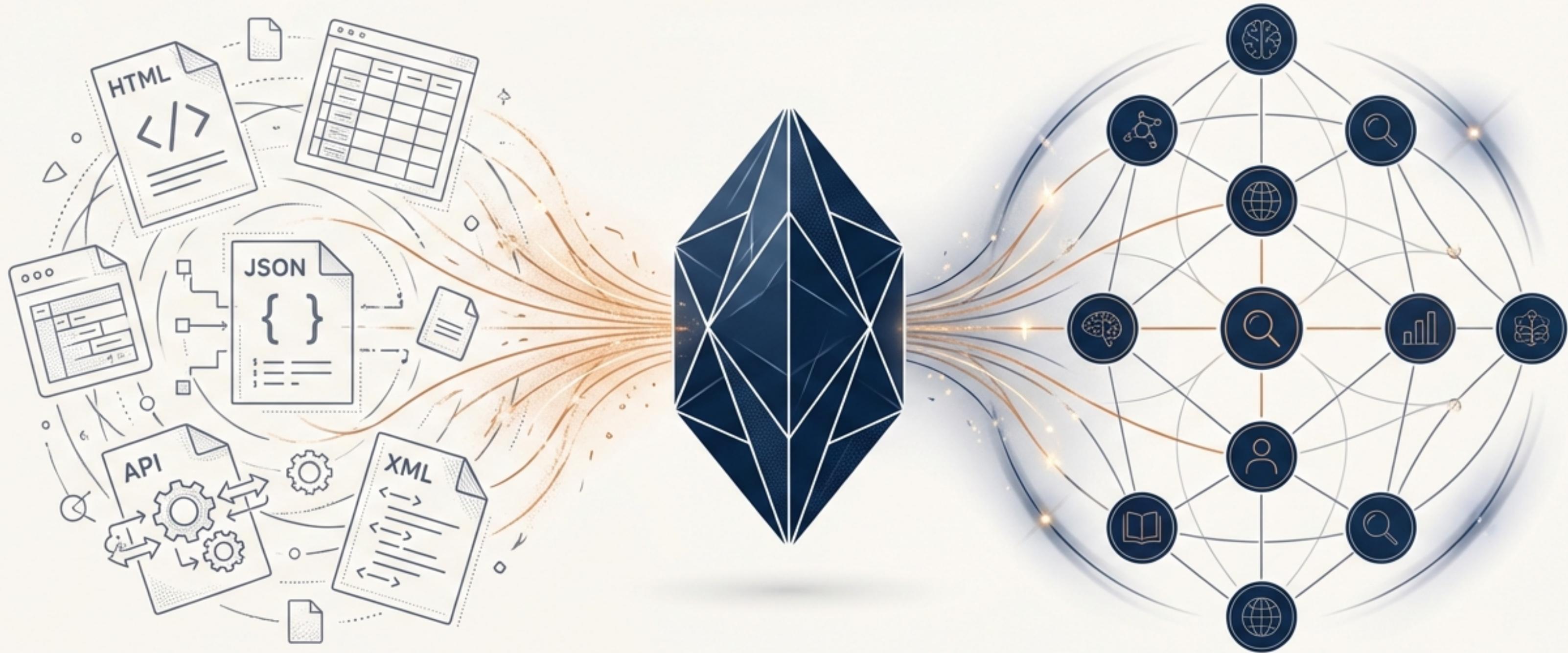


The Virtuoso Sponger: Transforming the Web into Your Knowledge Graph

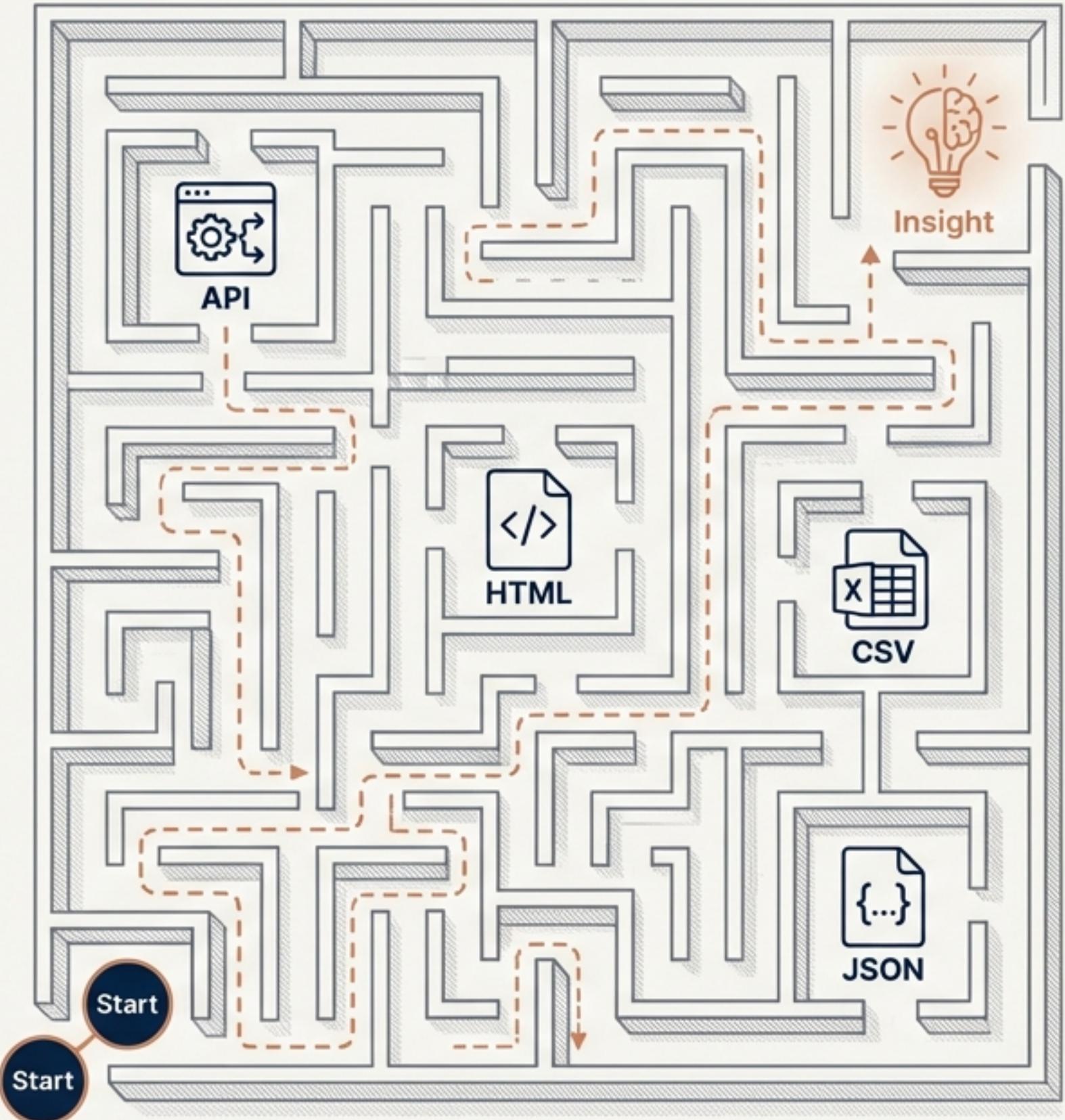
From data chaos to structured intelligence, on demand.



We Live in a Data Maze

The modern data landscape is defined by exponential growth and fragmentation.

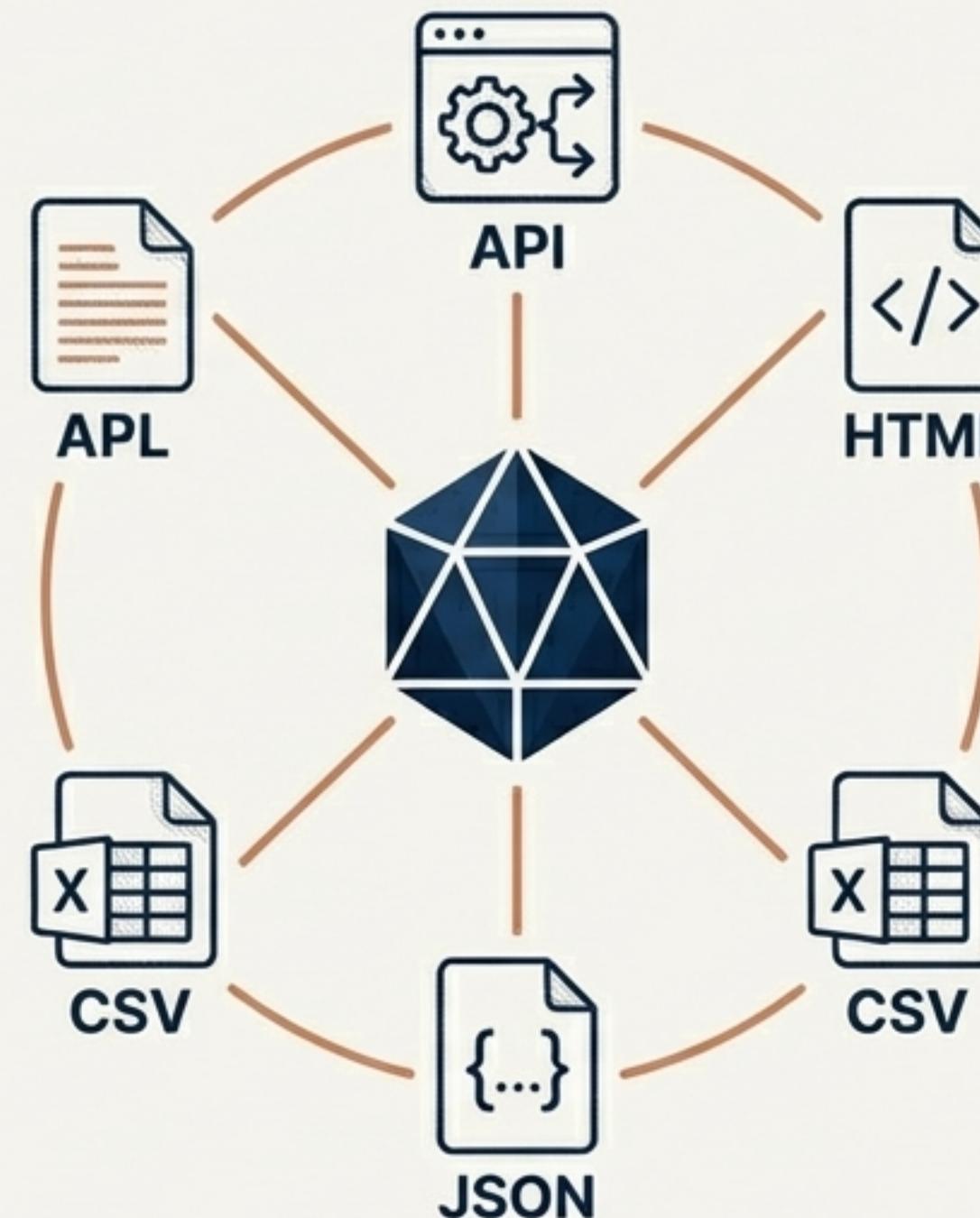
- **Data Silos Proliferate:** Information is trapped across countless disconnected sources, from public web pages and APIs to internal documents.
- **Complexity Spreads:** These challenges are no longer confined to the public Web; they are seeping into the private domains of organizations via the Hybrid Cloud Infrastructure.
- **Intelligence is Locked Away:** Extracting unified, actionable insights is a constant, resource-intensive battle.



The Sponger is the Unifying Thread

The Virtuoso Sponger is an **Extract, Transform, and Load (ETL) Middleware Layer**, built into every Virtuoso instance.

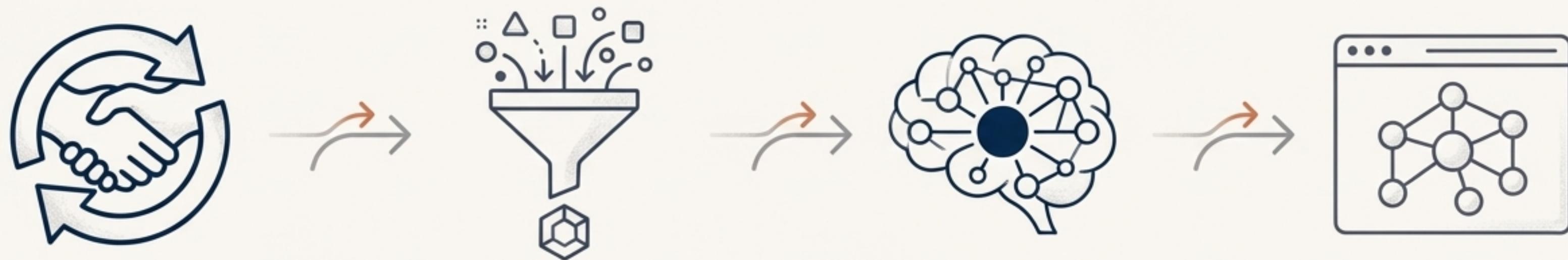
Core Function: It treats a vast array of document types and APIs as unified, structured data sources.



The Output: All structured data generated by the Sponger manifests as a **Knowledge Graph**—a collection of RDF statements woven together by hyperlinks, following Linked Data principles.

How It Works: A Four-Phase Transformation Process

Behind its simple hyperlink-driven interface is a powerful, automated workflow:



1. Negotiate

Sponger uses HTTP content-negotiation to determine the best format from the target source.

2. Extract & Transform

It applies relevant **Extractor Cartridges** (drivers for specific data types) to generate initial structured data.

3. Enrich

It then applies **Meta Cartridges** to perform advanced analysis like NLP for entity extraction and lookups against the Linked Open Data (LOD) Cloud.

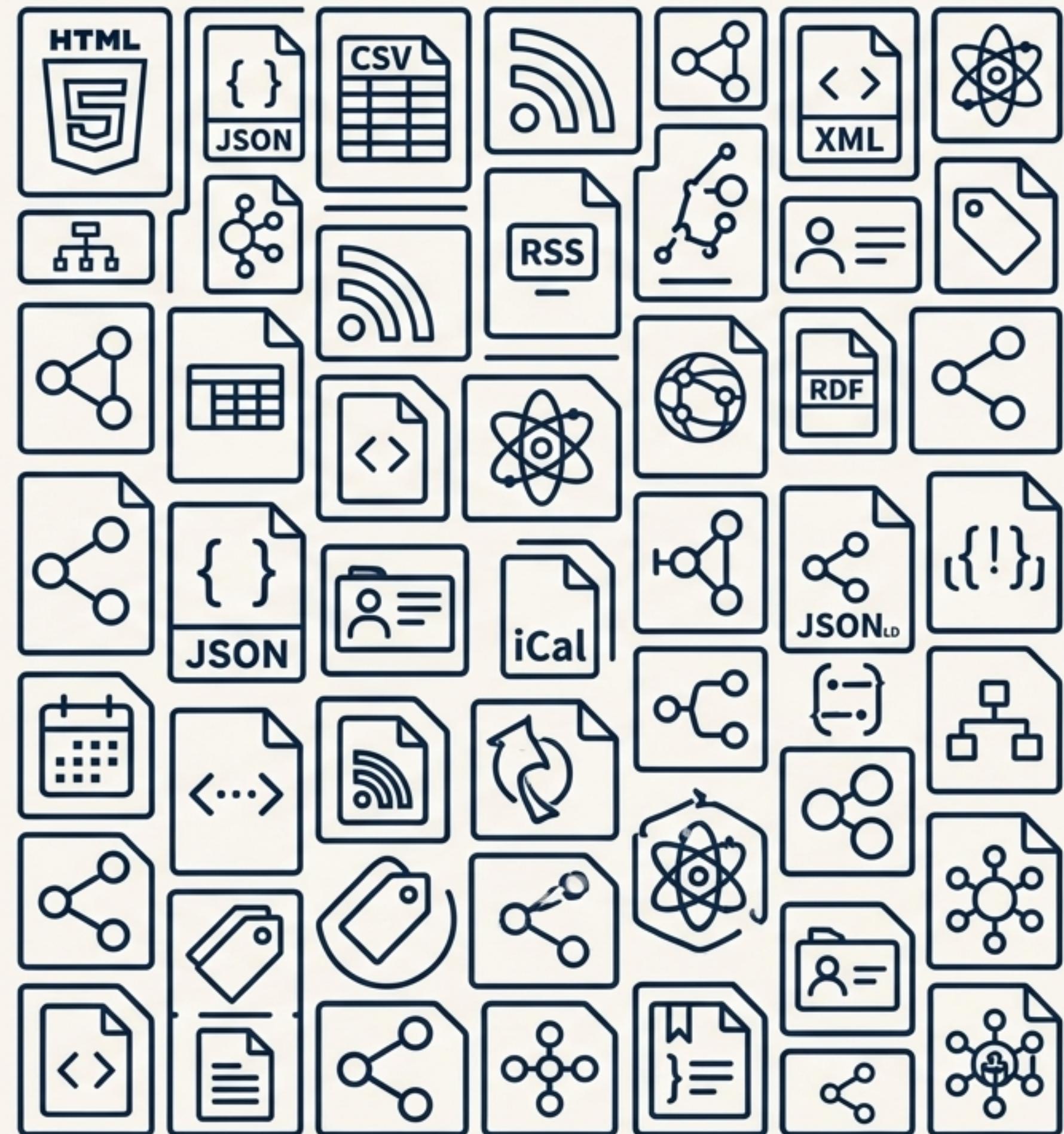
4. Present

The final, enriched knowledge graph is presented to the user agent (e.g., your browser).

The Power of Cartridges: Universal Data Connectivity

The Sponger's flexibility comes from its **Cartridges**—specialized drivers for transforming specific document types and APIs. The system supports over 70 combinations, including:

- **Web Standards:** HTML, Plain Old Semantic HTML (POSH), (X)HTML+RDFa, HTML5+Microdata, HTML5+JSON-LD
 - **RDF Formats:** RDF-Turtle, RDF-N-Triples, RDF-N-Quads, RDF-XML
 - **Data Formats:** CSV, XML, JSON
 - **Content Feeds & Calendars:** Atom, RSS, iCal/iCalendar, vCard/vCalendar
 - **And more**, with customization APIs available for further enhancement.



From a Standard Web Page...



Let's take a typical online article. To a human, it's content. To most machines, it's just a block of text and markup, with its deeper meaning and context locked away.

<https://www.fastcompany.com/90324660/how-disney-grew-its-3-billion-mickey-mouse-business-by-selling-to-adults>

...To a Rich, Structured Knowledge Graph

The Sponger ingests the URL and returns a machine-readable document describing the original article. It automatically extracts key entities and relationships, creating structured metadata.

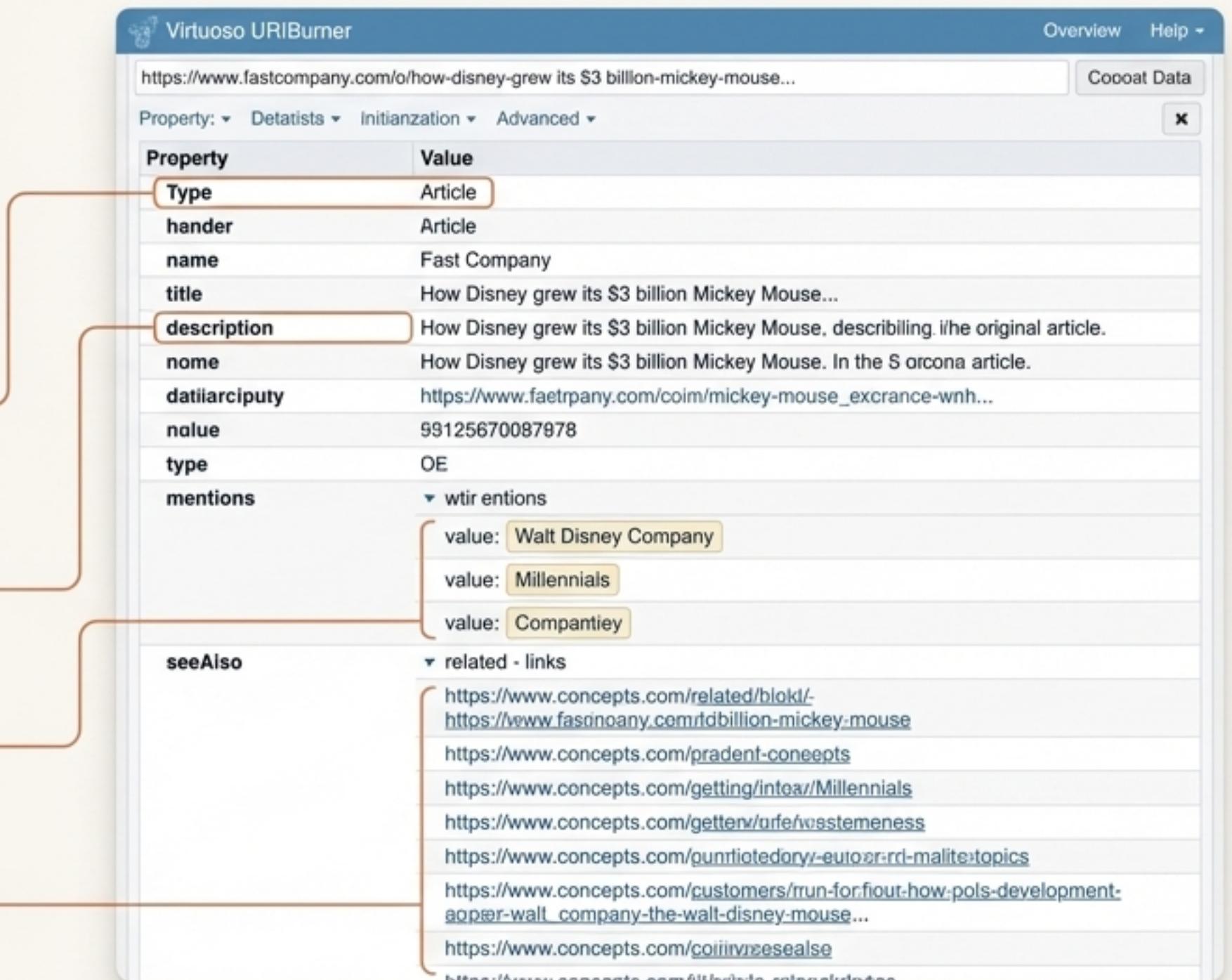
Document Type: `Article`

Primary Topic: The article itself

Title & Description: `How Disney grew its \$3 billion Mickey Mouse...`

Entities Mentioned: `Walt Disney Company` , `Millennials` , etc.

Related Topics: `SeeAlso` links to other relevant concepts.



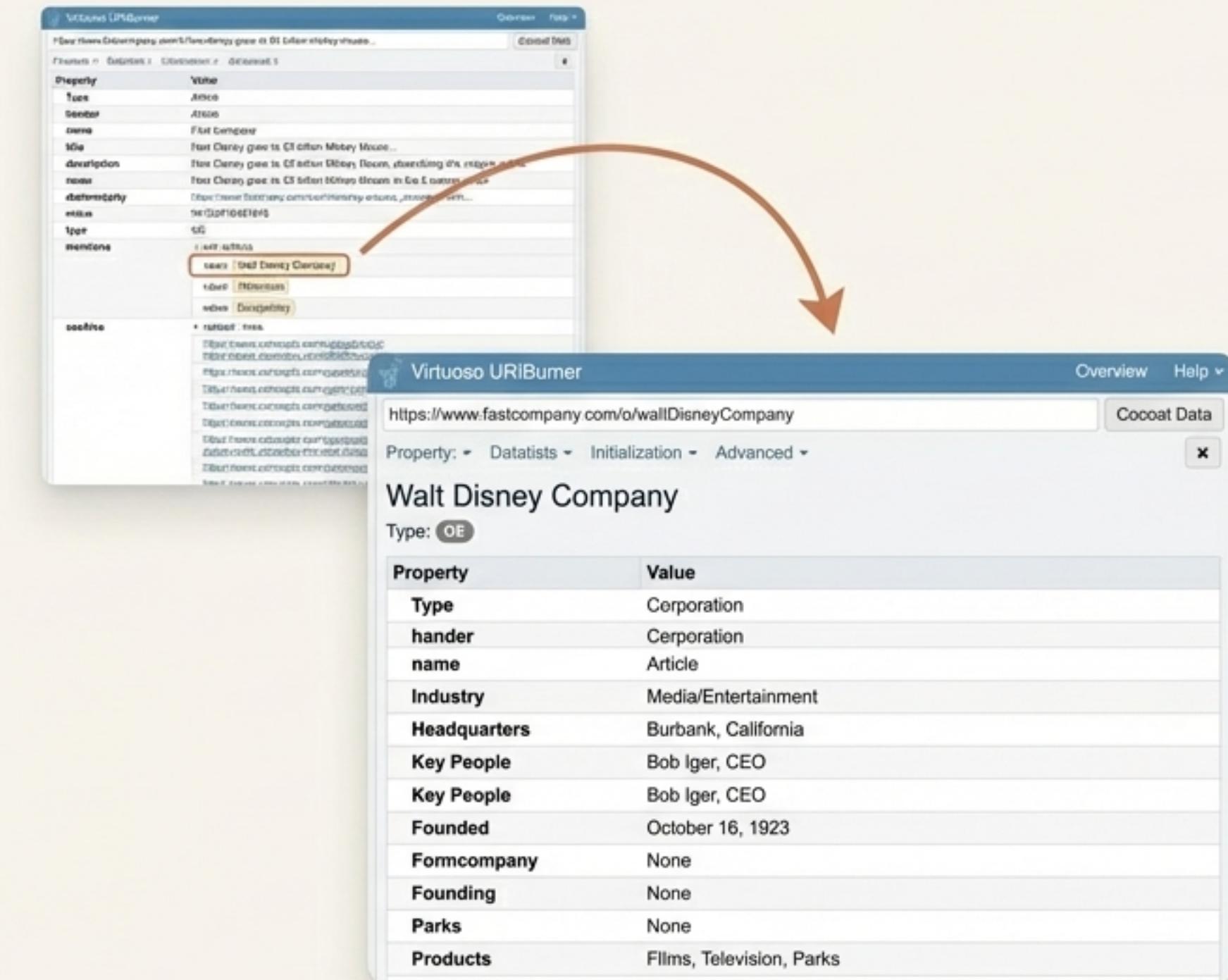
The screenshot shows the Virtuoso URIBurner interface with the following details:

- URL: [https://www.fastcompany.com/0/how-disney-grew-its-\\$3-billion-mickey-mouse...](https://www.fastcompany.com/0/how-disney-grew-its-$3-billion-mickey-mouse...)
- Property dropdown: Detatists, Initianzation, Advanced
- Property table:

Property	Value
Type	Article
hander	Article
name	Fast Company
title	How Disney grew its \$3 billion Mickey Mouse...
description	How Disney grew its \$3 billion Mickey Mouse, describiling ihe original article.
nome	How Disney grew its \$3 billion Mickey Mouse. In the S orcona article.
datilarciputu	https://www.faetpany.com/0/mickey-mouse_excrance-wnh...
value	99125670087978
type	OE
mentions	<ul style="list-style-type: none">value: Walt Disney Companyvalue: Millennialsvalue: Companiey
- seeAlso:
 - <https://www.concepts.com/related/blokt/>
 - <https://www.fasinoany.com/0dbillion-mickey-mouse>
 - <https://www.concepts.com/prudent-conecepts>
 - <https://www.concepts.com/getting/intoa/Millennials>
 - <https://www.concepts.com/getten/urfe/wosstemeness>
 - <https://www.concepts.com/guntiotedory-eutoor-rid-malite:topics>
 - https://www.concepts.com/customers/run-for:fiout-how:pols-development-soper-wall_company-the-walt-disney-mouse...
 - <https://www.concepts.com/corirvseesealse>

Hyperlinks as Super Keys for Deeper Exploration

Every entity in the generated knowledge graph is a hyperlink. This enables faceted browsing, where you can pivot from the document's metadata to explore the rich descriptions of the entities themselves. This is not just data extraction; it's the creation of an explorable data space.



The screenshot illustrates the Virtuoso URIBurner interface, a tool for generating RDF triples from various sources. The top window, titled 'Virtuoso URIBurner', shows a table of properties and their values for an entity, with a specific row for 'name' pointing to 'Walt Disney Company'. A red arrow points from this row to the bottom window, which is titled 'Virtuoso URIBurner' and displays a detailed entity description for 'Walt Disney Company'. The bottom window includes a table of properties and values, such as 'Type: Corporation', 'Industry: Media/Entertainment', and 'Products: Films, Television, Parks'.

Property	Value
type	Article
handler	Article
name	Walt Disney
bio	Walt Disney gave us Mickey Mouse...
description	Walt Disney gave us Mickey Mouse, starting the magical...
notes	Walt Disney gave us Mickey Mouse in Go East...
alternatively	Disneyland, Disney characters, Disney...
editors	Walt Disney
label	Walt Disney
spur	Walt Disney
mentions	Walt Disney

Property	Value
Type	Corporation
handler	Corporation
name	Article
Industry	Media/Entertainment
Headquarters	Burbank, California
Key People	Bob Iger, CEO
Key People	Bob Iger, CEO
Founded	October 16, 1923
Formcompany	None
Founding	None
Parks	None
Products	Films, Television, Parks

Example: Clicking on 'Walt Disney Company' in the previous view reveals a detailed entity description, connecting it to other related data within the Virtuoso database.

Querying the Live Web as a Virtual Database

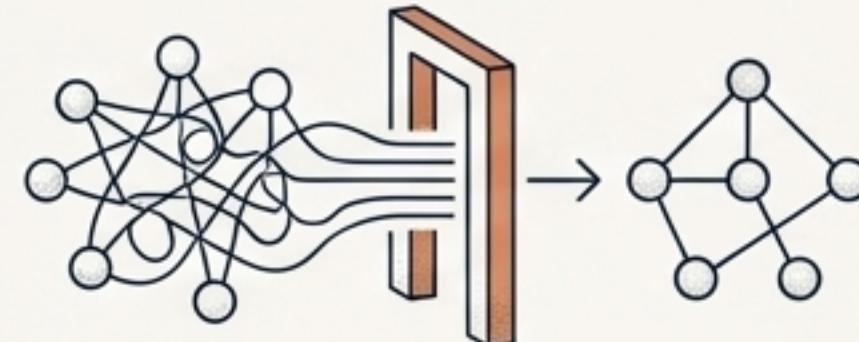
The Sponger's power is fully integrated into Virtuoso's SPARQL query service. A document URL can be used directly in the FROM clause, functioning as an external Data Source Name (DSN).



```
DEFINE get:soft "soft"
PREFIX cpu: <https://docs.google.com/spreadsheets/d/1NmrGjc8pcgh1S_0mFNABiQp
SNjY6Jxm1lA0mcxHaldg/export?format=csv#>
PREFIX dsn: <https://docs.google.com/spreadsheets/d/1NmrGjc8pcgh1S_0mFNABiQp
SNjY6Jxm1lA0mcxHaldg/export?format=csv>
SELECT DISTINCT ?s AS ?processorID xsd:string(?model) AS ?modelName ?cores
IRI(?amazonUrl)
FROM dsn:
WHERE {
  ?s cpu:Model ?model ;
  cpu:Cores ?cores ;
  cpu:Amazon_Link ?amazonUrl .
  FILTER (CONTAINS(STR(?amazonUrl),"https:"))
}
```

Example: This SPARQL query treats a live Google Spreadsheet of Intel CPUs as a relational table, extracting specific columns on the fly.

The Sponger Delivers Tangible Advantages



Simplified Knowledge Graph Exploitation: Takes the tedium out of Linked Data deployment.



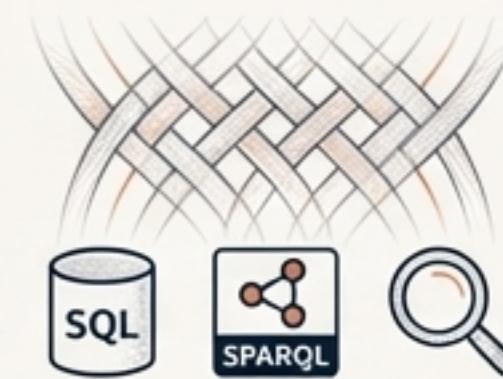
Ease of Use: Hyperlinks are the sole control mechanism for powerful data transformation.



Effortless Extensibility: New data sources are supported by adding new cartridges.



Broad 3rd Party Integration: Works with over 70 API and document type combinations.

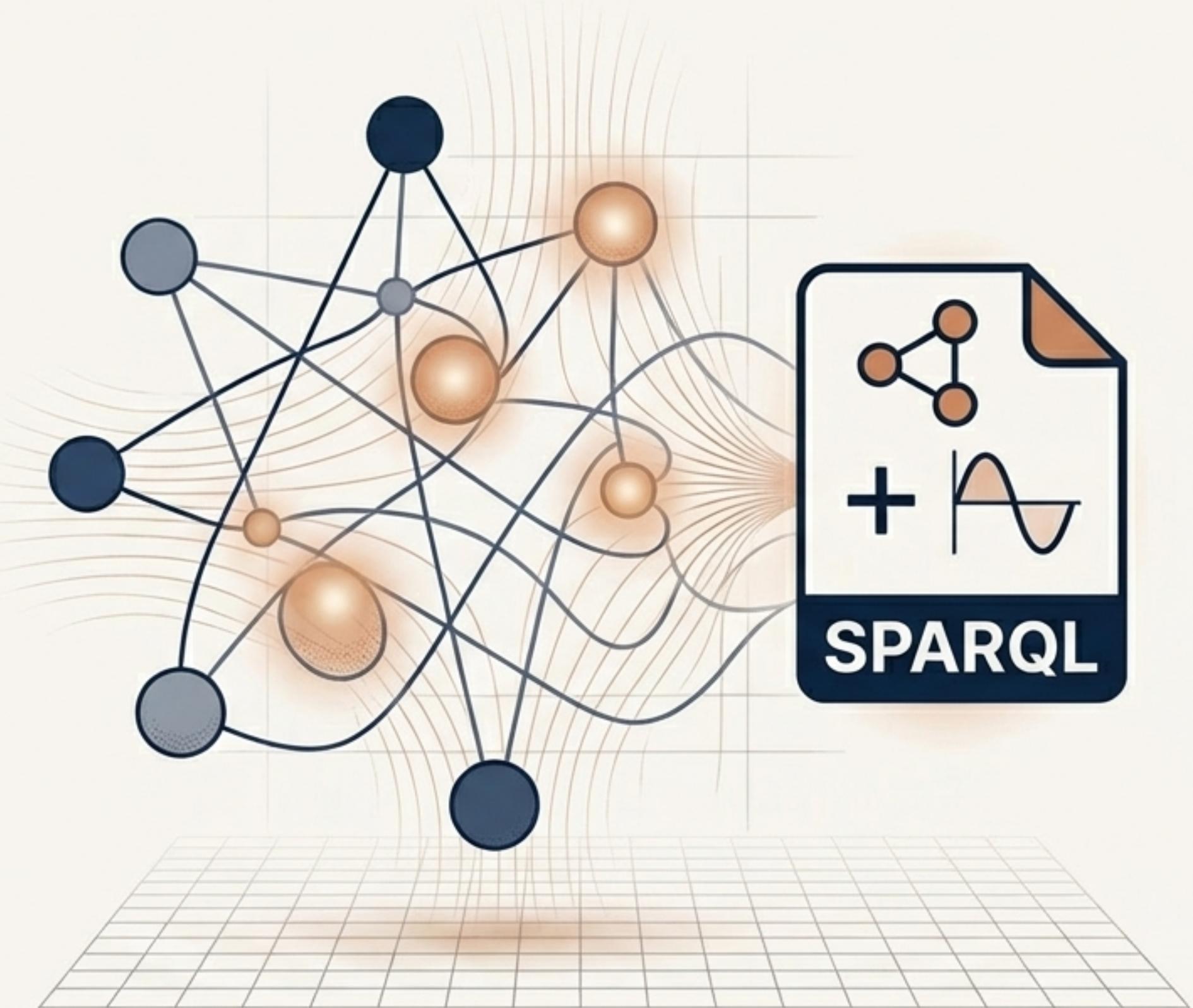


Powerful Data Meshing: A true data virtualization solution for SQL, SPARQL, and Free Text queries.

Supercharging Knowledge Graphs for the Age of AI

The Virtuoso Sponger is evolving. New **meta cartridges** now inject **vector embeddings** directly into the knowledge graph generation pipeline.

The Impact: This enhancement allows for powerful semantic analysis to be incorporated directly into your queries. You can perform **cosine similarity** analysis within a SPARQL query, combining structured, graph-based queries with semantic similarity searches.



Enterprise-Ready: Flexible Interaction and Deployment

The Sponger is a core, built-in component of Virtuoso, designed for modern infrastructure.

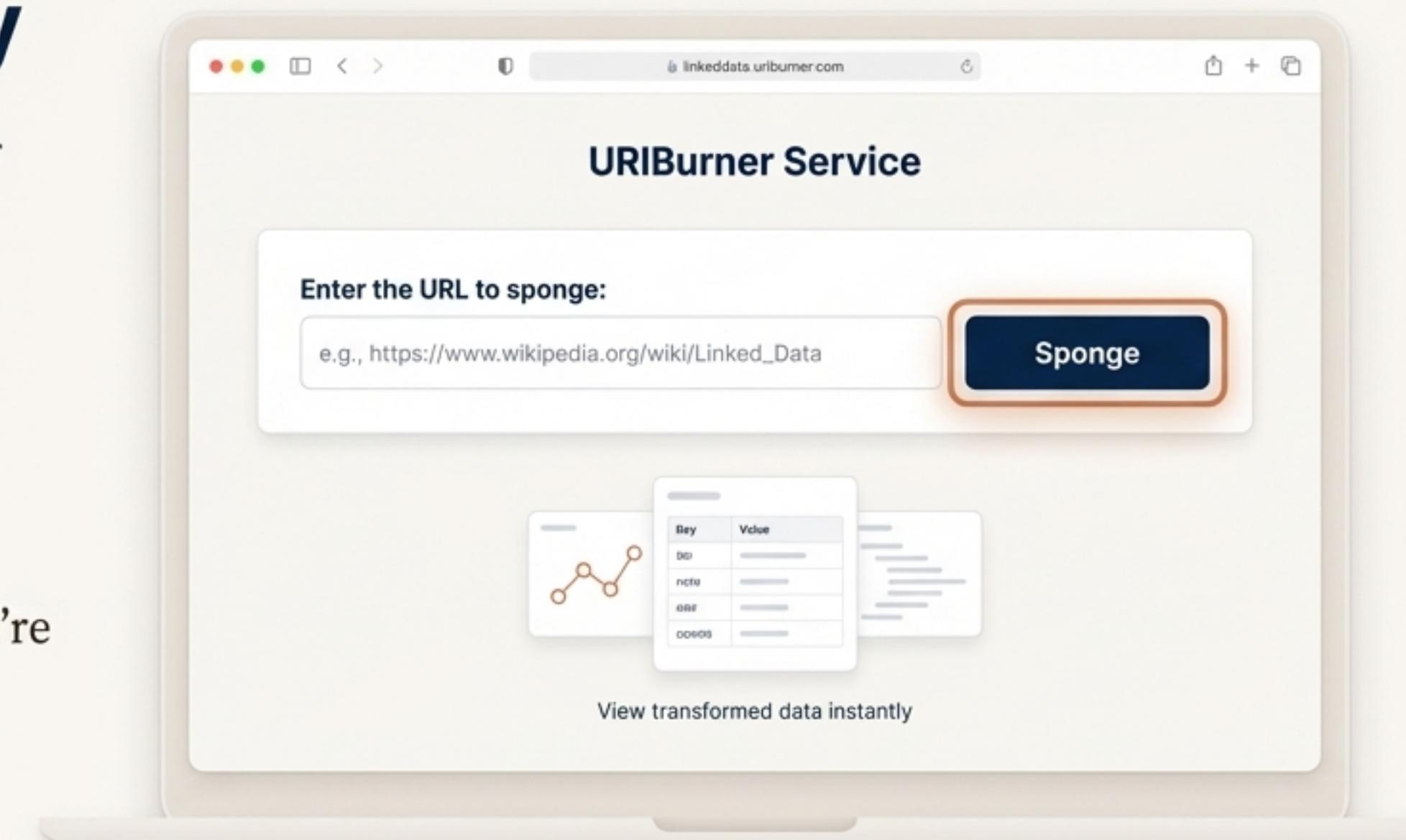
- **RESTful Interaction:** Can be interacted with directly using simple URL patterns for easy integration into any workflow.
- **Multi-Cloud Availability:** Virtuoso, including the Sponger, is available for deployment on all major cloud platforms.



Experience the Transformation Instantly

You can test the full power of the Virtuoso Sponger right now, for free.

- 1. Try the Live Demo:** Go to our URIBurner Service, a public instance of the Sponger running since 2007.
- 2. One-Click Sponging:** Install the OpenLink Structured Data Sniffer and Data Explorer Browser Extensions to transform any page you're viewing with a single click.



Your Path to a Unified Data Fabric

Ready to integrate the Sponger into your own environment?

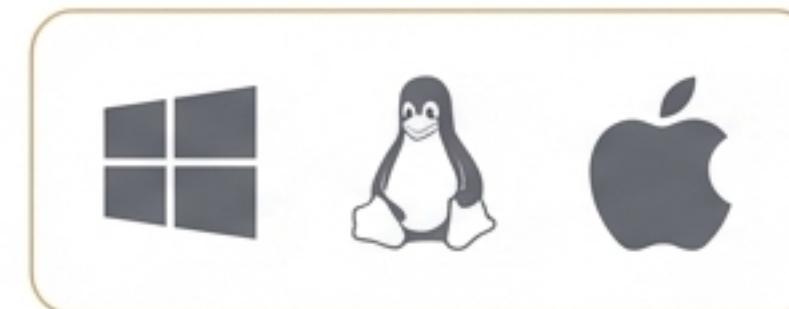
1. Explore

Try the live URIBurner service for instant results.



2. Evaluate

Download a free evaluation license.



3. Deploy

Deploy in the cloud or with containers.



Every Document is Now a Structured Data Source

The Sponger brings the magic of webby structured data to all your applications and services. It is a well-designed solution to data access and integration, built on the principle of deceptive simplicity. Your existing analytics dashboards and productivity tools can immediately morph into launch-points for exploring intelligent Knowledge Graphs, optimizing both personal and organization-wide agility.

