E-GOVERNMENT CORE VOCABULARIES

CORE BUSINESS VOCABULARY

USE CASES SUBMITTED TO THE WORKING GROUP

JOINING UP GOVERNMENTS

DOCUMENT METADATA

Property	Value
Release date	*16/02/2012*
Status:	*For public review*
Version:	*0.2*

This specification was prepared for the ISA programme by:
PwC EU Services

This draft specification can be downloaded from the ISA website:

TABLE OF CONTENTS

1. NATURAL PERSONS

Submitted by
Phil Archer (W3C).

Description
A natural person, Brian Ridgewell, does a variety of what might be considered contract jobs. In some cases he receives a regular salary, in others he is paid a gross amount that he must declare to the tax authorities. All income is personal income. He operates entirely as an individual and has not registered as a business and therefore does not have a legal entity separate from himself and does not have a business tax record or VAT number. Nevertheless, he does do contract work for public sector institutions and there needs to be a record of the fees paid to him as a service provider.

Comments
Chris Taggart: This is a common case, but I think will confuse the business vocabulary. Increasingly we are all getting income from non-salary work, whether it's consultancy, a bit of eBay trading, working on a short-term contract, or doing part-time casual work. In that case we are all businesses now, and this undermines the usefulness short term. In the UK this is likely to be tackled by the requirement that those who wish to work as a business for the government but as a freelance/consultant will have to publish a personal tax ID, which can then be published as an identifier. However, this is not without issues, and IMHO I think the problem here is not, how do you identify the business part of Brian's life (which from a tax and legal perspective inseparable from his personal life), but how do you identify Brian. I'm happy to leave that to the Core Person Task Force, as I think we have our hands full here!

2. OPEN GOVERNMENT DATA

Submitted by
Chris Taggart (Opencorperates)

Context
Around the world, governments have started to publish huge amounts of data under an open licence. The aim of this is to improve transparency, reduce barriers to innovation built on government data, and allow government to restructure itself using 21st century technology of data exchange, rather than the currently lingua franca of government, the emailed spreadsheet. A significant proportion of this relates to corporate entities. However, at the moment, the lack of a Core Business Vocabulary is making publication in a useful form problematic, and reuse and combination of that data with other datasets more difficult.

Example scenarios
Some of the examples are:

1. **Business Registers.** Companies House in the UK has already started publishing URIs for every registered company, and is committed to publishing a regular data dump of

the entire register under an open licence. We know of other company registers who are also intending to publish their core listing of companies as open data too. Because of the lack of a Core Business Vocabulary, the UK has had to use its own vocabulary for representing core attributes in the RDF representation. Similar problems will face other company registers wanting to publish their data

2. **Spending data.** Increasingly governments are publishing detailed information on their spending at the transaction level -- the UK now routinely publishes individual transactions for both central and local government; the EU has the Financial Transparency system; line item spending data is also published by a number of other national and regional governments around the world. However, a key problem is that the recipients are almost always identified by name. As well as being subject to representation errors (abbreviations, concatenations, notes, defunct names), these also need to be reconciled to legal entities in order to be useful. While this is being jointly done by OpenSpending and OpenCorporates, a better solution would be for the companies to be identified at source, which would also allow better data sharing between government institutions, allowing them to have better oversight over their suppliers, and bringing about efficiencies. While this is done in a limited number of cases, it is done using proprietary ID systems such as Dun & Bradstreet's DUNS numbers.
3. **Other official registers.** Many other registers list companies, from the EU lobbying register to the UK's Charity Commission, to the European Patent Office, to the European tender system, and need to identify not just the legal corporate entities, but other attributes too (registered address, entity type, status). This information is often already available as data, but its usefulness is severally hampered by the lack of a consistent way to represent it.
4. **Internal data use.** The UK has a large database called the Inter Departmental Business Register. This is created by the UK's Office of National Statistics to allow it to produce key statistics about the economy, particularly by sector, region and size of company. It is also used by the tax authorities as a key data source. Other countries will have similar such registers, but one of the obstacles to publishing it is that it uses proprietary data in identifying companies and to establish relationships between them, even though that proprietary data is sourced largely from public sources. One of the obstacles to removing the proprietary requirements is the lack of a Core Business Vocabulary, which in turn means the public data can't easily be combined.
5. **Citations.** One of the most common references to companies is in citations, whether for breaches of Health & Safety or Environmental regulations, or in court cases, or in rulings on takeovers and breaches of competition law. Increasingly this information is being published, but is rarely reused due to the lack of identifiers for the entities involved.

3. NGOS, CIVIL SOCIETY, DEVELOPMENT AID & ANTI-CORRUPTION

Submitted by

Chris Taggart (Opencorperates)

Context

Although these groups and requirements are in many ways quite disparate, they share many of the same issues -- that of identifying entities, and of sharing that identity information -- and a non-governmental status -- being outside government, yet forming a vital role to play in society,

and often working in conjunction with government. The increase in globalisation over the past 50 years and the move into a world of electronic data brought about by the internet has made the need of a Core Business Vocabulary, and the services built on it more pressing for these groups.

Example scenarios

1. **The International Aid Transparency Initiative (IATI)**. This voluntary, multi-stakeholder initiative includes donors, partner countries and civil society organisations and aims to make information about aid spending easier to find, use and compare. In conjunction with OpenCorporates, IATI held a workshop of organisational identifiers at the Open Government Data Camp in Warsaw in October, attended by many of the key stakeholder, including the World Bank, precisely to help solve some of these issues in a common and open way.
2. **Campaigners.** Many NGOs perform a vital role in monitoring the activities of multinational companies, whether in the environmental area, human rights or child labour. Publish What You Pay for example is encouraging and pushing for transparency in the extractive industries, working with governments and commodities companies to publish information about their activities.
3. **Anti-corruption.** The World Bank's recent report, The Puppet Masters, tellingly exposed how companies, particularly shell companies are used in case of grand corruption, and makes recommendations regarding the minimum information that corporate registries should collect and make publicly available about the legal and beneficial owners of legal corporate entities. Similar work is being done in the US, with bills going through both the Senate and the House to make the publication of beneficial owners a legal requirement. In addition the rise of Big Data and of data-driven journalism has meant the need and ability of journalists to use, combine and analyse company information has risen significantly, and a Core Business Vocabulary needs to support that in a lightweight but useful way.

4. FINANCIAL INSTITUTION REGISTER (EUROPEAN BANKING AUTHORITY)

Context

The Financial Institutions (FI) Register, as it is defined in Phase 1 of the project, is a list available on EBA's internet website (in .pdf-format and .xls-format) that will inform the public about the current status of Credit Institution licenses issued in the EEA. The legal basis for this is Art. 14 of the Directive 2006/48/EC (as amended by Art. 9 of Directive 2010/78/EU, aka "Omnibus"-Directive).

Scenario

It will be updated twice a year with data sent by National Supervisory Authorities through a portal. The Register will replace the list of credit institutions formerly published in the Official Journal of the European Union (OJEU) by the European Commission according to Art. 14 of Directive 2006/48/EC (see the last updated version of the EU commission's list).

The interesting fields for a core vocabulary and our experience in field size are the following:

- CODE XML
- NAME

- TYPE
- MAX LENGTH
- PRESENCE MA/OP/CO(1)
- CONTROLS
- COMMENTS

FUNAM	Full Name	Alpha	200	MA	Value must contain 3 characters at least.	In several countries you can have more as one name
TYPE	FI Type	Alpha	3	MA	The Type should have one of the following value : • **CRD** for Credit Institution • **BRA** for EEA Branch	A classification for the type of institution could be helpful

TOWN	Town	Alpha	50	CO	Mandatory if TYPE= CRD Optional if Value must contain 2 characters at least.

HSTAT	Head office State	Alpha	2	CO	Irrelevant if TYPE=CRD. Mandatory If TYPE = "BRA". As a reminder, State ISO-2 codes of EEA States have the following values : • **AT** for Austria • **BE** for Belgium • **BG** for Bulgaria • **CY** for Cyprus • **CZ** for Czech • **DK** for Denmark • **EE** for Estonia • **FI** for Finland • **FR** for France • **DE** for Germany • **GR** for Greece • **HU** for Hungary • **IS** for Iceland • **IE** for Ireland

					• **IT** for Italy • **LI** for Liechtenstein • **LV** for Latvia • **LT** for Lithuania • **LU** for Luxembourg • **MT** for Malta • **NL** for Netherlands • **NO** for Norway • **PL** for Poland • **PT** for Portugal • **RO** for Romania • **SK** for Slovakia • **SI** for Slovenia • **ES** for Spain • **SE** for Sweden • **GB for United Kingdom** • **GI for Gibraltar**

HTOWN	Head office Town	Alpha	50	CO	Mandatory if TYPE = "BRA" Irrelevant if TYPE= "CRD". Value must contain 2 characters at least.
MCR	Minimum Capital Requirements	Alphanum	4	MA	**Y :** Initial capital equal to or higher than 5 million Euros **N** : Initial capital between 0 and 5 million Euros **NULL** : No initial capital

Legend:
MA: mandatory
OP: optional
CO: conditional

A second phase will be launched next year, which will have more details on the Financial Institution Register.

5. USE BY COMPANIES THEMSELVES

Submitted by

Chris Taggart (Opencorperates)

Context

As companies become increasingly complex, and dependent on a large number of suppliers and customers in different jurisdictions the need for them to be able to map out those relationships, and to exchange that information has become critical, and the failure to do so can cause serious disruption, reputational damage and even put the whole business at risk. In addition, the ability of shareholders to oversee the companies of which they are owners has greatly lessened as their ability to understand the business has reduced.

Examples

1. **Supplier chains.** the 2011 Japanese Earthquake showed how many companies were exposed to risk of their suppliers failing to deliver or going out of business (increased by Just In Time manufacturing), as did the failure of Lehman's, exposing hidden counterparty risk. In another area, global fashion companies such as Gap regularly get hurt by the discovery that far down their supplier chain child labour is being used, causing a considerable reputational risk. Mapping those supplier chains at the moment is near impossible, still less being able to monitor what happens to the suppliers (bankruptcy, health & safety violations). Moving forward, it seems possible that many such companies will publicly map these supplier chains to prove their cleanliness, and to allow the monitoring to be done by the wider community.
2. **Customer risk.** When a large company fails, it takes down with it a number of other dependent companies. At the moment this is hard for companies to map, hard for governments and regulators to have visibility on and hard for credit ratings agencies to calculate.
3. **Efficiency**. As with government, there is a big potential win for large companies understanding their own supplier and customer base, and being able to tie that information to publicly available sources of information in a consistent way.
4. **Corporate Governance.** The past five years have brought so many examples of corporate governance failure, it's not necessary to list them all, from Lehmans to Enron to BP to the banks. Many companies which were previously thought to be successful, healthy, or behaving responsibly were discovered to be very different entities that were believed by the shareholders and regulators. Though this is a far bigger issue than just vocabularies. The first steps to solve this problem are finding a way of identifying companies and the relationships between them.

6. EUROPEAN BUSINESS REGISTER (EBR)

Context

In 2009 the European Commission adopted a Green Paper and launched a public consultation in order to assess the need for an improved cooperation between the business registers of the Member States of the EU. Business registers register, examine and store company information, such as information on a company's legal form, its seat, capital and legal representatives, and they make this information available to the public. Accordingly, they have a key role in ensuring transparency across the markets and thus restoring trust following the financial crisis.

Business need

When conducting business transactions companies need to have access to comparable, up-to-date and official information about other companies in order to ensure effectiveness and security of these transactions.

Scenario

Company A located in a European Member State needs to conduct a business transaction with Company B located in another European Member State. Company A needs to ensure that the information about Company is reliable, official and up-to-date. Company A accesses a single point of contact which provides access to the data stored in National Business Registers of Member States. Company A can check and compare data from across Europe that is made available through standardised reports which supports multi-language interface.

Derived requirements

The Core Business must specify:

- The minimal subset of metadata that must be exchanged with Business Registries and that are needed for the most frequent business transactions, for example:
  - legal name of a company [Business name]
  - Registered office [Address]
  - People representing the company (Managing Director, Board of Directors, holders of Proxy) [Person]
  - subscribed capital if any [Financial capital]
  - status
  - etc.
- Multi-language support
- [To be Completed]

7. XBRL EUROPE BUSINESS REGISTERS CORE TAXONOMY (XEBR)

Submitted by

Thomas Verdin (XBRL Europe**)**

Context

The xEBR Taxonomy has been developed by the XBRL Europe Business Registers Working Group for more interconnection between Business Registers in Europe (cross-border data exchanges) and an easier access and comparison of the company data for the end users.

XBRL (eXtensible Business Reporting Language) is the universal language for business and financial communication. It's based on XML.

XBRL Europe is a non-profit association that has been set up to foster European XBRL efforts and to implement common XBRL projects in Europe between its members and to liaise with European Authorities and organisations.

Members of the xEBR WG are Business Registers or Information Providers from more than 14 countries in Europe, with observers from India, Singapore, and Taiwan.

The xEBR Taxonomy is core reference taxonomy designed to connect local taxonomies (made by national registers) for company profile (identity and address), history and financial statements across Europe. It includes some core / common concepts that are linked with the local concepts in the national taxonomies. As each taxonomy contains presentation links

between its own concepts, the combination of the core (mapping) links (from the xEBR taxonomy) and the local (presentation) links (from the local taxonomies) allows to render (display on screen), compare (direct access to core concepts) or analyse (calculation of ratios) cross-border data.

The xEBR Taxonomy contains

- core reference tags for "Company Profile" = identity (including company address/location)
- core reference tags for "Financial Statements" ("annual accounts")
- core reference tags for "Company Officials" = persons (including officials address)

These core reference tags/concepts have been chosen after analysing the equivalent tags in national taxonomies and in the EBR schemes.

Most of the business registers in Europe uses the XBRL technology. In 2008-2009 each of them developed local taxonomies for company profiles (identity) and financial statements. These were individual initiatives. With the creation of XBRL Europe, they decided to share their experiences and defined common core taxonomy for linking their various local taxonomies. This taxonomy is known as the xEBR Taxonomy and includes common tags (concepts) for the company profile (identity and address), the company officials (persons) and the company history (events). Mappings between the core reference taxonomy and national taxonomies allows cross-border comparisons and data exchanges (business registers interconnection).

Business need

Similar to EBR (end users need to have access to comparable, up-to-date and official information about other companies). However, xEBR uses the XBRL technology to compare structured records rather than a limited number of predefined concepts: local links (in the national taxonomies) and common links (in the xEBR core reference taxonomy) can be combined (e.g. once the "address" tag has been identified in two national taxonomies, the presentation's links in those local taxonomies may be used to explain the content of an "address" in each country).

Usage scenario

The xEBR taxonomy is used between the national registers for cross-border exchanges (today: IT, FR, SP : the search of "AL ITALIA" on Infogreffe/France will connect you directly with the company profile on Infocamere/Italy). Other use cases have been presented for cross-border data analysis (financial ratios, company profiles...).

Derived requirements

The xEBR Core Reference Taxonomy includes concepts on company profile, company officials, company financial statements and company history. It has been developed by a XBRL Europe Working Group with representatives from 14 Member States (business registers, national banks, data providers).

Main content of the Core Reference Taxonomy

- Company ID [List]
- Company name [List]
- Company legal form [List]
- Company activity [List]
- Company address [List]
- Collective procedure [Presentation]

8. IDENTIFICATION OF A LEGAL ENTITY

Submitted by

Chris Taggart (Opencorperates)

Description

This is the big one. Unlike natural persons, which are physical things, a business is a conceptual entity created explicitly (sometimes implicitly) by a state. This allows it to have legal standing, and thus agree contracts, have a bank account, owe money, and have possessions. This legal entity is separate from its (current) legal name, and just like a person the name can change over the course of their life, and may also have subtly different but valid variations at any one time (Mr F. Flintstone B.A. Hons may be also known as Fred Flintstone; Bedrock Infrastructure Ltd may also be known as Bedrock Infrastructure Limited, and may also have several registered trading names).

In virtually every single jurisdiction (see On company identifiers, the web and reinventing the wheel), when these entities are brought into existence they are given unique, permanent identifiers (usually called company numbers). Separately from this core identifier and jurisdiction (both of which we need to find ways of describing), the legal entities have a number of attributes that are common in one form or another across jurisdictions, although with different names, and different requirements to publish.

1. **Legal name.** There is usually a single official name of the entity in the register. This frequently changes over time, and it's not uncommon for the same legal name to be used by different and unconnected legal entities over time.
2. **Registered address.** Note this is not the same as a headquarters address, trading address, or similar (although it sometimes is the same place), but is commonly an address where suit can be filed and accepted and where company documents can be inspected, and in practice in many places is the address of a lawyer, accountant, or local agent.
3. **Entity type.** What form of entity is it – company, partnership, corporation, etc. Note: It may be useful to start with the native name of the entity type as a string (in the UK 'Public Limited Company', 'Private Limited Company', etc; in Spain Sociedad Anonima, Sociedad Limitada, etc). We've been looking at abstracting some of these into generalised form (Publicly Owned Limited Company, Private Limited Company, etc), but we're still at the very early stages with this.
4. **Status.** The present legal status of the legal entity. I think this should be associated with a date, but maybe the date should be associated with the whole record. In theory we could use a period for this (from XXXX to YYYY), but I think in practice we would rarely know these dates, and they would often be open ended. Jurisdictions vary in the terms and types of status, using 'Active', 'Dissolved', 'Removed From Register', 'In

Liquidation', etc. with lots of local languages. OpenCorporates abstracts these into “Active” and “Inactive” by mapping the variable possible Status to these. It's possible there are better or more granular abstractions, but it's a useful filter.

5. **Registration date** (aka formation date, incorporation date). The date the legal entity was formed. Note that in the case of a foreign branch, this is often not the same date as the parent company's registration in the home country. We have come across situations that companies have been deregistered and then registered, but I think this is an edge case.
6. **Deregistration date.** When the company is dissolved, or permanently ceases for some reason.
7. **Trading names.** Sometimes these are registered, sometimes they aren’t.

9. RELATIONSHIPS BETWEEN ENTITIES

Submitted by

Chris Taggart (Opencorperates**)**

Description

This is more subtle and difficult than it might seem, often because you may know that the ultimate parent company of Foo UK Ltd is Foo Inc (from, say, SEC reports), but that it often isn't owned directly, or 100% owned, and many of the intermediate steps are obscured by offshore companies about which little is known (or knowable).
One of the main jobs with this is determining what these relationships are, and whether we need explicit inverse relationships (is_parent_of, is_subsidiary_of, is_foreign_branch_of, is_home_company_of), and whether shareholdings fall into this category, and whether we need terms for inferred relationships.

Merger & acquisitions

Submitted by Chris Taggart (Opencorperates**)**

Description

Also other changes to entities have to be taken in account (e.g. changing from one entity type to another). This is arguably an event (a sale of shares, for example) which leads to a change of status, or creation of a new entity. Along a similar line are appointment of liquidator, administrator etc. We've just started trying to understand this, and it may be better to put this in the next list.