

## **ORGPedia: Corporate Intelligence for the 21st Century**

### What is ORGPedia?

As a result of the burgeoning movement to open up access to public data, government agencies are making unprecedented quantities of information about regulated businesses available online in reusable formats. From data about corporate compliance with health, environment and safety regulations to data about companies that receive assignments of patents to financial disclosures from publicly traded companies, the opportunity now exists to “mash up” and visualize disparate data about organizations.

The aim of the ORGPedia project is to design and create:

- An **open corporate identification system** to enable datasets about companies to be compared. Currently, different agencies use different numbering schemes. An open ID will enable taxonomies to “talk” to one another.
- An **online platform to mash up and visualize authenticated government datasets** already collected about firms and organizations pursuant to statute or regulation.
- An **Application Programming Interface (API)** and supporting software libraries to make it easy for third parties to incorporate ORGPedia into their own systems.
- A **community to encourage public participation** in reviewing, annotating and contributing to collected government data whether by companies and organizations or by third parties.

### The Goals of ORGPedia

If successful, ORGPedia would create the opportunity to:

- **Reduce the Costs of Regulatory Compliance** - With a “wikipedia for firms,” companies will need to input information only once, bringing down the costs of regulatory compliance.
- **Increase Wealth and Investment** - With better access to information, investors can make more effective decisions.
- **Inform Consumer Behavior** - More robust knowledge of company environmental, workplace safety and other practices will help both regulators and consumers.
- **Promote Corporate Accountability** - With the ability to compare government contractor databases against lobbying records, for example, it will be possible to see when a parent company lobbies and its subsidiary gets a lucrative government contract.
- **Enhance Innovation** - With the ability to mash up Patent Office and other data, economists can get a better picture of the state of innovation in our economy.

- **Enable Research on Corporate Governance** - ORGPedia will offer an unprecedented research testbed for the study of business evolution.

In short, by amassing and visualizing formerly disparate data sets about the ownership, structure, performance and regulatory compliance of organizations, we can achieve reduced compliance and investment costs, improved corporate accountability, greater consumer protection, and create an important new research and reporting test bed that will foster critical, data-driven journalism and scholarship about the life of organizations.

### The Numbering Challenge

Despite the sudden availability of so much data, it is currently impossible to combine this information and transform it into new knowledge about the social life of organizations because of the [disparate](#) and non-transparent numbering schemes used for tracking organizations.

Today there are multiple schemes in place for entity identification, many of them considered inadequate, some of them proprietary, and none of which “talk” to one another. The complexity of corporate structures presents unique challenges in designing a universal mapping (e.g. comparing accounting vs. tax structures). As a result, most systems today serve specific purposes and user communities.

The Office of Financial Research has issued an RFC in response to its own stated desire to: “standardize how parties to financial contracts are identified in the data it collects.” The Office expressed an explicit “preference to adopt through rulemaking a universal standard for identifying parties to financial contracts that is established and implemented by private industry and other relevant stakeholders through a consensus process.” In parallel, the SEC is collaborating in a ISO standard-setting process to develop a legal entity numbering scheme.

Currently, Federal agencies alone are spending at least \$52.8 million in taxpayer dollars annually on Dun and Bradstreet, the owner of the proprietary identifier system used by many government entities.

### Next Steps

The Institute for Information Law and Policy and Center on Financial Services Law at New York Law School are co-hosting 5 workshops to bring together experts whose know-how will be vital to designing a successful open corporate intelligence system for the 21st century. The aim of these workshops is to understand the needs of these different communities; elicit the questions that must be addressed for the project to succeed; identify interested partners; and draft a specification for the development of ORGPedia.

Each workshop will discuss design, and governance and also have a specific thematic focus.

- I. Open Data and Corporate Governance - will address who needs to participate in order to sustain a legitimate and robust platform and community.

2. Open Business Intelligence in the 21st Century - will convene those who create and those who depend upon business intelligence in the financial services community to articulate what's needed and to learn from the experience in that sector.
3. Promoting Corporate Accountability with Open Data and Open Identifiers - will bring together regulators and groups with an interest in corporate compliance to discuss the impact of the project on government and consumers.
4. Designing an Open Numbering Ontology and Platform - will focus on refining the specification and identifying candidate technologies.
5. International Challenges - will take the discussion outside the United States and test the assumptions developed until this point. This workshop will be held in London.

### Some Assumptions About What We Need to Build

- 1) A multi-disciplinary, independent and legitimate governance structure to oversee the growth and development of the project and to address difficult questions, such as the appropriate policy for enabling correction of inaccurate information, privacy and trade secret protection and strategies to attract engagement and participation.
- 2) A network of experts who can address policy prescriptions for tagging government-mandated data collections to ensure that this data can be mashed up.
- 3) An open, non-proprietary system for assigning unique persistent identifiers for organizations and corporations with whom Federal and state governments interact. These include government contractors, grant recipients, academic institutions, regulated facilities/plants, financial and banking entities, patent holders.
- 4) Such a system must enable the public to identify both "parent" and "child" entities; ownership and governance relationships. We need to be able to track changes over time; e.g. changes in ownership, joint ventures and mergers, births and deaths of entities, name changes, state registrations, doing business as names and related trademarks.
- 5) This taxonomy must both provide a persistent and universal federal identifier and, at the same time, interoperate with other common domestic and international identifier schemes used in legacy systems, such as:
  - 1) IRS EINs
  - 2) DUNS numbers
  - 3) CUSIPS
  - 4) Census identifiers
- 6) As these disparate efforts proceed, we want to be sure that we are not further entrenching ourselves in systems that don't allow indicators to "talk to" one another. We want to be sure that we have both a universal identifier and a system that allows other common numbering schemes to interoperate.

7) Since all such numbering schemes cannot be identified up front, we want to design an extensible system that will encourage others to participate in the ecosystem. It should support linking, tagging and enhancements for scalability.

8) ORGPedia will develop a data dictionary to standardize approaches to dealing with different terms, such as Inc. and Incorporated, abbreviations and other common elements in the description of organizations.

9) This standardization will help with developing tools by which data can be harvested from federal and state databases using web services approaches. A web services/linked data approach will help to power organizational lookups and thereby improve accuracy and reliability.

10) Both the numbering standard and the website itself should be lightweight and standards-based. We should build quickly and grow the system rather than anticipate every eventuality.

At its core, ORGPedia will offer an open numbering scheme to organize and structure authenticated data about corporations scraped from government databases. This will enable the data presented to be used for systematic research. At the same time, ORGPedia will enlist public participation to review data quality and spot errors in public data sets as well as to contribute additional data (e.g. about a company's products, its sustainability or child labor practices).