

A Framework for Dynamic Data Source Identification and Orchestration on the Web



Alexander Berezovskiy and Dr Leslie Carr

4th International Workshop on Web APIs and Service Mashups
European Conference on Web Services (ECOWS 2010)

1 December 2010, Ayia Napa, Cyprus

Problem

- Modern Web can be seen as a “universal data machine”
 - Multitude of web applications, services and data provides
 - Almost every user has vast amount of data stored online
- Data is often duplicated
 - Users need to register, provide their details, upload photos...
- Ideally, we would like a way to universally manipulate the data
 - APIs are different, functionality depends on data source
 - We need to know what data source to use
 - Make adjustments for every data source

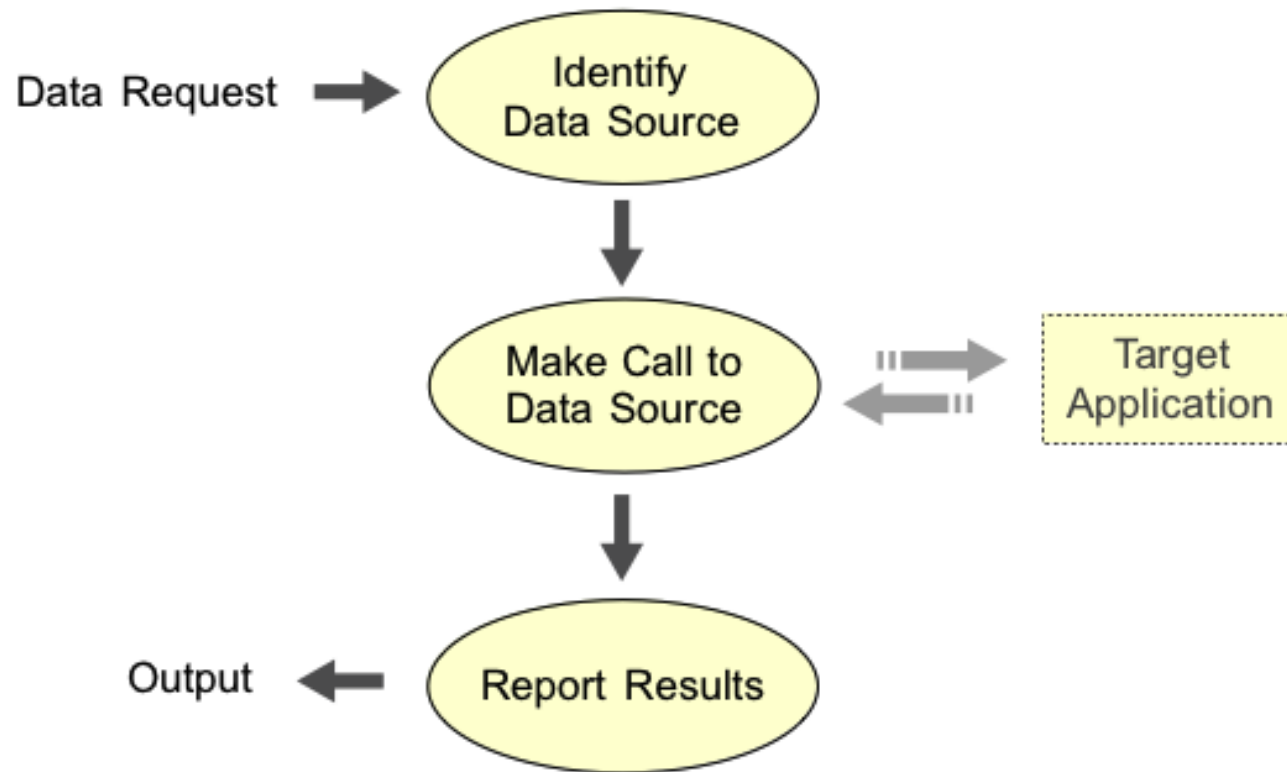
Current solutions

- OpenSocial
 - Focused on social applications
 - Requires adjustments in target applications
- Integrate every possible resource
 - Very time consuming
 - No guarantee the user will be satisfied with the choice
 - User interaction required to choose the right source

Proposed solution

- Identify the most appropriate data resource
 - Task is given, nature of required data is known
 - Without any user intervention attempt to identify most suitable data resource to perform the given task
- Execute the task (process data request)
 - When data resource is identified, hand the request over to the resource and report on the results (or, return results)
 - Allow full CRUD (create, read, update, delete) operations on data

Proposed solution (cont'd)

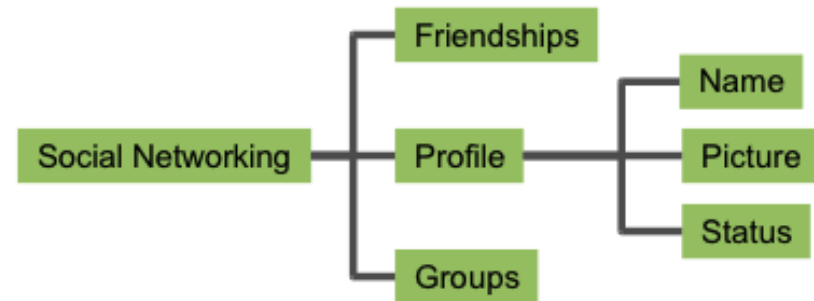


Data resource identification

- One visit to a web page can tell a lot about the user
 - Country, language, browser, operating system, ...
 - We assume all these parameters affect the user preferences and their choice of applications
- Use two-dimensional model:
 - Information about the user
 - **Country, Language**, Age, Occupation, Marital Status, ...
 - Usage information
 - **Browser, Operating System**, Web and Local Applications, ...
 - Some information can be obtained from a single HTTP request with no user intervention required

Data resource identification (cont'd)

- User information is grouped in a single entity called Environment
- Data can be structured as a tree:



- Identification algorithm:

Total score for data source a is defined as:

$$TSA(a) = TRS(a) + ERS(a) + TRAA(a)$$

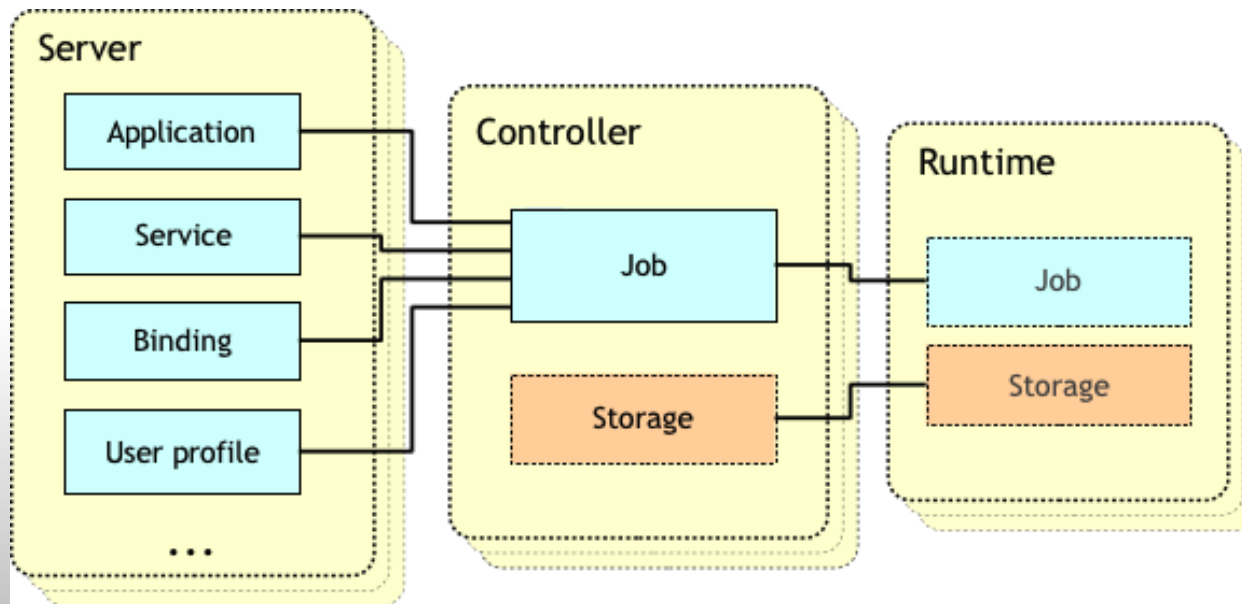
$TRS(a)$ - Total Relevance Score for a

$ERS(a, u)$ - Environment Relevance Score for a and user u

$TRAA(a)$ - Total Relevance Application-to-Application score for a

Data operations

- Each data resource serves data differently
- Data operations are performed by “bindings”
 - These are small chunks of code executed independently
 - We need at least one for each data resource
 - They can be written by anyone



Data operations (cont'd)

- Bindings return data in “raw” format
- The data can then be converted to almost any format
 - Currently available XML, JSON and Plain Text
 - Can be adjusted to serve RDF
- Example binding to return a user's name from Facebook:

```
import urllib
import simplejson

interface = {
    'fields': {'username': {'required': 'yes', 'type': 'text'}},
    'formats': ['html', 'xml', 'txt']
}


def run_binding():
    url = 'http://graph.facebook.com/' + str(job.input_args['username'][0])
    response = urllib.urlopen(url)
    user = simplejson.loads(response.read())
    return user['name']
```

Demo

Services / Nuvius

http://nuvius.com/service/ Google

Home Applications Service Map Web Tree Wiki First time here? [Sign in](#) | [Register](#)

 **nuvius**
preview

Search Applications:

[Cyprus / English](#)

applications services profile

Applications

[Recommended](#) [Most Recent](#) [Service Map](#)

Use **Service Map** to navigate around things available at **Nuvius**, and narrow down the list of recommended applications.

Business	Communication	Information	Multimedia	Productivity	Science & Education
Collaboration	Authentication	Encyclopedia	Audio	Bookmarking	Distance Teaching
Finance	Blogging	News	Station	Calendar	Research
Project management	Messaging	Article	Gaming	Event	Paper
	Social networking		Photo sharing	Contact book	
	Friend list		Video	Contact	
	News		Channel	Data storage	
	Profile			Folder	
				Notebook	
				Note	
				Office	
				Document	
				Presentation	
				Spreadsheet	


[About](#) | [Terms of Use](#)

Demo

Event / Calendar / Services / Nuvius

http://nuvius.com/service/calendar/event/ Google

Home Applications Service Map Web Tree Wiki First time here? Sign in Register




 **nuvius**
preview

Search Applications:

Cyprus / English

applications services profile

Applications

 [Recommended](#)  [Most Recent](#)  [Service Map](#)

Use **Service Map** to navigate around things available at **Nuvius**, and narrow down the list of recommended applications.



Productivity

Calendar

Event

- [Attendee](#)
- [Author](#)
- [Date](#)
- [Details](#)
- [ID](#)
- [Place](#)
- [Title](#)
- [Updated](#)




Data

-  [Data for Event](#)
-  [Generic interface](#)

Service ID: 35




Google Calendar

Google Calendar is a free time-management web application offered by Google. Its features include calendar sharing, device synchronization, SMS alarms and others. Users are required to have a Google Account in order to use the software.



 [Add to Library](#)  [Similar applications](#)  [Data for Event](#)

Hardtree

Hardtree is an online, integrated business and IT management suite. Hardtree comes with a multitude of modules including Service Support, Project Tracking and Sales Management and other.

 [Add to Library](#)  [Similar applications](#)  [Data for Event](#)

Tungle.me

Demo


Event / Calendar / Nuvius

http://nuvius.com/service/calendar/event/data

Google

Home Applications Service Map Web Tree Wiki



First time here? [Sign in](#) | [Register](#)

 Search Applications:

[Cyprus / English](#)

applications services profile

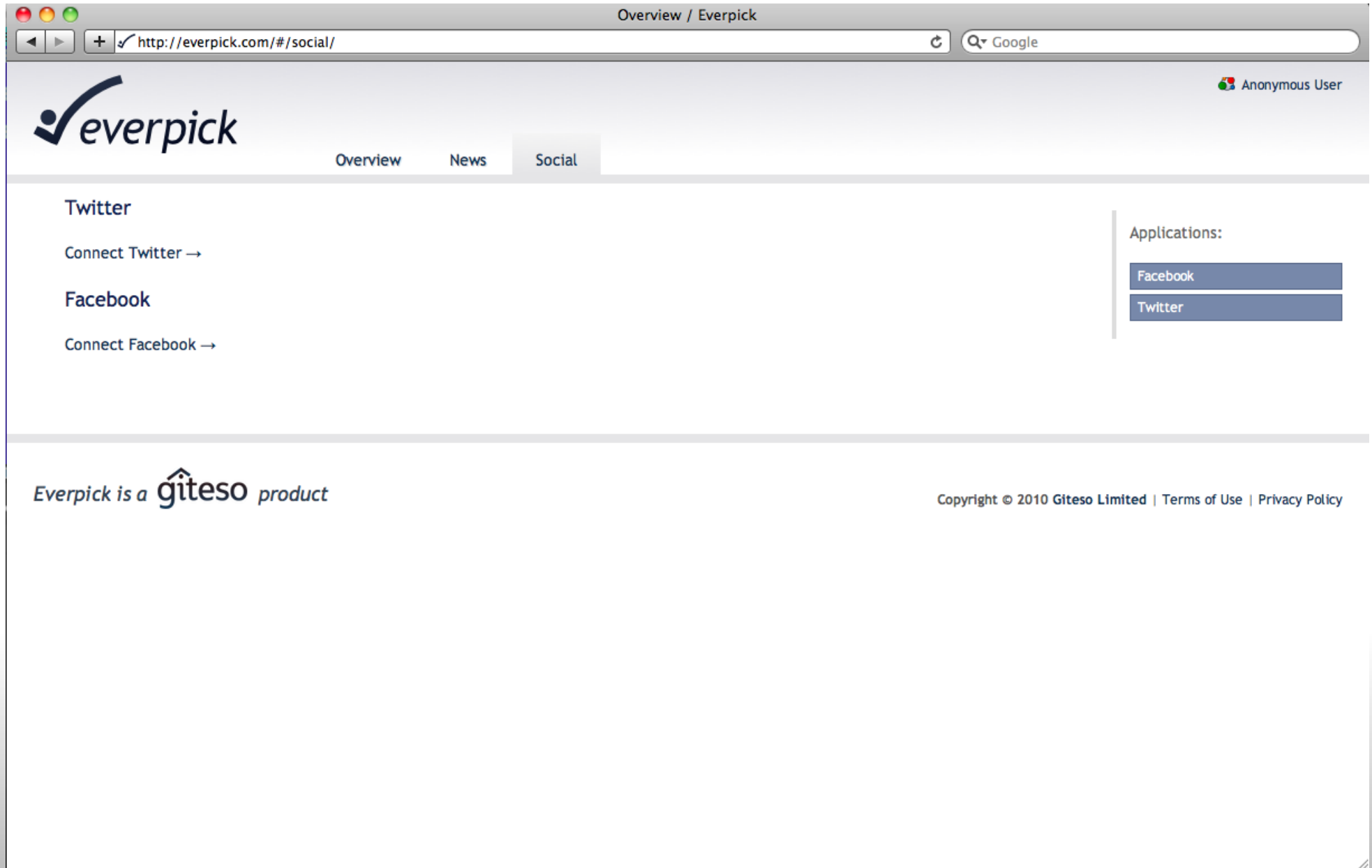
Data for Event

- Using application: ☐ [Google Calendar](#)
- Binding:  [GoogleCalendar by letoosh](#)
- Interface:  [Calendar / Event: Generic Interface](#)
- Format: HTML

The result is:

- attendee:
 - status:
 - picture:
 - name: letoosh@gmail.com
- title: Mashups'10
- author:
 - picture:
 - name: Alex Berezovskiy
 - email: letoosh@gmail.com
- updated: (2010, 11, 30, 11, 2, 56, 1, 334, 0)
- place:
 - latitude:
 - name: None
 - longitude:
- details: None
- date:
 - start: (2010, 11, 1, 9, 0, 0, 0, 305, 0)
 - end: (2010, 11, 1, 17, 0, 0, 0, 305, 0)

Demo



The screenshot shows a web browser window with the title "Overview / Everpick". The address bar displays "http://everpick.com/#/social/". The page features the Everpick logo (a checkmark icon followed by the word "everpick") in the top left. In the top right, it says "Anonymous User" with a small icon. Below the logo, there are three navigation tabs: "Overview", "News", and "Social", with "Social" being the active tab. The main content area is divided into two columns. The left column has a "Twitter" section with a "Connect Twitter →" link, and a "Facebook" section with a "Connect Facebook →" link. The right column has an "Applications:" section with two buttons: "Facebook" and "Twitter". At the bottom of the page, there is a footer that reads "Everpick is a giteso product" on the left and "Copyright © 2010 Giteso Limited | Terms of Use | Privacy Policy" on the right.

Overview / Everpick

http://everpick.com/#/social/

Google

Anonymous User

everpick

Overview News Social

Twitter

Connect Twitter →

Facebook

Connect Facebook →

Applications:

Facebook

Twitter

Everpick is a giteso product

Copyright © 2010 Giteso Limited | Terms of Use | Privacy Policy

Demo

The screenshot shows a web browser window with the title "Social Networking / Everpick". The address bar displays "http://everpick.com/social/". The page features the Everpick logo and navigation tabs for "Overview", "News", and "Social". The "Social" tab is active, showing a list of posts. Each post includes a user's name, a timestamp, and the text of the post. On the right side, there is a section titled "Applications:" with buttons for "Facebook" and "Twitter".

Time for a relaxing train journey back to Bath after a fun weekend at Southampton and Home.

Oliver Bills 30 November 2010, 11:22

Warming up for #online10 - "Aim for the Stars" no surprise its open data and the virtues of linking it #opendata #linkeddata #datagovuk

Nigel Shadbolt 30 November 2010, 11:12

"очень часто посещающие мысли :)"

<http://www.youtube.com/watch?v=bYw7P886upk>

Alina Klokova 30 November 2010, 10:32

Grr, reading documentation off a website which has just failed. Unhelpful!

Peter Brewer 30 November 2010, 10:17

#opendata Nigel Shadbolt calls for university computer science courses to prioritise open data and development skills - <http://bit.ly/danKz>

Applications:

Facebook

Twitter

Demo

The screenshot shows a web browser window with the address bar displaying `http://everpick.com/#/news/`. The page title is "Overview / Everpick". The browser's search bar contains "Google". The website's header features the "everpick" logo and navigation tabs for "Overview", "News", and "Social". The user is logged in as "Anonymous User".

The main content area displays a news article titled "Actor Leslie Nielsen Dies at 84" with a timestamp of "29 November 2010, 13:55". The article text reads: "An anonymous reader writes 'Famous actor Leslie Nielsen died of pneumonia at a Florida hospital Sunday evening. Leslie was renowned for his comedic roles in dozens of films and TV shows, such as The Naked Gun and Police Squad. His characteristic style and humor was always enjoyable, and he will be sorely missed.' Of course you might also remember him from The Forbidden Planet, which is classic sci-fi by any measure." Below the text are two share buttons: "f Share" and "share". A link "Read more of this story at Slashdot." is provided.

On the right side, under the heading "Applications:", there is a list of four items: "HabraHabr", "Google News", "Slashdot", and "Killer Startups".

Below the first article, a second article is partially visible, titled "Microsoft Patents Shape-Shifting Display" with a timestamp of "29 November 2010, 14:50". The text for this article begins: "Stoobalou writes 'In layman's terms, Microsoft's patent is for a special type of touch-screen display which includes a 'shape-memory' layer at its base. When activated by a special frequency of ultraviolet light, individual blocks — not-coincidentally the same size as a pixel on the display part — can be raised or lowered. lending the displayed image physical texture.'"

Demo

The screenshot shows a web browser window with the address bar displaying `http://alex.hardtree.com/#/calendar/settings/integration`. The page title is "Events / Calendar / Hardtree". The Hardtree logo is in the top left, and a user profile "John (logout), Settings" with "Messages (0), Notifications (0)" is in the top right. A search bar is located below the logo.

The main navigation bar includes icons for Home, Account, Administration, Calendar, Changes, Documents, Finance, Identities, Infrastructure, Knowledge, Messaging, Projects, Reports, Sales & Stock, and Service Support. The "Calendar" icon is highlighted.

The left sidebar contains links for "All Events", "Upcoming", "Calendar" (with sub-links for "Month", "Week", "Day"), "Settings", and "Integration" (which is the active section).

The main content area is titled "Calendar Integration" and includes a "Help" link. It shows "Active Resources:" with "None" listed. Under "Available Resources:", "Google Calendar" is listed with a "+ Connect" button. Below this, it states "Powered by Nuvius. Your Nuvius profile is Anonymous User".

A modal dialog box is open in the foreground, titled "Google Calendar needs your approval". It displays the "Name: Google Calendar" and "URL: http://www.google.com/calendar". At the bottom, it says "Please proceed to Google Calendar:" followed by a "+ Connect Google Calendar" button.

At the bottom of the page, the copyright notice reads: "© 2010 Giteso Ltd. All rights reserved. | Hardtree".

Demo

Month / Calendar / Hardtree

http://alex.hardtree.com/calendar/settings/integration/add/65#/calendar/month!date_year=2010&date_month=9

Google

John (logout), Settings
Messages (0), Notifications (0)

hardtree^{beta}

Search

Home Account Administration **Calendar** Changes Documents Finance Identities Infrastructure Knowledge Messaging Projects Reports Sales & Stock Service Support

All Events
Upcoming

Calendar

Month

Week

Day

Settings
Integration

Calendar [+ New Event](#) [Help](#)

Go to date: [Submit](#) [Go to Today](#)

← August 2010 **September 2010** October 2010 →

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
35	30 Office Viewing: Justin Downes	31 Office viewing: Alexandra Haddow Office Viewing: Tina Biswas	1	2	3	4	5
36	6	7	8	9	10	11	12
37	13	14	15	16 Cloud service management debate	17	18	19
38	20	21	22	23	24	25	26
39	27	28	29	30	1	2	3

Demo

The screenshot displays the Hardtree web application interface. At the top, a browser window shows the URL `http://alex.hardtree.com/#/identities/settings/integration`. The page header includes the Hardtree logo with a "beta" tag, a search bar, and a user profile for "John" with links for "logout", "Settings", "Messages (0)", and "Notifications (0)".

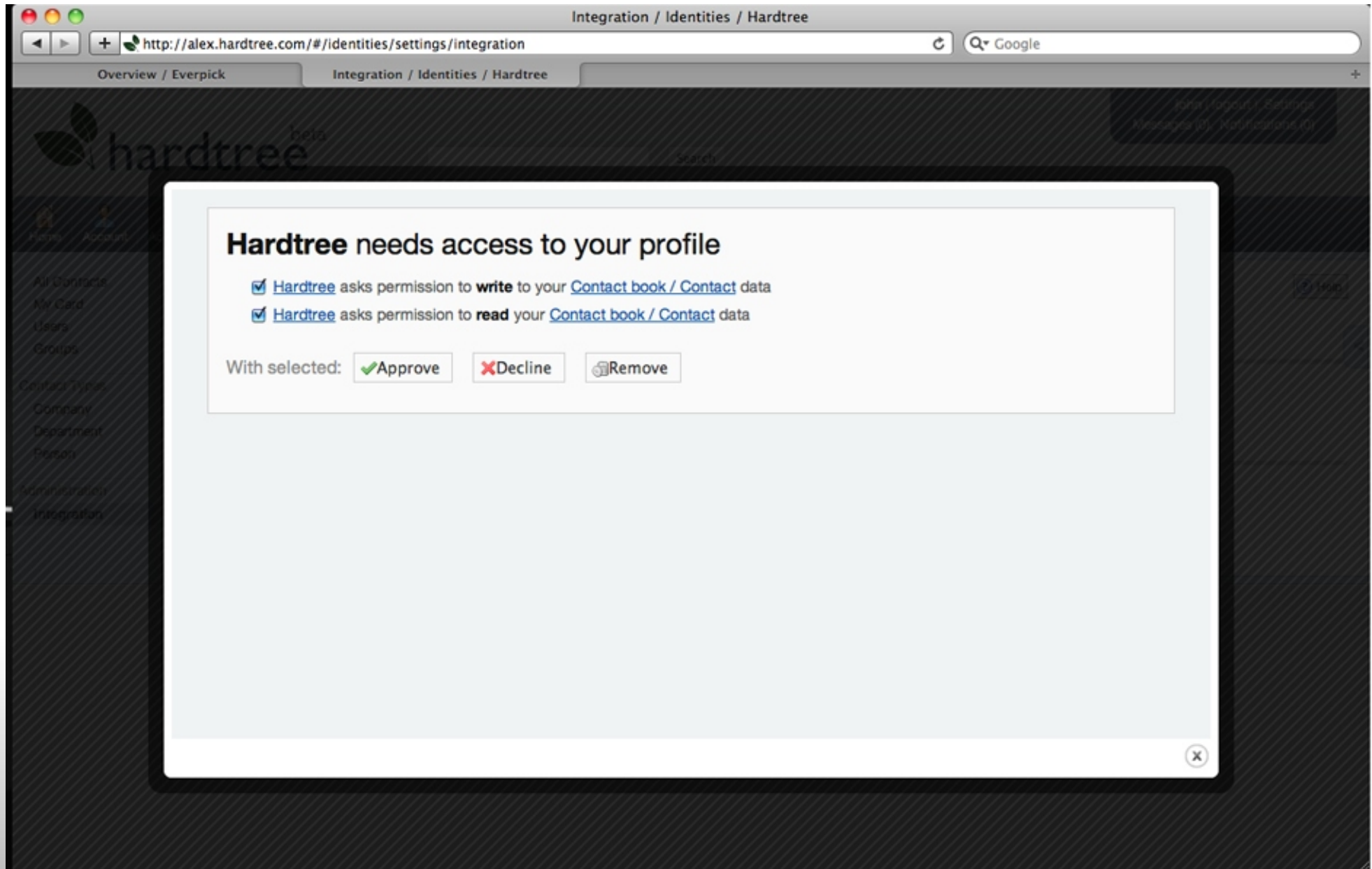
A horizontal navigation bar contains icons for various sections: Home, Account, Administration, Calendar, Changes, Documents, Finance, Identities (selected), Infrastructure, Knowledge, Messaging, Projects, Reports, Sales & Stock, and Service Support.

On the left, a vertical sidebar lists navigation options: All Contacts, My Card, Users, Groups, Contact Types (Company, Department, Person), Administration, and Integration (highlighted).

The main content area is titled "Identities Integration" and includes a "Help" button. It displays two sections: "Active Resources:" which shows "None", and "Available Resources:" which contains a yellow warning box stating "Hardtree does not have access to your profile." with a "+ Grant access" button.

At the bottom of the main content area, it says "Powered by Nuvius. Your Nuvius profile is Anonymous User". The footer contains the copyright notice: "© 2010 Giteso Ltd. All rights reserved. | Hardtree.com | Language".

Demo



Discussion

- Dynamic discovery of data sources
 - Data mining can help us read data
 - How can we do full CRUD on the data?
- Universal data addressing
 - How can we universally address data based on its nature?
- Semantic Web application
 - Can we derive ontologies from the available data?

Thank you

