
Webbox+Page Blossom: Exploring Design for AKTive Data Interaction

m.c. schraefel, Dan Smith, Max Van Kleek

Electronics and Computer Science
University of Southampton,
Southampton, UK
mc,ds,emax @ ecs.soton.ac.uk

Abstract

We give away our data to multiple data services without, for the most part, being able to get that data back to reuse in any other way, leaving us, at best, to re-find, re-cover, retype, remember and re-manage this material. In this work in progress, we hypothesize that if we facilitate easy interaction to store, access and reuse our personal, social and public data, we will not only decrease time spent to recreate it for multiple walled data contexts, but in particular, we will develop novel interactions for new kinds of knowledge building. To facilitate exploration of this hypothesis, we propose *Page Blossom* an exemplar of such dynamic data interaction that is based on data reuse via our open data platform Webbox + Active (active knowledge technology) lenses.

Author Keywords

Data interaction, Data sharing, Semantic Web

Copyright is held by the author/owner(s).
CHI'12, May 5–10, 2012, Austin, Texas, USA. ACM
[978-1-4503-1016-1/12/05](#).

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

We create data in increasing amounts daily: we post: a comment to a blog; a question on a forum; a URL to twitter; an observation to Facebook; a run in Runkeeper, our sleep at Zeo, our heart rates in HRM, our address and payment details at multiple stores, annotations in an eBook, bookmarks in lists. All that fantastic data about our lives we freely give away to what are, effectively, data silos: it is difficult, if not impossible, for a non-programmer to retrieve and reuse this data again in any other context. There are other consequences to these silo keepers: tweets disappear over time; Facebook posts are often difficult to re-find.

Various applications like snippet keepers such as Yojimbo¹ and Evernote² or bookmark/annotation services like Diigo³ do support data capture and sharing. They require us however first to actively deposit material into yet another container and second, those containers are also their own data silos or "walled gardens"[6]. In using these largely corporate social- or cloud-oriented applications, we almost

¹ <http://www.barebones.com/products/yojimbo>

² <http://www.evernote.com>

³ <http://diigo.com>

immediately lose (access to) the value of the data we have created: we cannot re-use it beyond the particular limits of its initial deposit. We may think we are pouring these streams into a bucket for future use, but in reality, we are pouring them into a sieve. The cost of the sieve is lost opportunities for personal, social or public knowledge building. I may wish to look at the overall number of tweets my colleagues and I have made around a particular topic relative to what eBooks we have been reading at that time. That data exists, but there is no facile way to query, store or manipulate this information because of the walled garden application data design.

In contrast to corporate data capture, over the past few years governments have been focused on Tim Berners-Lee's call to "raw data now" on institutional scales [2]. Lead by the US (data.gov) and UK (data.gov.uk) terabytes of spreadsheets of departmental statistics are being poured into the open, not only for greater transparency, but to support opportunities to create new knowledge by being able to pull together this data for comparison and analysis. This interlinking is facilitated by use of (semantic) web standards.

Representation of government open data to date, has largely been lead by private companies producing unfortunately only fixed interaction Web 2 style mashups around, for instance, crime rates⁴ or GDP spend per area,⁵ or hospital performance per region.⁶ Citizens cannot readily mix and match these open data / closed interaction sources to create our own queries, nor can we add new data to these sources.

⁴ <http://www.police.uk>

⁵ <http://www.usgovernmentspending.com>

⁶ <http://www.sepho.org.uk/extras/maps/NHSatlas2011/atlas.html>

Intriguingly, we see governments on the one hand pouring resource into opening their data in an attempt to exploit the opportunities of interlinking data, while the sheer scale of the data dump makes doing anything with the data from discovery to integration unwieldy for non-coders. On the other hand, corporate social networks, shopping and similar sites have been making highly interaction constrained, user-friendly sites that are walled gardens. Citizens massively generate and willingly toss our personal, social and public data over and into their data silo walls, though never to be fully recoverable by us, again.

We propose an alternative approach: webbox + AKTive (actionable knowledge technology) lenses⁷. Webbox may be thought of as one's personal cloud that we have implemented [7] to live, interlinked, on one's own devices and to communicate with other webboxes in one's network. Webbox leverages the advantages of web standards and open data interlinking (1) to support easy data storage of any data type, and (2) to enables private, social and public data sharing/querying to and from one's webbox (and webbox network) via AKTive lenses.

Our goal with webbox has been to develop a platform that will let us explore the hypothesis that, if we have tools to let us maintain control of all the data we create, and have ways to blend that with relevant social *and* public data, we will support far greater social benefit through active, sharable knowledge building and knowledge sharing than is currently possible with walled garden data, and poorly accessible open data in general. In the following sections we show how we propose to explore this hypothesis via the webbox and an example AKTive lens we call *page blossom*. *Page blossom* is a test case to show how tapping into unsiloed, sharable data

⁷ <http://webbox.ecs.soton.ac.uk>

can enrich everyday activities such as web browsing. Through this example we hope to demonstrate how new kinds of knowledge building interactions are possible once we reclaim our own data.

Webbox Data-Interaction Overview

To situate the page blossom data interaction, we first briefly overview the Webbox platform. Webbox is a platform that uses web standards to support

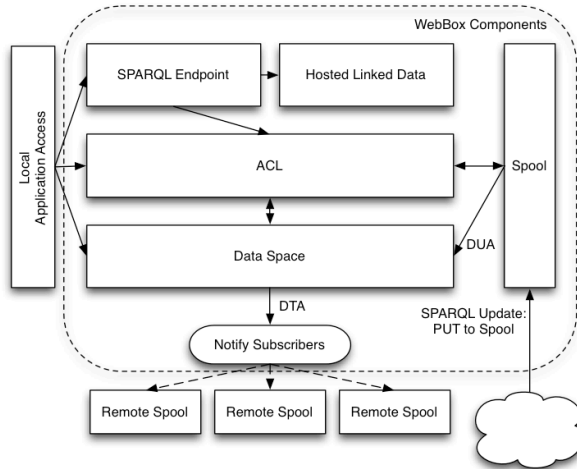


Figure 1 Webbox architecture from [7]

decentralized data storage and sharing. Its architecture is described in detail elsewhere [7]. The key purpose of a webbox is to *support our interaction with our data* via four main services, highlighted from the webbox architecture, described below:

1. The Secure Store: Data Sea not Silo

Unlike current cloud services that may or may not be encrypted, but will have company-accessible back doors, Webbox is a secure, personal cloud. One does not have to trust a third party with one's data. Also, most cloud services are focused on sharing files; we are

interested in structured and semi-structured data. Anyone authorized can deposit data (like a name, an address, a set of statistics) into or read data from that webbox. This exchange is accomplished via the publicly writable SPARQL endpoint in the webbox (as per webbox components, Figure 1). Sparql is the semantic web standard for metadata querying and modification of (semantic) data stores over the web.

2. Data Sharing

Data in a webbox can be shared with any other webbox. Webboxes send references (URLs, specifically)

of the shared data to the webboxes of collaborators. The collaborators' webboxes then subscribe to those URLs, and receive notifications whenever they are updated. The combination of WebACL access control with the SPARQL Update protocol also enables collaborators to edit data that is shared. When data is updated, all subscribers' webboxes are notified, and their webboxes fetch the new version, which can then be shown to subscribers in an appropriate context/lens.

3. Enriched Interactions: Aktive Lenses

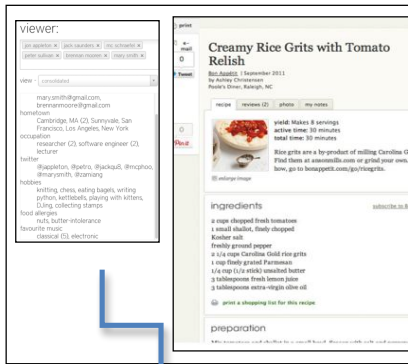
AKTive (actionable knowledge technology) lenses are the interaction layer of webbox (described as Local Application Access in Figure 1). They act similarly to how database views represent aspects of a database. Webbox AKTive lenses differ from views in two ways. First, they may run not only over a structured data source like a database or open data store, but also over semi-structured data, like a web page or social data service. Second, lenses will often have widgets (functions) embedded in them to provide multiple kinds of augmentations to the original data via interactions with those widgets.

4. Notebooks: Support for Work in Progress

Our goal with webbox is to support dynamic interaction with data to gather, annotate, share, blend and build data into new knowledge. Notebooks as typically spaces for our "work in progress" [5] represent collections of interoperable AKTive lenses and lens functions in support of a work in progress context such as managing projects, interacting with social networks and the possible interactions among these

Page Blossom: Interacting with Web Data

In this paper, we focus in particular on *the AKTive lens*,



the interaction layer within the webbox. To demonstrate the attributes of such interactions, we focus on a scenario, *page blossom*, in which a person annotates static content on a web site with information she needs about organizing a dinner party retrieved from her social network, including dynamic information provided by one of the attendees. The numbers in the scenario correspond to numbers of screenshot figures (on the left), as well as to a description of the brief technical underpinnings elaborated in the section that follows.

Scenario: Preparing a Grits Social Brunch

I have just come back from North Carolina where I was introduced to "grits." I plan to have a Sunday brunch with some friends where I wish to serve grits as a course. I seek out a recipe online. With the miracle of the web I find several quickly. Normally, I would be stopped here and take my chances in hoping to have found a recipe everyone will like. But I have a webbox lens, *page blossom*, which enhances this page for my brunch purpose.

On activating *page blossom* - in this case, pushing an extension button in a web browser (0) - a sidebar opens where I can quickly look up the individuals I want to invite. (1) I ask Webbox to produce a "consolidated" view of my invitees (Figure 2), which presents a unified and de-duplicated view of all the properties listed among their various social network profiles. This consolidated view makes it easy to identify whether and how many people have listed any explicit food aversions or allergies. I quickly discover that one of my friends has listed dairy in their "food allergies" property.

Not being an expert on Southern cooking, I realize I need more information about what to use to substitute

for butter in the recipe. The source website unfortunately does not list any alternatives, and a cursory web search for "grits butter alternative" similarly fails to pull up any solutions. None of my friends have explicitly listed Southern cooking as an area of expertise. So, instead, I ask Webbox to remind me of their personal histories - specifically, the time spent in various states in the US - to see if any have spent any time in the South. While this would have been very tedious to aggregate the location histories of my friends' Foursquare⁸ check-ins, Dopplr⁹ trips, Google Latitude¹⁰ History, Facebook Timeline records manually, Webbox has already captured and stored this information locally and instantly derives a high-level visualization from these records (Figure 3) (2)

From this visualization, I realize that Jon has lived in South Carolina for most of his life, and so he is probably a good person to ask. To request this information from Jon while providing him the context of the recipe, I right click directly on the recipe in the page (Figure 4), and ask Webbox to create a special type of shared annotation, called a Request, which can be used to ask friends for information of particular types. Such request messages get rendered as forms for the recipient, so that they can quickly fill out responses, which can then be aggregated on the sender's end. While requests can be of various structured forms (multiple select, text, or particular values), in this case I simply create a "text" type Request and explain the situation. (3)

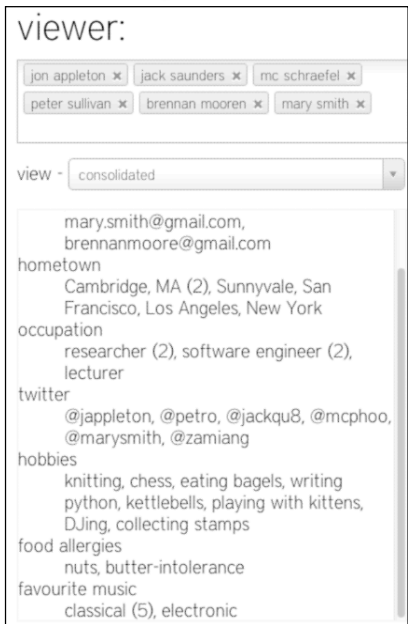


Figure 2 – Consolidated views of entities allows quick identification of properties. Duplicated values are shown with numbers (indicating order)

⁸ Foursquare – <http://foursquare.com>

⁹ Dopplr – <http://www.dopplr.com>

¹⁰ Google Latitude – <http://latitude.google.com>

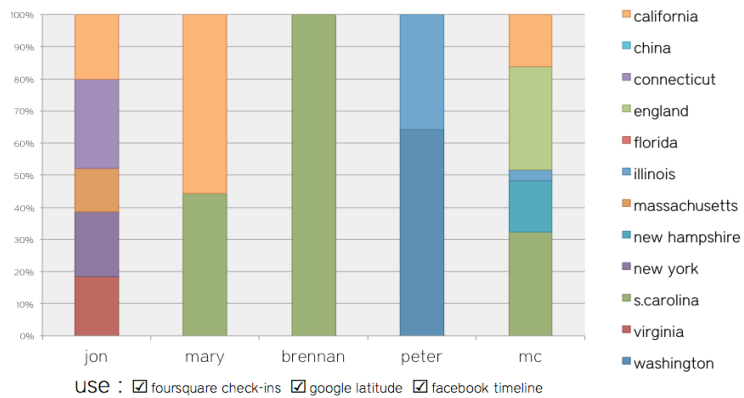


Figure 3 – Lens displaying how long friends have lived in different US states and abroad, as determined by location-related feeds from social sharing platforms such as Foursquare, Dopplr, Google Latitude, and FB Timeline

Jon receives a notification in his inbox (4) which he selects, taking him to the recipe page. Having found a video walk-through of a grits recipe that called for the use of sunflower oil instead of butter, he sends the link to the video back. This video then appears alongside the original request annotation in the source page (Figure 6) (5).

Technical Overview of Scenario Components

The page blossom prototype currently draws on the following components to enable the above interactions.

0. Capturing data in, on, and around webpages:

Unlike tools like Diigo which strictly provide textual webpage annotation capabilities to browsers, Webbox’s AKTive page blossom lens is designed to let users grab arbitrary content (structured – e.g., RDFa, lists, tables, or unstructured – text, images, etc) in pages, or add content to the page. Either way, these annotations become first-class Webbox objects which can be directly shared, retrieved, viewed and edited in any other AKTive lens context – such as rendered in a microblog view or in the Webbox explorer interface. In this way, information captured and created by users is designed to be shared, re-discovered, re-purposed and re-used as much as possible.

post a request:

jon appleton ✕

dear jch,

I'm thinking about preparing Grits thursday night but was wondering if you could advise on a butter substitute, because Peter has a butter-intolerance.

thanks
mc

type:
text

Send Cancel

Figure 4 - Sending a textual request for information from Jon, which renders as a response form embedded in the page context on his side.

- Social network sites** like Facebook or Google+ are (currently) used to bootstrap the user’s Webbox contact list, although contacts can so be added directly by their WebID. Information from the user’s friends’ social network profiles is imported and aggregated wholesale into Webbox so it can be visualized and made use of by any webbox lens.
- Deriving high-level views using social network histories as “sensors”**– Lenses like Figure 2 demonstrate how social network streams can be used as “sensor streams” and used to derive visualisations. In this example, each friend’s complete history of Foursquare check-ins, Google Latitude location events and Facebook Timeline/Places check-ins are used by the lens as to infer the approximate duration of each person’s stay within a particular US state or country.
- Direct social elicitation** – To reduce information friction associated with asking people for information, AKTive lenses supports “Request” messages within the information contexts that such needs occur. These requests are shared annotations that can request various kinds of input, from simple textual responses (which make them work like e-mail) to surveys and polls for specific, structured information such as availability, (e.g., Doodle¹¹) or opinions (e.g., formspring¹²).

¹¹ Doodle – <http://www.doodle.com>

¹² Formspring – <http://www.formspring.com>

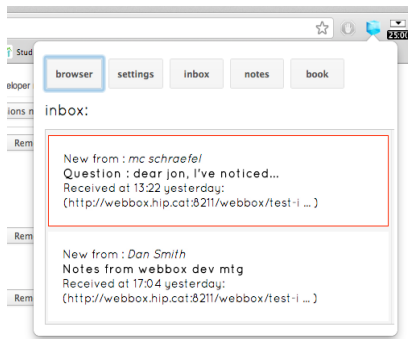


Figure 5 – Jon’s browser receives a message containing the Request for input

4. **Integrated messaging** – On the receiving end, new messages are queued up and displayed in the user’s browser in a notifications bar. For information items such as Requests that have an associated information context (e.g., a web site or lens), these requests also directly appear within the context (when allowed to do so) so that common information tasks can be addressed at once, reducing context switching.

5. **Page blossoming** – Jon’s response is attached directly to the original request for information, and made available next time I access the recipe page. Collectively, the ability to augment pages with structured and unstructured, static and dynamic content, including views derived from the user and the user’s social network means that sites that were formerly static become *surfaces* for collaborative sensemaking. This, we refer to as *page blossoming*.

Conclusion: A sea of Interactive Potential

At time of writing, interaction with our data is largely one way: out of our hands and into data silos, curtailing reuse and so restricting knowledge building opportunities. The page blossom example shows how interaction with even common information types like web pages gains enriched novel interaction possibilities being able to access and re-purpose data that, rather than residing only with third parties, also stays in our hands. While we have described new interactions supported by only one lens and associated widgets that run on this unsiloed webbox data, many more are possible. We are, for instance, testing a transparent

lens to capture data to a webbox simultaneously as we add it to any site and a faceting lens to let us explore our webbox data according to any parameter from deposit time to data type to concept. We present this Work in Progress note now to demonstrate the potential for new, data-oriented interaction design and, we hope, with our open webbox, AKTive lenses and notebook architecture, to facilitate community participation in design of and research in active data interaction.

References

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. *The Semantic Web*, 722-735, 2007.

[2] T. Berners-Lee. Tim Berners-Lee on the next Web. In TED2009, 2009. http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

[3] J. Hoart, F. Suchanek, K. Berberich, E. Kelham, G. de Melo, G. Weikum, F. Suchanek, G. Kasneci, M. Ramanath, and A. Pease. Yago2: a spatially and temporally enhanced knowledge base from wikipedia. *CACM*, 52(4):56-64, 2009.

[4] Miller, R. C., Chou, V. H., Bernstein, M., Little, G., Van Kleek, M., Karger, D. and schraefel, m. c. Inky: A Sloppy Command Line for the Web with Rich Visual Feedback. *UIST 08*, Monterey, CA, 131-140, 2008.

[5] m. c. schraefel. What is an analogue for the semantic web? *HT '07*, 123- 132, 2007.

[6] The Economist. The Web’s New Walls. Sep 2nd 2010, <http://www.economist.com/node/16943579>.

[7] M. Van Kleek, D. Smith, N. Shadbolt, and mc schraefel. A decentralized architecture for consolidating personal information ecosystems: The WebBox. PIM Workshop, CSCW 2012, Seattle, WA, 2012.

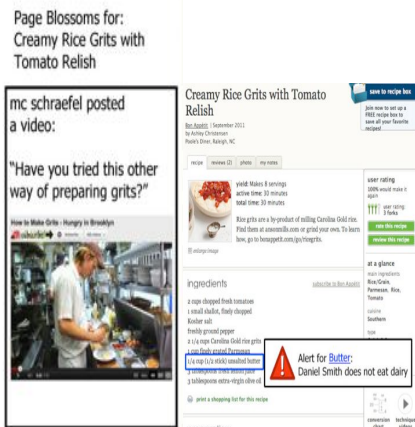


Figure 6 – Original recipe is annotated with Jon’s response