The Sustainability Hub: An Information Management Tool for Analysis and Decision Making

Steve Cayzer
Hewlett-Packard Laboratories
Bristol, UK
BS34 8QZ
steve.cayzer@hp.com

Chris Preist
Hewlett-Packard Laboratories
Bristol, UK
BS34 8QZ
chris.preist@hp.com

ABSTRACT

Sustainability is becoming an increasingly important driver for which decision makers – consumers, corporate and government – rely on principled, accurate and provenanced metrics to make appropriate behavior changes. Our assertion here is that a Sustainability Hub which manages such metrics together with their context and chains of reasoning will be of great benefit to the global community. In this paper we explain the Hub vision and explain its triple value proposition of context, chains of reasoning and community. We propose a data model and describe our existing prototype.

Categories and Subject Descriptors

H.4.3 [**Communications Applications**]: Information browsers

General Terms

Management, Measurement, Design,

Keywords

sustainability; metrics; context; chains of reasoning; community; information management; provenance

1. INTRODUCTION

Sustainability is becoming an increasingly important driver in all aspects of society. According the Hartman Group, "88 percent of consumers define themselves as participating in sustainable behaviors" [1]. Similarly, governments are interested in metrics that guide and monitor policy decisions; a good example might be companies' emissions data as demanded by the UK Carbon Reduction Commitment, which begins in April 2010. Non governmental organizations such as Greenpeace collect and use sustainability data to assess how effective other stakeholders are being, to identify laggards and possibly target for campaigning.

For industry, sustainability is simultaneously a constraint, a risk and an opportunity. Firms need to ensure that they are meeting legislative requirements for standards such as EnergyStar or Blue Angel, to design more environmentally friendly products and use this as a competitive market advantage, to change business models, or even enter new businesses that are emerging as sustainability factors (particularly climate change) become more of a business driver. When these decision makers – individuals or organizations –evaluate a course of action, they rely on metrics. There are many such metrics ranging from the scientific data published by the IPCC[1] through to online carbon calculators, carbon labels, lifecycle analyses and a myriad of ICT standards.

We cannot hope to agree on a global standard for all metrics, but what we can do is to help decision makers understand the provenance and semantics of these metrics. Consumer oriented metrics, like carbon labels [2], have drawn criticism for being opaque, confusing or even untrustworthy. Our assertion here is that metrics do not exist in isolation; they carry with them a great deal of context that is needed to fully understand them. We propose a *Sustainability Hub* to capture, manage and share this context. This Hub (as we will call it for brevity) will expose the quantitative reasoning behind metrics and allow users to create, examine and reuse this reasoning. Hub users will be empowered to examine, rate, annotate and enrich metrics so that a measure of community confidence can be applied to the metrics.

In the remainder of this paper we will explore the challenges of managing metrics, and will show how the Hub mechanisms of context, chains of reasoning and community will address these challenges. We propose a data model and describe a prototype that illustrates our approach.

2. Challenges for sustainability metrics

Sustainability data, more so than many other types of data, suffers from being more 'questionable'. This is for two reasons

1 .The art of collecting the needed data is in its infancy. This means that much of the data is missing, contradictory, has inbuilt assumptions associated with its collection method or usage, etc.

2. Different stakeholders have different motivations, which means that consciously or unconsciously, data collection and usage by an organization may be skewed. For example, a company might be keen to appear as 'green' as possible, while a pressure group may want to demonstrate the importance of their particular issue.

SIGMETRICS '09, June 15-19, 2009, Seattle, WA, USA.

[1] Intergovernmental Panel on Climate Change http://www.ipcc.ch/

To make some of these issues apparent let us consider the quote which appears on the CFP of this workshop:

"According to the Climate Group, the ICT sector is responsible for 2% of the worldwide carbon footprint"

When one really tries to understand this data, questions abound. One might ask 'where did this figure come from?' (the original source is Gartner[2]) or indeed 'what is ICT' (does this include mobile phones? Communications infrastructure? Entertainment consoles? TVs?). One might also ask how much of ICT's lifecycle is included (eg mineral extraction, production, distribution, use and end of life). In fact, even the notion of a carbon footprint is not necessarily intuitive. Typically people are concerned with the global warming potential, so other gases such as methane and nitrous oxide are also relevant, and converted to a CO2 equivalent (CO2e) over a certain timescale (eg 20 years, 100 years). Many metrics carry with them some sense of uncertainty and it would be good to know how confident we are in this figure. Is there a lower or upper bound? What year does this relate to? What studies confirm or refute this figure? The list goes on.

If one can include, along with metrics, some information on the methodology, caveats and assumptions, together with the provenance and community rating of the source, then decision makers will be better able to reach an informed and appropriate conclusion. In essence, making decisions requires access to robust, provenanced, verifiable, quantitative data

Thus, we can see that there are many difficulties with accessing, using and sharing sustainability data. We categorise these difficulties as follows:

- **Access** – extraction of metrics from complex documents which contain detailed environmental analyses.
- **Comprehension** – capture and presentation of contextual information needed to understand the metrics.
- **Integration** – the appropriate combination of sustainability metrics in order to reach a conclusion.
- **Comparison** – of (potentially contradictory) metrics relating to the same underlying product, process, market or behavior
- **Sharing** – a conclusion of a metric-driven analysis
- **Validation** – allowing a user, or a community, to assess the reliability of a metric or conclusion
- **Community enrichment -** enabling and encouraging users to move from passive consumption to active enrichment through data creation, feedback and annotation

We believe there is a need for an innovative solution to address these challenges.

3. The Sustainability Hub Vision

The Sustainability Hub is our vision for addressing these problems. Put at its simplest, the Hub is a way to gather, manage, leverage and share sustainability data to support a global community. The Hub can be thought of as an ecopedia for sustainability data, providing these 3 key values:

The Hub manages sustainability information together with its **context**.

[2] http://www.gartner.com/it/page.jsp?id=503867

Users can search for data items of interest, examine their derivation, and combine them in a variety of ways to support their own **chains of reasoning**.

This new information is in turn made available to other Hub users in the **community** who enrich the information with feedback, annotations and ratings

We ground our vision in a number of use cases which will be used to illustrate Hub features in the remainder of this paper.

1. **Interactive LCA**: Life Cycle Analysis (LCA) is a technique for calculating the lifetime impact of a product or service (for example a printer), from mineral extraction to end of life. It is thorough and well grounded, but communicating the results effectively is a major challenge. The Hub presents the results in an interactive, understandable and reusable form.
2. **Scenario Analysis**: Company strategists are keen to explore the impact of likely policies (eg carbon cap and trade) on their activities. By providing an intuitive way to perform 'what if' analyses, the Hub makes it possible to assess the likely impact under a range of different assumptions and predictions
3. **Structured Discussions**: The environmental debate is engaging consumers from all walks of life. The Hub captures the reasoning that leads to different claims being made, together with the context which makes effective analysis and comparison possible. These structured discussions are stable, reliable and reusable, as opposed to blog comments and forum discussions for which redundancy, repetition, mistakes, and transience are all too typical characteristics.
4. **Energy Plans**: Governments are being urged to make 'energy plans that add up' [3]. The Hub allows public bodies and consumers alike to examine the mix of energy options available, how much (and why) each technology is expected to contribute (much of this will necessarily be coarse grained estimation) and thus to formulate, explore and communicate alternative plans.

3.1 Context

Sustainability data does not exist standalone. Rather, in order to interpret and use this data in a principled manner, one needs a considerable amount of contextual knowledge. Context is the surrounding information that is needed to make sense of a figure such as:

- **Provenance**. Where does this data come from? Is it derived from a peer reviewed piece of research, a respected NGO or corporate report, or possibly a (named) Hub user?
- **Certainty**. Is this piece of data known with some level of certainty (perhaps through real world measurements) or is it simply a 'finger in the air' estimate?
- **Range**. Is there a lower and upper bound to this figure? This will help enormously in performing sensitivity analyses
- **Scope**. Scope can be both spatial and temporal. Spatial scope relates to geography: for example global or regional (European vs United States). Temporal scope allows us to differentiate current figures from historical ones, from multi year averages, and from future based estimates (ie predictions).
- **Caveats**. In many sustainability assessments the authors attach caveats. For example, in an EU study on printing [4],

there is a caveat around the permissible generalization of the assessments due to product specific assumptions.

- **Assumptions**. Often, it is necessary to assume something about the world in order to calculate a figure. Typically this is a simplification; for example we may assume that a piece of equipment will need no maintenance over its lifetime; or that consumer behavior is simply governed by price, or that all long distance employee travel is by plane.
- **Parameters**. Most data are dependent on a number of quantitative assumptions, or parameters, such as typical lifetime of a printer, or predicted carbon market price.
- **Commentary and Explanation**. Sustainability data is not always intuitive. In many cases it is necessary to provide commentary and explanation to help consumers understand the data (indeed, this is why most LCA reports are so long!)

3.2 Chains of reasoning

Sustainability data rarely exists in isolation, as we have seen. However, in addition to the surrounding context needed to make sense of a figure, most data depends on, or is derived from, other data. We call this derivation a *chain of reasoning* which is comprised of quantitative data, mathematical operations, and the supporting contextual information.

For example, the environmental impact of a printer is the sum of its manufacturing, distribution, use and end of life impacts (including possible recycling). Paper consumption is one part of this, and this in turn is derived from 2 figures; the amount (weight) of paper consumed, and the impact of that paper. Although the first figure may be observed directly (or estimated), the latter figure will be derived from what is likely to be a complicated set of figures, together with their own caveats, assumptions and parameters (see for example [6] and [7]). The Hub provides chains of reasoning which capture this complexity in a quantitative, grounded, illuminative and interactive manner. Making these chains of reasoning explicit, open to examination, shareable and reusable will provide enormous benefit to those with an interest in publishing, using and understanding sustainability. Typically a chain of reasoning will take the form of a mathematical formula eg (to take a simple example):

```
CO2_emissions
 = (tonnes_production) * (CO2_emissions_per_tonne)
```

The Hub will allow users to:

Examine; a chain of reasoning; see how a figure was derived from other data

Create; their own chain of reasoning and share it with the community

Enrich and Reuse; add more data or context to a chain of reasoning; build on it to extend the chain.

Comment and annotate; record discussion, debate and peer review

3.3 Community

In a web environment, data should be open, provenanced and subject to peer review and commentary. Community feedback is vital. Users will want to know if a figure is generally perceived as reliable and useful. Comments, annotations and ratings will allow users to collaboratively perform a sort of community peer review. Going beyond this, users should be able to reuse and build on the published chains of reasoning, to create new insights and conclusions. Thus the Hub will support and encourage community enrichment of sustainability data. Users can:

- *Examine, create, extend and reuse* a chain of reasoning
- *Create and publish* more data
- *Link* existing data together
- *Annotate*; chains of reasoning, data and even other users
- *Rate*; collective decisions about the most reliable and useful data, users and chains of reasoning.

4. The Sustainability Hub Prototype

Using the design principles outlined in this paper, we have built a Sustainability Hub prototype. We have applied it to a number of cases, including an EU study on the impact of printers [4].

4.1 Hub Data Model

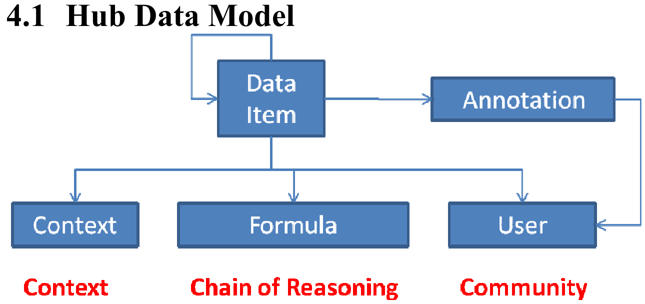

Fig 1. Sustainability Hub Data Model

Figure 1 shows a simplification of the data model which supports our Hub prototype. Central to the model is a DataItem which is supported by Context (there are different types of context, of course). The Formula underlying the DataItem comprises the top level of that DataItem's chain of reasoning. Each Formula has a number of constituent DataItems, which in turn have their own formulae and this tree extends until 'leaf' DataItems, which are simply discrete values supported by context. DataItems are contributed by an user, and may have annotations which are contributed by the same author or by other users. Thus community support is the third pillar of our data representation.

DataItems may be linked to other DataItems. This can be due to explicit relationships (they are involved in the same chain of reasoning), by user annotation, by indirect linkages or by calculated similarity.

Figure 2 shows how the Hub supports putting these figures into context. In this model the impact of printing is dominated by the manufacturing, paper and (especially) use phase (eg electricity)

4.2 Context and Chains of Reasoning

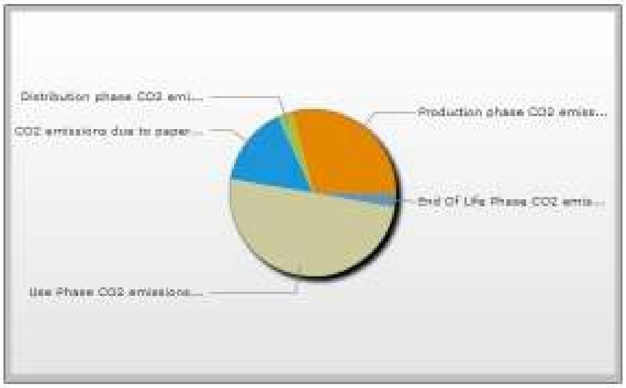

Figure 2: Context diagram for printer lifecycle impacts

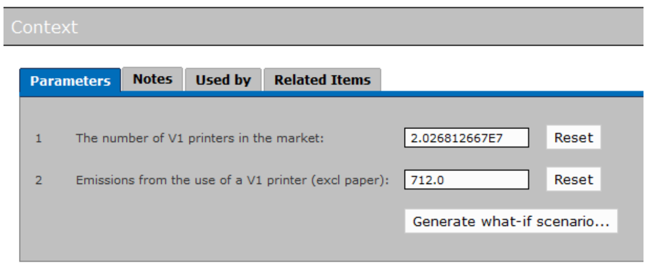

Figure 3: Tabbed view of context

Figure 3 shows how the diversity of contextual information can be presented in an intuitive way to the user using a tabbed view. Note that the parameters can be changed by the user in order to perform a 'what if' analysis.

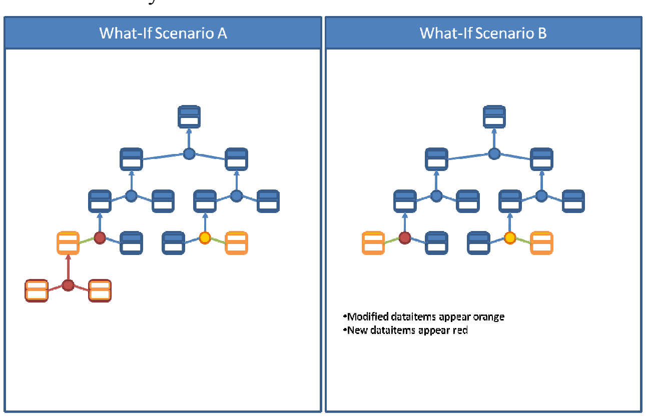

Figure 4: Visualization and comparison of 2 chains of reasoning

Figure 4 illustrates one approach to visualizing and comparing chains of reasoning. We are investigating a range of such approaches to cover a variety of data scales and user needs.

5. Related Approaches

5.1 Sustainability Web Portals

There are a number of sustainability portals in existence. For example, Wiser Earth[3] and Ecomotion[4] are green social networking sites, which deal with catalyzing a global community hence playing a synergistic role to the Hub. Wikiagreen[5] offers a community powered source of sustainability information, but is focused towards the qualitative and subjective content that is again complementary to the Hub. The CITRIS climate navigator[6] offers a bespoke innovative visual scenario analysis tool (Gas-CAP) which is analogous to the sort of functionality we are building to explore chains of reasoning on the Hub.

5.2 Sustainability Data Sources

Sources of provenanced, quantitative sustainability data already exist on the web. For example AMEE[7] is a neutral source of data for carbon footprints, while ecoinvent[8] is a source of life cycle inventory information. Such sources are key enablers for Hub functionality, but do not themselves offer user interaction with the chains of reasoning built from their data.

5.3 Software tools

Sophisticated tools like SimaPro[9] allow users to perform lifecycle analyses and track some of the contextual information we have described above, such as uncertainty and temporal scope. However, such tools are complex pieces of software, requiring expert users. Their results can often be difficult to interpret and they operate in a closed (desktop) model. The Hub is concerned with a wider range of analyses, offers more intuitive interaction with chains of reasoning and, perhaps most importantly, involves a global user community through the web. In our opinion, it is crucial to publish and share sustainability metrics so that consumers can explore, understand, validate and interact with it. We believe that this is what will allow the Hub to scale globally.

5.4 Online calculators

In stark contrast to LCA software, the proliferation of online tools such as carbon calculators[10] has allowed non expert users to access sustainability analyses in an intuitive, visually appealing way. The Hub uses this approach as a design goal; however it is designed to tackle a wide range of quantitative reasoning tasks, not just carbon footprints. In addition, the online calculators use a wide variety of parameters and assumptions that reduce consistency and transparency [5]. The Hub captures and exposes the context underlying the calculations, in a way that can be examined, commented upon, tested and even changed. Through feedback and annotation, the Hub provides community rating on those analyses which are most (or least) reliable and trustworthy.

5.5 Context Handling

Central to the notion of the Hub is a data representation format capable of handling provenance, uncertainty and other context. For the Hub, provenance is more about the transparency of the source rather than workflow management as is the focus in

---

[3] http://www.wiserearth.org/

[4] http://www.ecomotion.org.uk/

[5] http://green.wikia.com/wiki/Wikia_Green

[6] http://www.citris-uc.org/climatenavigator

[7] http://www.amee.com/

[8] http://www.pre.nl/ecoinvent/

[9] http://www.pre.nl/simapro/

[10] Google offers one at http://www.google.co.uk/carbonfootprint/

eScience [8]. Nevertheless, some of the techniques for attribution of data transformations [9, 10] are clearly relevant to context handling of complex chains of reasoning.

The semantic web [11] is a framework for data representation on the web. It handles relationships, provenance and context in a flexible and extensible manner. There is also the possibility of using ontologies (formal logic based models) to represent chains of reasoning. But the semantic web also suffers criticism of being over-complex and brittle [12]. Semantic web databases are still nowhere near as common place as their relational predecessors. Nevertheless, the semantic web could provide relevant data representation and exchange technologies.

5.6 Visualization

An important design criterion for the Hub is the use of innovative visualisation techniques to explore complex data. It is important that the Hub interface is intuitive, interactive and comprehensible without sacrificing rigor and completeness. The approach shown in figure 4 is a graph based exploration, which is one useful technique but it will not be appropriate for all needs. There is ongoing, relevant research on visual metaphors for communicating environmental impact [13] as well as prototype systems like Google's PowerMeter[11] .

Another challenge is search and browse interfaces. Faceted search [14] is one approach which uses orthogonal dimensions of metadata (topic, location, unit, time, impact category and so on) to effectively narrow down browse over global scale collections.

5.7 Data privacy and confidentiality

In order to build an industry wide picture, it is necessary for the Hub use data that may be company confidential. One possibility is for such data to be used internally to create aggregate statistics, which themselves are not sensitive and may safely be exposed. There are relevant techniques that allow the publication of such statistics while protecting both the initial data [15] and its source [16]. Although the Hub is expected to be a trusted intermediary, privacy guarantees are likely to enhance takeup.

6. Discussion and future work

We are building our prototype in consultation with our partners in Hewlett-Packard's printing business. The intention is to mature the Hub into an enterprise, and ultimately global-scale application, which can be used to inform product design and strategic decision making both inside and outside the company.

The future work involves a number of interesting research challenges. Dealing with uncertainty in a quantitative yet intuitive way is still a good example. A related challenge is that of visualization. We are investigating modalities for users to explore data at scale without being overwhelmed. These and other research activities supplement ongoing work to build up a critical mass Hub community of active users.

7. ACKNOWLEDGMENTS

Our thanks go to Tom Wilcox who provided valuable insights in the preparation of this paper.

[11] http://www.google.org/powermeter/index.html

8. REFERENCES

[1] The Hartman Group 2009 Sustainability: The Rise of Consumer Responsibility

[2] Carbon Trust 2008 Working with PepsiCo and Walkers: Product carbon footprinting in practice. Case Study CTS058 http://www.carbontrust.co.uk/Publications/

[3] Mackay, D. 2008 Sustainability Energy - Without the Hot Air UIT ISBN-10: 0954452933

[4] Fraunhofer IZM and PE Europe 2007 EuP Preparatory Studies 'Imaging Equipment' (Lot 4) Final Report on Task 5 'Definition of Base Cases' November 2007

[5] Padgett, J. Paul, Steinemann, Anne C., Clarke, James H., Vandenbergh, Michael P. 2008. A comparison of carbon calculators. Environmental Impact Assessment Review 28 106–115

[6] Framework for the development of Carbon Footprints for paper and board products CEPI (Confederation of European Paper Industries) report September 2007

[7] Larsen, H.F., Hansen, M.S., Hauschild, M. 2009. Life cycle assessment of offset printed matter with EDIP97: how important are emissions of chemicals? Journal of Cleaner Production 17 115–128

[8] Simmhan, Y. L., Plale, B., and Gannon, D. 2005. A survey of data provenance in e-science. *SIGMOD Rec.* 34, 3 (Sep. 2005), 31-36.

[9] Cui, Y. and Widom, J. 2003. Lineage tracing for general data warehouse transformations. *The VLDB Journal* 12, 1 (May. 2003), 41-58.

[10] Buneman, P., Chapman, A., and Cheney, J. 2006. Provenance management in curated databases. In *Proceedings of the 2006 ACM SIGMOD international Conference on Management of Data* (Chicago, IL, USA, June 27 - 29, 2006). SIGMOD '06. ACM, New York, NY, 539-550.

[11] Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., and Stephens, S. The semantic web in action. *Scientific American*, (December 2007).

[12] Shirky, C. 2003 The Semantic Web Syllogism, and Worldview. http://www.shirky.com/writings/semantic-syllogism.html

[13] Birt, A. (2006) Context Connection:Visual Indication of Food Backgrounds Thesis: Man + Humanity Masters Design Academy Eindhoven, The Netherlands

[14] Hearst, M. (2006) Clustering versus Faceted Categories for Information Exploration, Communications of the ACM, 49 (4).

[15] Xiong, L., Chitti, S., and Liu, L. 2007. Preserving data privacy in outsourcing data aggregation services. ACM Trans. Interet Technol. 7, 3 (Aug. 2007), 17.

[16] Ringberg, H., Applebaum, B., Freedman, M.J., Caesar, M., and Rexford, J. 2009 Collaborative, Privacy-Preserving Data Aggregation at Scale. Cryptology ePrint Archive, Report 2009/180 http://eprint.iacr.org