What does It Look Like, Really? Imagining how Citizens might Effectively, Usefully and Easily Find, Explore, Query and Re-present Open/Linked Data

mc schraefel

[1] IAM Group, Electronics and Computer Science
University of Southampton
Southampton, UK, SO17 1BJ
mc+iswc@ecs.soton.ac.uk

Abstract. Are we in the semantic web/linked data community effectively attempting to make possible a new literacy - one of data rather than document analysis? By opening up data beyond the now familiar hand crafted Web 2 mash up of data about X plus geography, what are we trying to do, really? Is the goal at least in part to enable net citizens rather than only geeks the ability to pick up, explore, blend, interrogate and represent data sources so that we may draw our own statistically informed conclusions about information, and thereby build new knowledge in ways not readily possible before without access to these data seas? If we want citizens rather than just scientists or statisticians or journalists for that matter to be able to pour over data and ask statistically sophisticated questions of comparison and contrast between times, places and people, does that mission re-order our research priorities at all? If the goal is to empower citizens to be able to make use of data, what do we need to make this vision real beyond attending to Tim Berners-Lee's call to "free your data"? The purpose of this talk therefore will be to look at key interaction issues around defining and delivering a useful, usable *data Exploratron* for citizens. In particular, we'll consider who is a "citizen user" and what access to and tools for linked data sense making means in this case. From that perspective, we'll consider research issues around discovery, exploration, interrogation and representation of data for not only a single wild data source but especially for multiple wild heterogeneous data sources. I hope this talk may help frame some stepping stones towards useful and usable interaction with linked data, and look forward to input from the community to refine such a new literacy agenda further.

Keywords: interaction, design, user experience, linked data

1 Introduction

What does interacting with the Semantic Web or Linked Data actually look like? And if we understand that, what are the challenges in making those interactions possible? And for whom do we imagine we design these interactions to support? Whose problems do we solve with any of the tools we create? When are these imagined users of our work actually real people? How do we know?

Some of my colleagues and I have been thinking about these semantic web and user interaction questions since the first eponymous meeting at the WWW 2004

conference launching the First International Semantic Web User Interaction (SWUI) Workshop. The series kicked off with a memorable head to head session between Jim Hendler and Ben Shneiderman. Violent agreement rocked the sessions.

Since then, we have had a workshop on the themes of identifying interaction challenges in a semantic web context annually at venues from CHI, the ACM's annual Conference on Human Factors, to a virtual workshop between MIT and Zurich, and more frequently, here at ISWC.

One of the high notes of this annual series was the 2006 SWUI in Atlanta where Tim Berners-Lee revealed during his talk the revised Semantic Web Layer Cake that included, at last, a user interaction layer on top no less.
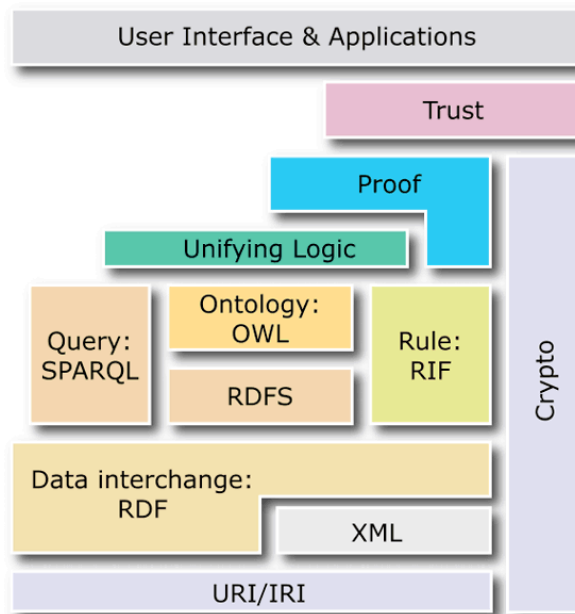
Revised Semantic Web Layer Cake

6 years on from that first outing, I've been asked to give an invited talk about Human Computer Interaction here at ISWC, for which I am honored. We still struggle to get good HCI oriented papers submitted to ISWC. Some of us who do submit them still struggle to get our them accepted at ISWC.

Indeed, I'd like to take this opportunity to acknowledge some of our colleagues who have been an ongoing steering committee for SWUI: Duane Degler, Lloyd Rutledge, Avi Bernstein, David Karger, Jennifer Goldbeck. I'd also like to acknowledge the Web Science Foundation as a constant sponsor for the workshops we've held, more often than not, the founder of the post workshop feast (or at least coffee and nibbles during sessions). In particular, Wendy Hall, Nigel Shadbolt and Tim Berners-Lee.

While I am delighted to have the opportunity to talk for half an hour or more on my favorite concerns about the intersections of interaction and linked data research, I would rather take this space to give some time to other researchers in the field whose

thinking in this space is already well grounded in practice. Indeed, David Karger and I recently enjoined HCI and Human Computer Information Retrieval (HCIR) researchers who deal with large data sets of various domains to help us frame an agenda to excite other HCI/HCIR researchers to consider the opportunities for new research in this area of massive open data.

These conversations have encouraged me to reach out to these experts again, specifically to hear their formative thoughts on what they see as key challenges to make open data/linked data to in particular useful and usable by regular citizens. In particular we asked that they consider what may be new or special about this kind of data that brings new research opportunities to HCI that might also be of interest to the researchers who seek ways to tame this data for functional use for the machine.

This paper spotlights responses from 5 of these researchers across industry and academia. Before we consider five of the expert responses, in the next section, allow me to set the scene of the questions asked. As is apparent reading through the replies, a few themes for consideration recur.

2 Eliciting the main interaction challenges for the linked data/semantic web's interaction success

What are 1 or 2 key priorities you think must be addressed that will aid citizen-focused manipulation of open data sources for personal/social knowledge building?

2.1 Focus: Tools for the Citizen User

The focus of the question is around the Citizen User: a citizen user is not a domain expert (necessarily) - but is someone who has an interest in some information, and the (structured) data is publicly available to help build up an answer to the question, and they are happy to be able to make use of the data for sense making - for building new knowledge. They don't expect "the answer" but want appropriate data to build up a sense of an answer.

So, we are not expecting to create an interaction system that provides *The Answer*, but rather facilitates:
- discovery,
- interrogation,
- manipulation,
- annotation,
- representation of heterogeneous open data sources.

What, therefore are key challenges that in your view we MUST address/prioritize to support citizen based exploration of the freed data of sites like data.gov, data.gov.uk and related sources?

Example Scenarios:
> o *Where Should i Live?*

- where data exists on pollution, hospital waiting times, transportation, political representation in a region, crime stats,
   o for whom should i vote?
- where there may be data on a party's voting record and individual members' voting records, regional crime stats, etc
   o Is this a good school?
- where data may come from league tables, student reports on their views of instructors from all over, house prices, grocery locations, transport
   o What about drug reactions?
what other drugs have people taken with my condition, and what's been the success rate whether self-reported or by other measures?

To make such citizen-based exploration possible, what should be our research agenda? Our concerns from the back end to serve the front end? How do we move from the current high geek expert tools to citizen tools? How do we help people used to thinking about issues with the data to think about issues *for the person using the data* as the best path into solving problems of serving the data. Not all of you may think that that IS the best way, which is fine. Alternatives requested, too.

Some examples of issues we encounter regularly within data:
-   geographical boundaries in different data sets don't match up (hospital trusts don't map to crime regions)
-   not always clear what information in the data is - meaningless labels
-   data is incomplete or messy

2.2 Thesis & Background

My rationale in posing these particular questions is the following: with the emphasis on "freeing data" it seems we are *de facto* potentially establishing or requiring a new literacy - a literacy about data rather than documents; that we've moved from the page if you will to the cell. And that requires new kinds of knowledge - what to do with the data.

In the pre-printing press era literacy was the purview of a select few - the religious cast - who had access to manuscripts. With the press (and the middle class) literacy of documents becomes more wide spread. Is the era of linked data going to be the same now, where data and what to do with it has been the purview of statisticians or those trained in statistics - have access to the data, and produce the results for the rest of us?

If the goal is to believe that access to data is a Public Good, what does that mean for interaction? For a basic data literacy? Does this understanding of data in the 21st Century start with mash ups for all? Where do we as technologists / researchers /designers begin? Similarly, in order to apply that knowledge of data manipulations to the Interface, we also need services to enable normal web-literate citizens to engage the data - find it, explore it, manipulate it, and re-present it where that "it" may be sourced from many heterogeneous sources

2.3 Audience: At The Coal Face of Digging the Semantic Web

Most of the researchers in the Semantic Web community work in terms of dealing with representing the data for the machine efficiently rather than thinking primarily about people accessing and manipulating the data directly.

The goal of these interviews is to help people working on the problems of machines processing data to find it meaningful to connect potentially instead with what citizens need to be able to do with the data, where those citizens are not any more geeks than are the current users of the web.

3 Expert Responses

The following responses are direct reports in their own words of responses to the above framing.

3.1 Daniel Tunkelang, Technical Lead, Google

Here's some of my admittedly US-centric thinking about patent and census data:

Patents. Despite the availability of patent data through public (e.g., USPTO, WIPO) and private (e.g. Google) repositories and the regular appearance of patents in the news, the average citizen (at least in the US) seems to have little ability to either understand or influence how patents work. Some things that we could do to make this data more accessible:

Exposing the links between related resources (e.g., patent applications and prosecution histories). Even people familiar with patents may not be aware of resources like PAIR, the USPTO portal that offers the full history of a patent or pending patent application. And the interfaces make it inefficient and painful to navigate among resources.

Relationships between patents and entities. People invent patents; patents are assigned to companies; people work for companies; companies acquire other companies or their assets.

Connections between similar patents. Even the simple classification system used by the US patent system is not well exposed in interfaces. But I'm thinking of far more than that: connecting patents using link analysis of the citation and entity graphs and computing content-based similarity using information extraction.

Patent law itself is pretty complex, and there's more required here than exposing the raw data in a nicer interface. For example, technical terms should be linked to glossary entries where possible. Links to non-patent-art should also be connected to published documents where possible. And ultimately the value of all of these efforts would require policy changes that would make it easier for citizens to participate. But

there's a chicken-and-egg problem: today's citizens are ill-equipped to participate, so there is little motivation for policy change.

Census Data. It should be straightforward for the average citizen to access public demographic information, whether at a national level, a neighborhood level, or anything in between. But I'm not aware of any interface that makes it easy to do so. The best tools are designed for professionals who invest time in developing reusable queries for generating reports. But it's not just that the tools are complicated.

There's a vocabulary problem--Census data is classified using codes that are familiar for government agencies but not necessarily for citizen users.

Another challenge is that data is collected a varying geographical and temporal granularities, so users need to be able to explore to discover the data that best matches their information needs (i.e., they might not find it at precisely the granularity they had in mind).

3.2 David Huynh, Research Scientist, Google

I don't think I can tell you the *main* interaction challenges, because every interaction challenge seems roughly equally important. But anyway, here are a few thoughts.

(1) URIs are for machines to unambiguously identify entities to operate on, but URIs are practically useless for humans to perform the same task. For humans, images and identifying details (race, gender, birth year, profession for a person; industry, size, location for a company; etc.) are what help us unambiguously identify entities. Sure, there's clearly a realization that raw URIs shouldn't be shown to users, but there is not yet a realization that we need something else to do for humans what URIs do for machines--unambiguously identifying entities. The Freebase Suggest (search) widget provides an example of how to do this:
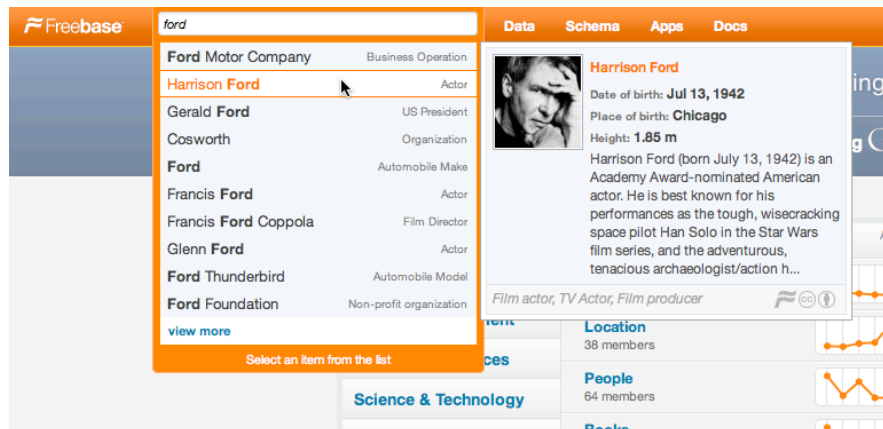
Freebase Interaction image

Note that each suggested entity is labeled ("Actor", "US President", etc.) and described with brief but identifying details on hover. Semantic UIs should strive to communicate to their users that things on the screen are representations of entities rather than mere text, and help users unambiguously identify entities.

The iPhone interface has nuances that make it so pleasant and fun to use, nuances that somehow let the user "feel" the interface. In the same way, semantic UIs must somehow get users to "feel" the semantic entities. After using my first iPhone for a few months, one day while reading a paperback book, upon reaching the end of a page, I instinctively placed my thumb at the bottom of the page and pushed it upward, only to realize that it's not an iPhone. That's what we want here with semantic UIs. After using semantic UIs for a while, when a user sees a plain piece of text like "ford", the reflex should be, "it's ambiguous / raw / bare".

And yes, it's about details, details, details.

(2) Direct manipulation techniques for en-masse data editing seem to be a game changer for a class of users--folks who can handle Excel but are not familiar with scripting or don't have time and patience for scripting. My work on Gridworks has already started to address some of the design challenges.

Just because data is open doesn't mean it's clean or it's formatted in a way that you can use. This inconvenience seems to be swept under the rug sometimes. The semantic web community has focused so much on semantics that perhaps not enough effort has been spared for addressing syntax. But obviously, without syntax, there is no semantic.

It's also quite important to make sure that these tools are generic, rather than RDF-specific. Our goal here isn't to shoehorn everything into RDF. The goal is to increase awareness, desire, and demand for structured data, potentially linked. Let each user decide which format might serve their own purpose at this time. As people use these tools more and more, gradually, structured data, linked, will become natural to them.

(3) Entity Reconciliation. One surprising thing I've learned from Gridworks is how ready people are to want entity reconciliation. And not just with Freebase but with their own databases. This persuaded me to generalize the reconciliation support in Gridworks, as per the Reconciliation Service API (http://code.google.com/p/freebase-gridworks/wiki/ReconciliationServiceApi)

I could almost claim that, by making reconciliation easy, Gridworks makes it obvious why one would want reconciliation. It's almost as if the tool makes people think in a certain way.

So, don't start with "RDF". You would have already lost. Start with what users ask for. Then if possible, let the tools nudge their thinking toward RDF or whatever that's ideal in the long term.

(4) Help build the upcoming structured data web. Or do research on the semantic web. Pick one (since you can't do both).

3.3 Ed Chi, Principal Scientist, Augmented Social Cognition, Xerox PARC

The issues you raised was precisely the inspiration for my Ph.D. Thesis work on creating a visualization spreadsheet. The idea was that if people can easily use spreadsheets, then they ought to be able to take that model further and start creating visualizations using them, and the thesis was an exploration to find out how to design such systems. I think of ManyEyes, and Jeff Heer's later works to be in the same direction.

We have since learned a lot about user-contributed content on systems like Wikipedia, Delicious, Twitter, and they show a very interesting participation architecture that consists of readers, contributors, and leaders. Not all users want to be leaders, and not all users want to contribute. We have sometimes use the derogatory term of "lurkers" to describe "readers", which I think is a bit unfair. Ronald Burt's work have shown that a lot of us would like to be brokers of information among social groups, but there are also need for an audience, or followers, who might become brokers later, but not everyone all at once.

I believe that data manipulation of open data sources to follow the same curve. Yes, some cancer patients will want to read all they can about their condition, and do the analytical work, and others (not necessarily because of tool limitations) would prefer to take a backseat, and let others curate the information for them. What's interesting is that they might want very simple interactions that enable for basic sorting of data, or maybe even services that interpret the data for them (e.g. doctor experts), but they would prefer someone else does the bulk of the work (even if it becomes very easy due to tool development).

Consider a typical usage scenario: I am reading several medical journal articles. Data all in tables, in PDF format. need to extract the data from the tables and plot them. Ahem! Good luck. Let's go and type them all in by hand.

So, given that, what can we do?

First, it's quite clear that much of the hard work remains in data import and cleaning. To democratize data analytics and manipulation, the bulk of the difficulty is dealing with data acquisition. Unfortunately, most of this is engineering and not sexy research, so there aren't really innovative work in this area.

By not exciting I mean, I don't know of a single tool that enables me to grab tables out of PDFs. Worse still, if I have browse around on the net, and I find the data I need in web pages, often they're in HTML tables that are very hard to cut and paste into my excel spreadsheet. What tool is really out there for my information extraction tasks? Tables are just one example. Other problems include things that are locked in databases, but barely visible to end-users:

- say I want to analyze all of the flights from US to Europe over the last month. How do I get the data? Do I perform lots of searches on travel websites to extract that? Do I go to airlines one by one and examine their schedule (in PDF or HTML format), and get the data that way?
- say I want to plot the price of harddrives by dollar per MB in the last decade. Again, where do I get the data? How do I clean it, so that I can plot them?

Some information extraction (AI-style algorithms, and some machine learning techniques) are making some inroad in this area. Some recent work on entity extraction with human in the loop seems pretty good. So if I have a document, and

want to find all interesting entities in them, and make a cross-index of related entities, there now seems to be some good research tools that do that. I also believe that mixed-initiative research for data import is sorely needed. We're doing a bit of this work in my lab at the moment. That is: human in the loop. The machine does some extraction, then human says, ah, that's not quite right, fix it this way, and machine do more, and then human fix again.

Second, there is the issue of data literacy. What kind of visualization works with what kind of data? What analytic technique is appropriate? Early work by Jock Mackinlay pointed to the possibility of automating some of these design choices, and we haven't made a huge amount of progress in this area.

Some tree Viz seems pretty automate-able, as are stacked graphs, population analyses, tables. I tend to favor simple visualizations that are understandable to lay people these days. Visual literacy is a huge problem that will take decades to overcome, so I favor simple vis these days.

Wizards, try-visualization-refine loops have all been tried in research. We need to stop inventing new visualizations, but actual usable tools for people here. By going to vertical domains, we will learn how to solve this problem. We need curriculum in visuzliation that is part of basic education.

That said, Vis researchers need to work on real scenarios more often. Go into medicine, and you see a lot of data analytics problems that are huge, and often not about visualizing generic trees. Often, it's visualizing protein interaction networks, or seeing evolving relationships. Go into another field, say, transportation, and you realize you need to combine infovis with geo-viz. It's often not about new visualizations, but about how to put vis components together.

3.4 Lloyd Rutledge, Computer Science, Open University, the Netherlands

In summary: less emphasis on grand new interfaces. More on familiar interfaces, but under user control and independent from the data. The user doesn't notice anything, thus no new literacy. The user only stops noticing that information access doesn't work the way it obviously should. Can we thus take large-scale data from multiple civil sources and have users access it in a way so unified and quick that there is nothing remarkable about it (finally!)? A "new literacy"? Computers as devices require(d) a new literacy. The Web didn't: users of the Web feel it acts they way they always knew it should (even though they actually couldn't imagine it beforehand). We do new things with the Web in new ways, but they feel familiar once you start.

To me, it seems that the same will be true of the end-user front-end applications of the type of use of the Semantic Web for which we dream and strive. The users will not notice the difference. They won't really notice anything. What will happen is:
- They want some data and they ask for it in a reasonable commonsense way, probably in ways they already (think to) ask for information
- Appropriate and correct data comes back.
- It comes back in the form of a presentation that makes perfect, common sense. The form of presentation itself is not remarkable apart from the data it presents.

This all happens with interfaces users have long been familiar with. The end user won't notice. At least on the per interaction basis. Perhaps over a longer period the user will day "It seems getting information used to be buggier and clunkier". The end users have no new literacy to learn. Their current literacy just works better.

This was just about data access. Could we argue the same way for data input and sharing? I think so. Users add data using means they already know. They and other users get this data back in ways that make sense. This data gets combined with other data, but that makes sense, of course? Only Semantic Web researchers know how remarkable this last step is. No one else will notice a thing. Researchers strive for a grand new SW interface that will put SW in the mainstream like a magic bullet application. But why make a new interface "paradigm"? A more appropriate challenge is to get tried-and-true interfaces to work the way they should with data placed on and accessed from the Semantic web the way it should.

There is no new literacy. There is new technology, good practice and science on the side of Semantic Web developers to make existing literacy work with data that is the way it should be.

One problem is that familiar interfaces are often controlled by non-user parties who also own and isolate the data. Challenges are thus having user control the form of interface and unrestricted access to public data, and have these two control issues be separate from each other. And to have the users not notice anything: they just have their interface, and they just ask for and browse data. And they just simply get it. No barriers based on who is providing the interface or who is providing the data or if one needs to be linked on the other.

A second challenge: allowing seamless combination of public, institutional and private data, all in the same interface, but with the corresponding security and data sharing/blocking. An example is combining civil databases on medications and medical services with your own medical records.

I [don't] mean to poopoo the work of [previous SWUI presented work]. But I think we need to encourage other challenges now. For one, there are too many submissions for new types of interfaces to the Semantic Web that aren't research and don't work, as we've seen in various journal submissions.

Making a new interface is often an attractive project for programmers, but most fall short of burden of proof, and there are only so many new interfaces possible. Not only unproven, many proposed new interfaces just don't work (such as big fat graphs).

But even the successful new interfaces aren't that new and aren't necessarily attached to the Semantic Web. All the new SW interfaces work with any amount of data of any origin. Their newness is more about what computers make possible for data access than was the SW makes possible.

These successful new SWUI's also don't require new literacy. They are natural extensions of familiar interfaces, some of which go back to paper. Like hypertext, they have a "retrospective obviousness", despite being hard to imagine beforehand (and hard to develop the first time). Thus no new literacy needed. SWUI that do require new literacy, like large RDF graphs (BFD's) and queries, even assisted, and even visually assisted (mostly), tend IMHO not to catch on.

So what is new is the type and scale of data that gets to these interfaces. In the civil service data example, what we need is multiple civil service branches to have their

data in familiar, even in 21st century SWUI's like facet browsers and autocompletion, but seamlessly. When another institute doesn't have their data on the SW/LW, end users should find it strange. Not "Why do they have any RDF files?" but "Why can't I get at this information? Why do I have to use their website to see it? Why do I have to jump back and forth from their website to my (semantic, but they don't know it) browser?"

3.5 Abraham Bernstein, Dynamic and Distributed Information Systems Group Univesrity of Zurich

There is plenty of data out there, but as you point out the linking is abysmal. I am not sure where I read it but even the connections between the LOD datasets is only very brittle ... I am not even talking of data repositories such as data.gov or even worse department of statistic excel sheets from different geographical regions etc...

So, the single most important question is how to integrate a multitude of sources.
ASIDE: Yes, it is true, we are very far away from actually understanding what the "best" way to interact with linked data is (assuming there is such a thing) and approaches such as faceted browsing, David and your stuff, NLP, etc. are a only a first step. Lost of work needed here - mostly of a good UI nature. The crux of the Semantic Web is that it adds heterogeneous (even previously unknown) data sources to the mix. Most of the UI approaches so far assume that the data already has been integrated "nicely" into one data-set. Exhibit, e.g., is great, but the most difficult work has already been done: the data integration. So if we really ask ourselves what the Semantic Web brings new to the picture in contrast to "just" interacting with Graph-based data then it is the data-integration problem.

So if we want to bring the Semantic Web to fruition we need to think how we can help our citizen user to combine heterogeneous data sources. My hunch is that it will need a combination of (possibly novel) UI metaphors, a sprinkle of good AI, some social computation, good software engineering. How can I substantiate this hunch?

- I think the first point is clear: We need to find out what UI metaphor is best used to integrate information. Personally, I have no clue if anybody has systematically explored citizen user data integration. I am aware of many projects doing it for pros, but not a lot of work on casual users.
- A sprinkle of AI is needed, as I believe that mixed-initiative might help to ease the bruden of data integration. To that end some statistical processing (e.g., for finding candidates for joins), maybe some rules (e.g., to encode otherwise collected background knowledge), and guided interaction (e.g., using planing techniques) might be helpful.
- social computation will probably help the enterprise by enabling the exchange of integration recipes.
- Good software engineering is needed to build some robust prototypes to test these ideas.

So finding the right interaction metaphor for integrating data seems to be the single, biggest challenge.

3.6 Others in the Discourse

The above commentaries represent specific contributions requested for this presentation of voices. In related conversations, a few more relevant points emerged that are germane to this discussion. Ben Shneiderman, Computer Science, University of Maryland, maps the process articulated above of discovery, exploration, interrogation, and re-presentation with parallel discussion going on in the visual analytics (VA) world, where a similar process (discovery, exploration, interrogation, presentation) is central. He recommends Thomas & Cook's online book "Illuminating the Path" (nvac.pnl.gov). There is also a 16-step process model in the *Readings in Information Visualization* (1999) that is also useful for construing stages within data engagement to be mapped. Since then, he notes, a variety of process models (e.g. Systematic Yet Flexible) have been described, tied to different data types. Shneiderman continues,

> The current term for this [data processing for sense making] in the VA world is "data wrangling" to describe the rough & tumble effort to get, clean, merge, filter, convert, extract, present, and share. Also part of this process will be discovering what is missing in the data or when the meaning has changed for an attribute or attribute value. In many cases, natural language processing methods are needed to clean messy text data, network analysis helps (as in DDupe for entity resolution http://www.cs.umd.edu/projects/linqs/ddupe/), and increasingly Mechanical Turk workers are being engaged.

Steve Drucker of Microsoft Research poses the question about how to sustain general UI's versus application-specific approaches in this citizen-user context: can general or data agnostic interfaces be adapted to specific conditions? By "condition" we might consider different capabilities or specific devices. "What would be the logical workflow to enable this to be a convenient and compelling usage condition?" An example case may be adapting Huynh's Exhibit to work within Excel.

Natasha Noy of Stanford University's Biomedical Information Research queries the representation challenges of making visible distinctions that may need to be made about data access in terms of provenance and trust: how does exposure of provenance and trust get represented across these borders? For example there is data that may be linked but which most citizens would not get near: health record data, even, intriguingly, if anonymized may remain protected. Thus there may be data known to exist, but not accessible. How might these cases be incorporated into tools that would expose sources for possible querying?

4 If you build it…

Several strands emerge across these responses to the challenge of what is the interaction for the data web to be like?

David Huynh is keen to foreground entities free from URLS, as URI's are codes for machines, not concepts meaningful for people. Likewise under the for people heading, making practices easy for people to perform makes new processes' value almost "obvious." Make alignment operations easy fast and intuitive, Huynh argues; use familiar interaction approaches like direct manipulation, and people want it.

Lloyd Rutledge also talks about naturalizing what seems to be new now – having raw data sources from organizations available – into practices that simply illuminate a gap if they are not there. Like Huynh, Rutledge suggests that the machine readable remain machine readable: one should not wonder "where's the RDF" – but "where's the information I can use for this problem." While Huynh evolves facetted browsing into more spaces, Rutledge plays down the need for new interaction or new interaction paradigms. For Rutledge, simply getting the anticipated right data back from an interaction is a big win. He postulates this experience not as a "new literacy" but as the "current literacy just working better."

Ed Chi turns the focus away from manipulating extant data to supporting capture of new data, whether personal or public. He likewise owns that not all citizens will want to do the raw data manipulation work anymore than all readers of Wikipedia contribute content to it. Hence services that enable data basics like simple sorting may be invaluable for light touch exploration. Sometimes, just getting data out of a fixed source may be the win: simple tools to remove and convert tabular data in PDF's to new metadata encoded data may also be a boon for personal use, and the ability to share/contribute new open data or raw data or linkable data – rich data – to the world quickly. This table cutter may be applied manually, but Chi also makes the case for more AI type scrapers to go and make the data that has not yet been formally freed. Then being able to wrap visualizers around the data semi-automatically at least via wizards may also help make the information accessible now as information, not just data.

Daniel Tunkelang laments the fact that in his experience there may be copious amounts of data already freed, such as patent or census data, but that its availability offers little or no opportunity for the citizen either to explore it or influence the process. Tunkelang wants obvious relationships between data better excavated, exposed and presented, such as patents and prosecution histories. Services already exist for histories of patents; these have yet to be linked to the patents themselves. Perhaps even more relevant to the searcher, similar patents are not obvious. Prior art could also be linked automatically.

Finally Abraham Bernstein echoes Tunkelang's and Chi's sentiments to say there is already a lot of data available that screams out for almost native linking of well expressed interactions (semantic zooming, eg nation to city to neighborhood to street) to AI to blend with mixed initiative to both find associated data (patents to their prior art components or drawings), to human computing/mixed initiatives to help enrich data where there are gaps.

Drucker wishes to see the use case that will show how general UIs may help work in specific contexts and Shneiderman shows where there are existing paradigms in visual analytics that may be useful to frame the practices to be represented (discovery, exploration, interrogation, re-presentation), and Noy suggests we consider representation issues for boundary conditions of the fully open to the partially exposed.

Intriguingly, there are few examples here of particular interaction designs. The closest we get is Huynh's approach to facets. Another opportunity for research in the SW/UI space may be to taxonomize the approaches that may be useful for the types of exchanges rich meta/data affords beyond facet browsing. What, as well, do mixed explorations look like that blend documents and data?

5. Concludium

The above exchanges are background or subtext to the formal conference presentation to be presented at ISWC 2010, and I heartily thank the participants who agreed to share their voices in this context.

From these, readers can see that those of us who are investigating how to support rich data sources exploration are intrigued by a variety of different properties in the space. All of us, however, seem to come from a core starting point: what are desirable and sensible processes for people? If one is looking at patents, what data is associated with patents and why not bring those sources together? Similarly, if one is looking at a census, why not make it possible easily to add new annotations to that data or connect related sources or represent relationships? If there is a hole in a data set, similarly why not find ways to automate citizen-directed scraping to enrich such sparse data sets?

Most of these questions have been acknowledged at least in conversations within the semantic web community, and many of them predate the semantic web, going back to hypertext. So we may wish to ask ourselves: where are the great semantic web applications that are meaningful to citizen users by doing these apparently simple, obvious, things?

This is the 9th Semantic Web conference. If we do not have these kinds of apparently simple and sensible interactions by now, is it time for us to look at our program and ask if there's something we should be doing differently? And if not why not? And if not, how else do we get to a useful and usable semantic web of linked data for citizen users? Which of these challenges will we own to say that by the tenth anniversary conference, we will prove not that we can manage a even more triples in less time, but that we can delight a citizen by solving one of her data-related problems?

May year ten be the year of the Semantic Web Citizen (?).