Why the Web Never Took Off

Les Carr
IAM / EPrints
Abstract

• Many people think that the Web has revolutionised the lives of researchers and academics, but to a large extent it has failed.

• This seminar will describe
  – where it has fizzled out
  – explain some of the work that we are engaged in to fan the flames again
  – seek to re-ignite the explosive potential that the web architecture provides.
I Want to Tell You A Story

• It is a story of us
  – Researchers, students, academics
  – Knowledge workers
• It is very partial and ideosyncratic

If you are reading this on the Web, then you’re probably wondering how a Web researcher can possibly claim that the Web has failed to have an impact on researchers.

This talk is about how researchers are failing to engage with the Web as writers, disseminators and information producers. This is evidenced in the worldwide attempts to provide Open Access via repositories and in the low priority that researchers give to maintaining accurate and up-to-date home pages and project pages on the web.

This talk pins the blame on the researchers’ relationship to the publishing industry, and their dissociation from their own intellectual outputs. It raises the issue of responsibility in the context of research knowledge.
Plot of the Story (Part I)

- The literature
- Not accessible, subscription etc
- Open Access, Budapest
- Institutional Repositories
- Empty - too difficult to fill
- Conclusion - Open Access is wrong, repositories are too hard.
- Or is it?
The Literature: As We Imagine

- Integrated
- Available
The Literature: As It Is

- Inaccessible
- Disjoint
The Twin Peaks Problem

- 24,000 journals with 2,500,000 articles/yr
The Budapest Open Access Initiative

• Old tradition of scholarly publishing + New technology of the Internet =

• Public good: free and unrestricted access to peer-reviewed journal literature

Budapest, December 2001
An Open Access Strategies

- **Green: Self-Archiving**
  - Journals continue as normal
  - Authors deposit a copy of their papers into an ‘institutional repository’
  - Public copy is a supplement to the publishers official article for those who can’t afford a subscription
  - Also an institutional record of its work for sharing, reuse, marketing etc
Fast Forward to Open Access

• The Optimal and Inevitable for Researchers.
  – The entire full-text refereed corpus online
  – On every researcher’s desktop, everywhere
  – 24 hours a day
  – All papers citation-interlinked
  – Fully searchable, navigable, retrievable
  – For free, for all, forever

  *Stevan Harnad, Les Carr*

  *OpCit International DLI Project Proposal (1999)*
Problem

- ECS repository, 11,000 records, 4,000 full text, 80-100% open access to our research output.
- Average repository, 300 items, 200 full text, negligible research output
- Recent NIH request for OA achieved 4% compliance
Cliffhanger

• Open Access is just too difficult

• Repositories are too difficult

• Only OA nutcases care
Plot of the Story (Part II)

- It’s not just OA Repositories that aren’t well used
- Neither are home pages, or project pages
- We don’t make web dissemination a priority
- In a world of ubiquitous web use, we start to disappear
Home Sweet Home Page?

• 30% of ECS academics don’t have working home page
• Neither do 20% of MIT professors
How Knowable Is Our Work?

• How does the outside world (external agents, e.g. journalists, industrialists, students, researchers) find out about us?
  – Google, linking

• Can you find joined up examples
  – BBC story links to researcher’s home page links to their research links to opportunities for MSc and Postgraduate study - bingo!
  – No.
  – Not just us look at BBC tech stories about any uni.
  – Some researchers feel they don’t need to care about the web.
Plot of the Story (Part III)

- Hypertext researchers go back to Vannevar Bush
- Let’s also go back to HG Wells, a decade before
- Who is responsible for knowledge and how can it be found?
- Enter the villain!

- Encyclopaedias of the past sufficed for the needs of a cultivated minority
  - universal education was unthought of
  - gigantic increase in recorded knowledge
  - more gigantic growth in the numbers of human beings requiring accurate and easily accessible information
Permanent World Encyclopaedia

• Discontent with the role of universities and libraries in the intellectual life of mankind
• Universities multiply but do not enlarge their scope
  – thought & knowledge organization of the world
• No obstacle to the creation of an efficient index to *all* human knowledge, ideas and achievements
Vannevar Bush, As We May Think
Atlantic Monthly, July 1945

- Director of the Office of Scientific Research and Development in USA, coordinating 6,000 American scientists during WW2
- Make our ‘bewildering store’ of knowledge more accessible
- “For many years inventions have extended man’s physical powers rather than the powers of his mind.”
The Memex (never built) was to be a mechanised device to allow a library user to
- consult all kinds of written material
- organize it in any way the user wanted
- add private comments and link documents together at will.

A personal library station which held all written articles and journals on microfilm.
- system of levers allowed users to add links
- create trails
Wells vs Bush

- **Wells**: a centralised, managed global knowledge repository to combat fragmenting academic authority.
- **Bush**: a *cross-disciplinary* scholarly paradigm to combat fragmenting scientific knowledge.
The Big Bad Wolf

- 1960s Robbins Report / expansion of higher education & expansion of science budget
- After the war Robert Maxwell decided to publish scientific journals and set up Pergamon Press which was quickly and hugely profitable. (BBC News)
- Up to this point, journal publishing was done by university presses and scholarly societies
- The New Demand made for a very profitable system - with an increasing number of commercial publishers moving into STM.
Plot of the Story (Part IV)

- Result is that universities further abdicated on their Wellsian responsibilities
  - Knowledge dissemination outsourced
  - Ownership of research materials given away
- Scholarly communications now largely in the hands of commercial concerns
- Popular imagination is that publishers do reviewing
  - Researchers are passive consumers of publishers output
Recapitulation Story V = I

• Who takes responsibility for curating the knowledge of the world?
• Back to OA & repositories - we do!
• The Institutional repository is a place where the members of an institution can curate their intellectual outputs / knowledge capital
  – Share
  – Use
  – Reuse
• The real Web revolution of ubiquitous knowledge will arrive.
Our Research (ECS)

Only a small part (7%)

Hurrah! The goodies are winning!
Personal academic web pages *ad hoc* and 2 years out of date (on average!)
Portal pages: draw together all official info from many databases
Semantic web portal pages: equivalent information published as RDF. Policy establishes ontology, individuals and persistent URIs.
Repository pages: metadata and data deposited by user on a regular basis (backed by policy) for Open Access.
Semantic repository pages: equivalent information published as RDF.
Semantic repository pages: adhere to discipline of referring to URIs established by external authorities, ie authors aren’t strings, they are individuals who can be tabulated.
The Semantic Institution

• The Repository is a single player in an institution’s knowledge publishing activities
  – Artefacts
  – Agents
  – Activities

• Different domains of authority over different kinds of information

• Should be an integrated policy for information publishing
  – Who assigns ids to what?
  – Who makes statements about what?
The Semantic Nation

• Semantic Individuals
  – researchers who curate their own intellectual capital using repositories

• combine to make Semantic Institutions
  – organisations that defines policies on naming, identification, information sharing and preservation
  – see recent ECS initiative

• which co-operate to form a Semantic Nation
  – loose co-operation via SPARQL endpoints allow agents to query and combine all the information exported by every institution
Semantic Mashups

Mashups are impressive only because they are difficult

AKT award-winning CS AKTivespace involved ACM classification, EPSRC database, mining entire national CS home pages, RAE database, publication list & making proto Google Map
A Misty Look into the Future: Plot of the Story VI

• Ubiquitous knowledge
• No need for ignorance
• What are the implications for devices, services, us, education.
Personal Experience

• While talking on the phone (e.g. to a potential collaborator) I can appear not too uninformed if I use Google to look up all the projects and people that they mention.
  – I can make links between what I discover and what I know
  – I can do it fast enough to keep up with the conversation
  – BUT the keyboard clicks make it obvious to the other party
A Turing Test

• Can you tell the difference between a real expert and someone faking it with Google?
  – or other more advanced knowledge tools
• What are the limitations on chicanery?
  – How unintrusive can the technology be?
  – How should the queries be made?
  – How should the results be shown?
  – What sort of answers can be faked?
    • cf Bloom
Experimental Proposal

• Who Wants to Cheat at Millionaire?
  – How well can you perform if one of your lifelines is 30-seconds on Google?
    • A minute, two minutes
    • All your lifelines are Google
    • Every answer can use Google

• This is Bloom’s lowest layer - RECALL.
Experimental Proposal

• Open Book Exams?
  – How well can you perform on “understanding” questions in a typical Part I undergraduate exam (with Google)?
  – How long do you need?
  – How much do you retain?

• Work your way up through Bloom
Why is this Important

• In a world with ubiquitous calculators, no-one needs to know how to use log tables or Napier.
• In a world with ubiquitous knowledge sources immediately to hand, no-one needs to know anything?
• What are the limits on knowledge support - what do people need to learn.
What’s the Difference?

• Clever Chap + Learning = Expert

• Idiot + Google = Expert?
The End

A “Total Imagination” Production

by

Leslie Carr