Footprinted.org: experiences from using linked open data for environmental impact information

Jorge Luis Zapico[1,3], Bianca Sayan[2,3], Leonardo Bonanni[3], Marko Turpeinen[1] and Steve Young[2].

[1] Center for Sustainable Communications, KTH Stockholm Sweden
[2] University of Waterloo, Canada
[3] MIT Media Lab, Cambridge, Massachusetts.

{zapico,bianka,leo}@media.mit.edu, turpeinen@kth.se, sb.young@uwaterloo.ca

Sustainability science relies heavily on information (Allenby 2000, 2006). And as sustainability is gaining weight in decision-making, good and accessible environmental information is needed (Goleman 2009). This is true both at an institutional level, like when deciding the materials for building a product, and at a personal level, deciding between chicken and salmon at the supermarket. However, most of the environmental information is closed, based on proprietary software, expensive or in text documents that are not possible to process.

We believe that it is necessary to bring open data concepts from the web to environmental impact information (Davis et al, 2010; Zapico et al, 2010). This would increase transparency, openness, and make it easier to create sustainability services on top of the data.

Exploring these ideas we have created Footprinted[1], a web service that is trying to solve these problems by opening up the information using linked data, focusing in life cycle assessment information. This article presents the finished application, the experiences developing it, and the first usages.

1.   The Application

The key concept for Footprinted is **making environmental information open, linked and usable**. This service is the result of a design research process (Brown, 1992; Barab and Squire, 2004), where the application is a way to test the theoretical ideas in practice, and at the same time to produce a functional product that can be used in the real world. It is a collaboration between KTH, Sourcemap inc and MIT Media Lab[2].

Footprinted uses linked data (linked data will be explored in section two) for storing environmental impact information so it can be easily created, presented, shared and reused. The resources in the repository are life cycle assessment (LCA) or other environmental impact analysis methodology (Baumann and Tillman, 2004). For the same product or service there can be different resources, for instance a LCA of leather produced in Brazil for 2010 and other of leather from Spain for 2005. There can even be two different competing results for the same product. The application's goal is not to provide one correct answer, but a forum to access

---

[1] Footprinted is still in a beta stage. For more information see http://footprinted.og

[2] More info at http://footprinted.org/about/team

different answers and a way of selecting and improving the best data. It can be seen as a *bazaar* of LCA instead of a *cathedral* or unique answer approach[3]. The best way of presenting Footprinted is to look at its main functionality:

1.1 Homepage

The start page is designed to present several hand picked life cycle assessments. It's directed for casual visitors, users that are visiting the site without a concrete purpose. We want to awake their interest and direct them to some resources that they can start exploring. It presents the essential information in a compact way: the name, source, main impacts, year, location and category.
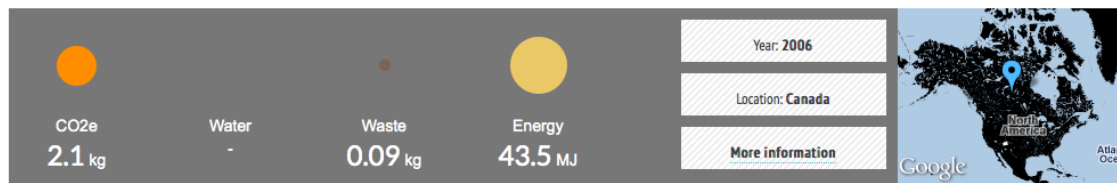

Figure 1: Compact view of resource in homepage

1.2 Search

The search page[4] is designed for users that are looking for something in particular in the repository. It can be navigated either by keywords or using a category tree. It presents the resource in a minimal way for easy comparison between the results including key data as name, year, location and impacts available for each one.
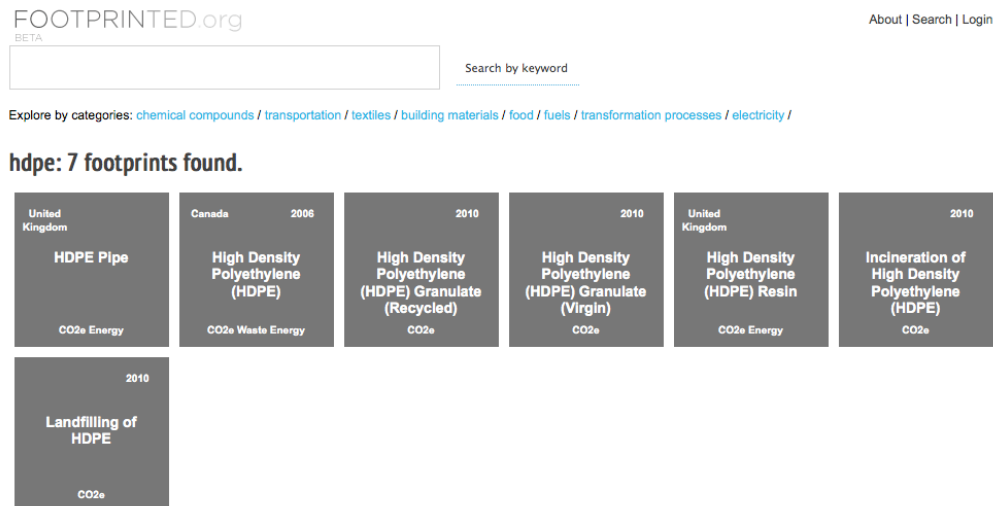

Figure 2: Search results for keyword HDPE

---

[3] For the origin of this analogy related to open source, see Raymond, 1999.

[4] See: http://footprinted.org/search

1.3 Fact sheet

The fact-sheet or main view[5] of a resource shows all the existing data connected with it. It has several parts:

Figure 3: Fact-sheet for a LCA of HDPE with zoom in detail.

- *Name and functional unit*: what is being assessed and what is the unit used. For instance: one kilogram of HDPE, one kilometer traveled by bus per passenger. This is by default normalized to one (of whatever unit is used) and if possible transformed to the metric system.
- *Impact Assessment*: the environmental impact of the functional unit in different categories. Typical impacts are: global warming potential (CO2e), water and energy consumption, waste, toxicity.
- *Flows*: inputs and outputs of the process or the production of the product (energy, raw materials, waste, water).
- *Applies to*: the year and location that the assessment was performed and that it is most relevant for applying it.

---

5 For instance: http://footprinted.org/HighDensityPolyethyleneHDPE40221568

- *Reference*: the original source from the information, author, link, title, year.
- *More info*: extra semantic information connected to the resource such as categories to which it belongs and concepts that are equivalent.
- *Export*: links to save the complete information in a machine readable format, either RDF or XML.
- *Share*: shortcuts for sharing the resource in social media sites such as Twitter and Facebook.
- *Comments*: basic comment feature so registered users can discuss the validity and problems of the resource.

All these sections are not available for all the resources and it depends on how much the user that created it uploaded to Footprinted or how much was available when importing the data set.

1.4   Contributing

There are two ways of contributing data[6] to Footprinted, either manually filling up a form with the required information as in figure 4, or uploading an EcoSpold[7] file that is automatically imported. This functionality is reserved for registered users, focusing mainly on environmental practitioners and researchers.

Figure 4: Form for creating new resources.

1.5   API

The Application Programming Interface[8] is the platform that developers use for integrating Footprinted data into other services. It allows to search and retrieve all the information from the repository in machine readable formats such as JSON, XML and RDF.

---

[6] More information at http://footprinted.org/about/contribute

[7] The format from EcoInvent http://www.ecoinvent.org/ecoinvent-v3/ecospold-v2/

[8] More info at: http://footprinted.org/about/api

http://www.footprinted.org/api/search?name=hdpe&key=xxxxx&encode=html

array(7) { [0]=> object(stdClass)#19 (12) { ["uri"]=> string(16) "hdpe-pipe4490130" ["name"]=> string(9) "HDPE Pipe" ["unit"]=> string(2) "kg" ["year"]=> string(1) "0" ["geography"]=> string(14) "United Kingdom" ["co2e"]=> string(4) "2.52" ["water"]=> NULL ["energy"]=> string(4) "84.4" ["waste"]=> NULL ["category"]=> string(18) "building material;" ["ref"]=> string(87) "Hammond, G. Jones, Craig. () Embodied Carbon - the Inventory of Carbon and Energy (ICE)" ["public"]=> string(1) "1" } [1]=> object(stdClass)#21 (12) { ["uri"]=> string(35) "HighDensityPolyethyleneHDPE40221568" ["name"]=> string(32) "High Density Polyethylene (HDPE)" ["unit"]=> string(2) "kg" ["year"]=> string(4) "2006" ["geography"]=> string(6) "Canada" ["co2e"]=> string(3) "2.1" ["water"]=> NULL ["energy"]=> string(2) "43" ["waste"]=> string(8) "0.085711" ["category"]=> string(18) "chemical compound;" ["ref"]=> string(54) "Haight, Murray. (2006) Canadian Raw Materials Database" ["public"]=> string(1) "1" } [2]=> object(stdClass)#22 (12) { ["uri"]=> string(57) "high-density-polyethylene-hdpe-granulate-recycled-6124186" ["name"]=> string(53) "High Density Polyethylene (HDPE) Granulate (Recycled)" ["unit"]=> string(2) "kg" ["year"]=> string(4) "2010" ["geography"]=> string(0) "" ["co2e"]=> string(4) "0.52" ["water"]=> NULL ["energy"]=> NULL ["waste"]=> NULL ["category"]=> string(18) "building material;" ["ref"]=> string(30) "(2010) Okala Design Guide 2010" ["public"]=> string(1) "1" } [3]=> object(stdClass)#23 (12) { ["uri"]=> string(55) "high-density-polyethylene-hdpe-granulate-virgin-8943783" ["name"]=> string(51) "High Density Polyethylene (HDPE) Granulate (Virgin)" ["unit"]=> string(2) "kg" ["year"]=> string(4) "2010" ["geography"]=> string(0) "" ["co2e"]=> string(3) "0.8" ["water"]=> NULL ["energy"]=> NULL ["waste"]=> NULL ["category"]=> string(18) "building material;" ["ref"]=> string(30) "(2010) Okala Design Guide 2010" ["public"]=> string(1) "1" } [4]=> object(stdClass)#24 (12) { ["uri"]=> string(43) "high-density-polyethylene-hdpe-resin5835386" ["name"]=> string(38) "High Density Polyethylene (HDPE) Resin" ["unit"]=> string(2) "kg" ["year"]=> string(1) "0" ["geography"]=> string(14) "United Kingdom" ["co2e"]=> string(4) "1.93" ["water"]=> NULL ["energy"]=> string(4) "76.7" ["waste"]=> NULL ["category"]=> string(18) "building material;" ["ref"]=> string(87) "Hammond, G. Jones, Craig. () Embodied Carbon - the Inventory of Carbon and Energy (ICE)" ["public"]=> string(1) "1" } [5]=> object(stdClass)#25 (12) { ["uri"]=> string(54) "incineration-of-high-density-polyethylene-hdpe-7831096" ["name"]=> string(48) "Incineration of High Density Polyethylene (HDPE)" ["unit"]=> string(2) "kg" ["year"]=> string(4) "2010" ["geography"]=> string(0) "" ["co2e"]=> string(3) "1.4" ["water"]=> NULL ["energy"]=> NULL ["waste"]=> NULL ["category"]=> string(23) "transformation process;" ["ref"]=> string(30) "(2010) Okala Design Guide 2010" ["public"]=> string(1) "1" } [6]=> object(stdClass)#26 (12) { ["uri"]=> string(26) "landfilling-of-hdpe7568055" ["name"]=> string(19) "Landfilling of HDPE" ["unit"]=> string(2) "kg" ["year"]=> string(4) "2010" ["geography"]=> string(0) "" ["co2e"]=> string(5) "0.014" ["water"]=> NULL ["energy"]=> NULL ["waste"]=> NULL ["category"]=> string(23) "transformation process;" ["ref"]=> string(30) "(2010) Okala Design Guide 2010" ["public"]=> string(1) "1" } }

Figure 5: Example of API response.

2.   Linked data and semantic web

Linked data[9] is a method of publishing data using web standard so it is interlinked and readable both for human readers and computers. The basic concept is:

- Things are identified by unique internet addresses (HTTP URIs).
- When looking at an URI people and computers get useful information about that thing in a standard and open format.
- That information includes links to other URIs of things related to that information.

2.1.   Implementation of Linked Data architecture

Footprinted implements this architecture:

- Each resource has a unique and permanent address[10].
- When accessing that URI a human will get a fact sheet with the information visualized in a readable format. When a machine (such another web service) accesses that URI, it would get the information using a machine-readable format such as RDF or JSON so the data can be easily interpreted and reused without human intervention.
- The information retrieved contains other URIs for concepts. For instance when it says kilogram, it's not just a word, but an URI from Qudt.org that contains a description, abbreviations, conversion rates...

Adding semantic information and working with ontologies provide opportunities for disambiguation and machine readability. It allows Footprinted to use existing information and repositories. The power of linked data can be seen in this simple example:

- We have a resource for Primary Aluminum Ingot:
  http://footprinted.org/PrimaryAluminumIngot84706627
- This has the triple owl:sameAs indicating that it is equivalent to Aluminum:
  http://sw.opencyc.org/concept/Mx4rvVi2O5wpEbGdrcN5Y29ycA

---

[9] More info at: 2009. Tim Berners-Lee. Available: http://www.w3.org/DesignIssues/LinkedData.html

[10] For example Aluminum ingots has the URI: http://footprinted.org/PrimaryAluminumIngot84706627

- This resource gives extra information about Aluminum, like that it is a Type of Metal.
- It is also linked to *http://dbpedia.org/page/Aluminium* where we can know that the Japanese name is アルミニウム.
- Putting all things together, a application looking for アルミニウム could get its environmental impact, without any human intervention.

More info

Reference flow: Primary Aluminum Ingot

Same as: aluminum

A piece (i.e., specific collections of nearby molecules) of aluminum; typically, these will be tightly bound into solid form.

from http://sw.opencyc.org/concept/Mx4rvVi2O5wpEbGdrcN5Y29ycA

Same as: Alluminio

Aluminium or aluminum is a silvery white member of the boron group of chemical elements. It has the symbol Al and its atomic number is 13. It is not soluble in water under normal circumstances. Aluminium is the most abundant metal in the Earth's crust, and the third most abundant element, after oxygen and silicon. It makes up about 8% by weight of the Earth's solid surface. Aluminium is too reactive chemically to occur in nature as a free metal.

from http://dbpedia.org/resource/Aluminium

Category of: chemical compound

Figure 6. Information gathered via linked data for Primary Aluminum Ingot

3.   Experiences from the development

Footprinted is developed using Open Source software and standards such as PHP, HTML, SPARQL, JQuery, ARC, RDF, XML and JSON[11]. Footprinted itself is released as Open Source[12], both for giving back to the software community but also as a way of increasing the transparency and openness of the service. The development of Footprinted.org is in the cutting edge of web technologies and as the tools have not yet matured the development work had its shares of problems.

3.1   Database performance problems

The first iteration of Footprinted is using ARC[13] and a relational database for storing the linked data triples. This setup worked well for individual resources, but poorly for more complex queries as accessing all resources and their carbon impact. Our solution has been to cache the repository into a relational database containing a summary of each resource (name, link, category, impacts…). Footprinted uses this table for the functions that need complex queries with good performance such as search. The complete information of a resource accessed using the URI is

---

[11] For more information and links see: http://footprinted.org/about/code

[12] The code for Footprinted.org can be found at: https://github.com/zapico/Footprinted under the Affero Gnu Public License (AGPL).

[13] ARC is a PHP library to use a regular relational database as repository for RDF triples. More info at: https://github.com/semsol/arc2/wiki

retrieved directly from the triples repository. In the future we see the need of migrating to a native triple repository such as sesame[14].

3.2 Performance of external repositories

One of the main pillars of linked data is the reuse of existing data repositories. So for instance, when using units such as Kilograms, Footprinted uses the existing QUDT ontology[15]. The problem is that when accessing the information of other repositories remotely, the performance of the application suffered, this is particularly true when making queries in big ontologies such as OpenCyc[16]. Our solution has been to duplicate locally these repositories and accessing them locally. We hope however that this will change in the future. In our own case, as seen in the previous point, we also hope that the performance of the API is good enough to avoid this problem.

3.3 Semantic errors

The environmental impact information being added to Footprinted does not have semantic information embedded. The data has to be parsed and interpreted for attaching meaning. This can be done automatically in many cases, for instance interpreting kg as the concept kilogram. But in many cases it can go wrong (see Figure 7), or there are unavoidable ambiguities, and manual parsing is needed.
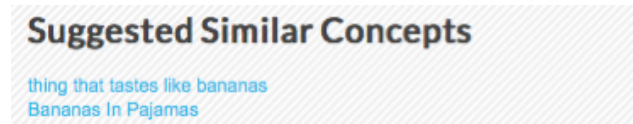
Suggested Similar Concepts

thing that tastes like bananas
Bananas In Pajamas

Figure 7: finding similar things to Bananas in the ontology gives back some unexpected results.

4. Existing and expected uses of the service

Footprinted is designed with different uses in mind. Some are already ongoing and others will start in the near future:

4.1. Serving as a repository for external applications

The aim of Footprinted is to be used as a repository for other applications that need environmental impact information. By using the existing API other web services can easily add environmental information to their services. The first and biggest user of Footprinted is Sourcemap.com[17], one of the pioneer web services using internet for sustainability. Sourcemap is a tool for making supply chains transparent. It has been used by organizations and companies for increasing their transparency (Bonanni et al, 2010) with the motto: "people have a right to know

---

[14] Sesame: http://www.openrdf.org/

[15] A ontology coming from NASA, more information at: http://qudt.org/

[16] OpenCyc is used as a general concept ontology.

[17] More information at http://sourcemap.org and http://sourcemap.com

where things come from and what they're made of " (Bonanni, 2011) The new version of Sourcemap uses Footprinted as repository for the environmental impact of materials and products, as it can be seen in Figure 8.
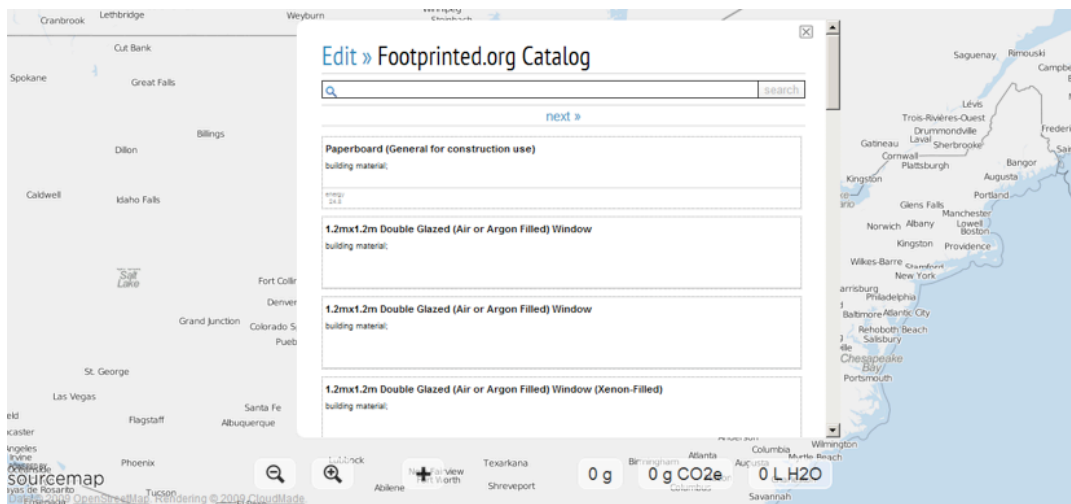

Figure 8. Sourcemap accessing Footprinted

4.2. Triggering innovation

Opening up the information can empower creators to develop new ideas and applications based in the data. If the data is available and usable for more people than just environmental experts, uses of the data can be crowdsourced. We will explore this by the organization of Green Hackathons[18], events where developers gather and develop during a limited amount of time interesting ideas connected to a goal, in this case sustainability.

4.3. Sharing LCAs and community of practitioners

Footprinted is initially populated with a collection of existing datasets[19] but the goal is to have practitioners and researchers uploading directly their results to the website. Social features like comments and ratings are also implemented, with the objective of providing quality assessment and a forum for improvement of the data.

5. Discussion and conclusions

In this article we have presented Footprinted, the ideas behind, the development experiences and the expected uses. Footprinted is now released and available online as a beta version. The registration necessary for using the API, creating data and using the social features is private, and an invitation code is needed. During the following months we will make a public release and start promoting the site.

---

[18] More information at http://greenhackathon.com

[19] Information about the datasets can be found at http://footprinted.org/about/data

We believe that the open approach of linked data can be a powerful tool for environmental impact data. Making this information available for a wider audience, easy to understand and use, will hopefully increase public awareness and create new unexpected uses of the data. The first usage of the API by Sourcemap is being positive, but it is still to be seen who will be use, how, how much, which services will be created and if they will have any impact.

It's also in the future if and how the community of practitioners will adopt Footprinted. Their usage of the application will be key in the future development of the site. We see many challenges ahead in assessing and visualizing data quality. But we believe that the focus on transparency and the open approach of creating and assessing in a social way instead of centralized verification can be more powerful.

As previously mentioned, environmental science and sustainability is heavily based on information. The innovations coming from semantic web and linked data open new opportunities in the field, both for practitioners and for widening up for other users. Footprinted is our way of exploring the presented ideas of transparency and openness into practice, and also a real service for making use of the opportunities ahead.

Bibliography

Allenby, B. (2000): Industrial Ecology, information and sustainability. Foresight/ vol.02, no.02, apr.00. The journal of futures studies, strategic thinking and policy.

Allenby, B. (2006): The ontologies of industrial ecology? In Progress in Industrial Ecology, An International Journal 3(1/2): 28-40

Barab, S. and Squire, K. (2004): Design-Based Research: Putting a Stake in the Ground, Journal of the Learn-ing Sciences,13:1,1 — 14

Baumann, H. and Tillman, A-M. (2004): The hitchhiker's guide to LCA : an orientation in life cycle assessment methodology and application. ISBN 91-44-02364-2

Bonanni, L et al. (2010): Small business applications of sourcemap. Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. Atlanta, Georgia, USA, 2010. 937. c2010 ACM.

Bonanni, L. (2011): Sourcemap: eco-design, sustainable supply chains, and radical transparency. Magazine XRDS: Crossroads, Volume 17 Issue 4, Summer 2011. ACM New York, NY, USA

Brown, A. (1992): Design Experiments: Theoretical and Methodological Challenges in Creating Complex In-terventions in Classroom Settings. The Journal of the Learning Sciences. 1992;2(2):141-178.

Davis, C., Nikolic, I. and Dijkema, G. P. (2010): Industrial Ecology 2.0. Journal of Industrial Ecology, no. doi: 10.1111/j.1530-9290.2010.00281.x

Goleman, D. (2009): Ecological Intelligence: How Knowing the Hidden Impacts of What We Buy Can Change Everything. Crown Business.

Raymond, E. (1999): The Cathedral & the Bazaar. O'Reilly. ISBN 1-56592-724-9. Available online at: http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/

Zapico, J. L., Brandt, N. and Turpeinen, M. (2010): Environmental Metrics. Journal of Industrial Ecology, no. doi: 10.1111/j.1530-9290.2010.00272