Jones, Barbara; Goff, Mhoraf

LEARNING TO LIVE WITH DATA DELUGE AND WHAT THAT MEANS FOR EDUCATORS

Universidad de Salamanca
Salamanca, España

Available in: http://www.redalyc.org/src/inicio/ArtPdfRed.jsp?iCve=201021400002
APRENDIENDO A VIVIR CON EL ALUVIÓN INFORMATIVO Y SUS IMPLICACIONES PARA LA EDUCACIÓN

Resumen: Una característica importante de nuestra época es el aumento exponencial de los datos. Grandes cantidades de información, en formato digital, son consumidas para la investigación, y a la vez derivadas de la misma. Internet ha hecho la "información" accesible -no solo datos- a aquellas personas que tienen los recursos necesarios para acceder y utilizar Internet. Como resultado, se ha generado para las instituciones educativas un nuevo conjunto de requisitos para dotar a sus profesores y alumnos con habilidades para acceder y aprovechar efectivamente la información en este entorno digital, bien para propósitos de aprendizaje, o para alcanzar éxito en esferas personales o laborales. Este trabajo explora este nuevo paradigma, que proporciona un estudio del horizonte actual, con ejemplos de trabajos recientes donde se tratan estas áreas de interés, finalizando con la elaboración de conclusiones sobre el estado actual y las posibles direcciones futuras.

Palabras clave: Internet; ciencia de la información; habilidades de búsqueda de información; nuevos aprendizajes; gestión de datos; pedagogía
LEARNING TO LIVE WITH DATA DELUGE AND WHAT THAT MEANS FOR EDUCATORS

Abstract: A major characteristic of our age is the exponential rise of data. Vast quantities of information in digital form are both consumed by and generated from research. The Internet has made “information” accessible to those with the resources to access and use the Internet. As a consequence a new set of requirements has been generated for educational institutions to equip both their teachers and learners with the information skills to access and effectively exploit this digital environment for learning purposes and for success in employment and personal spheres. This short paper explores this new paradigm providing a review of the current landscape including examples of recent work which addresses these areas, and drawing some conclusions about current status and possible future directions.

Keywords: Internet; information science; information seeking skills; new learning; data management; pedagogy;
LEARNING TO LIVE WITH DATA DELUGE AND WHAT THAT MEANS FOR EDUCATORS

Fecha de recepción: 18/11/2010; fecha de aceptación: 10/02/2011; fecha de publicación: 31/03/2011

Barbara Jones
Barbara.Jones@mbs.ac.uk
University of Manchester (UK)

Mhoraf Goff
mhorag.goff@manchester.ac.uk
University of Manchester (UK)

Introduction

We start with the widely held proposition that, “A fundamental characteristic of our age is the rising tide of data – global, diverse, valuable and complex” (European Union, 2010) and that this poses an unprecedented challenge in terms of how knowledge is organised, and by implication how knowledge in our society can be advanced (Floridi, 1995). The ‘information age’, presents us with a new way of conceiving contemporary societies with information as a defining feature: this has emerged as quantitative changes in information are bringing about qualitative changes in social systems (Webster, 1995). This process will continue as quantities of information continue to increase exponentially (Office of Science and Innovation, 2007). The challenge for education and educators is that digital information is no longer static and sequential in form, as a text would be, but more often of mixed content and dynamic, needing a different set of skills to handle and exploit it (Jones & Miller, 2007). The Internet has revolutionised how knowledge is packaged, shared and disseminated (Rosenberg, 2001). The internet has been the key driver of this paradigm shift because it has made possible the transition from analogue to digital information (whether digitized or born digital), and this has affected how knowledge is produced and disseminated through the process of research. The democratising (in the sense of opening up) and globalizing effect of the internet means that this knowledge is more widely accessible than ever before. The internet as a network of information resources has shifted from read-only to read-write capabilities via Web 2.0 enabling applications such as Facebook, Twitter and Wikipedia. These allow users to create and disseminate their own content and to customise and repackage content other people have created, such as mash-ups.
Information in a variety of forms and formats whether text, video, audio, numerical data or software code produced by others, re-worked and recombined is being innovatively exploited to create new products and services. This is a significant development because it challenges the view that new ideas and innovation flow from privileged centres of creativity which reinforce existing hegemonies of power relations (McMasters and Wastell, 2005). Such developments are changing the nature of education as a mechanism for transferring knowledge and facilitating learning, both in terms of content and the pedagogical process. In the educational and research context this paradigm shift now entails the need for new curricula and new methodologies to exploit the deluge of accessible information and technologies, which can to help achieve learning, and also most importantly to equip students with the competencies necessary to exploit such resources. In research, it is the domains of science and education, which are in the vanguard of knowledge creation and in the consumption and use of technology where this change can be seen most starkly. The „data deluge” as described by Hey and Trefethen (2003) in the context of e-science is perceived as affecting the whole practice of research across all disciplines. Education institutions as places where information and knowledge are created, transferred and disseminated are thus strategic players.

The possibilities for researchers and teachers

In the field of research, advances in instrumentation and the incorporation of computing based technologies into research practice in the sciences has brought not only greater throughput of research data but changes in the nature of research. Jim Gray of Microsoft talks about the “Fourth Paradigm” of data intensive research (Hey & Tansley, 2009) which is creating new sub-disciplines of science; such as „informatics” which collects and analyses scientific data on a large scale from multiple experiments, and also „computational” sciences which seek to use computer programmes to model and simulate natural processes. These sub-disciplines are a level of abstraction one step on from the empirical tradition of collecting observational data about natural phenomena; instead the collection of data becomes the aggregation of the outputs of multiple experiments. As costs for digital storage decrease and volumes of digital research data increase, there is greater potential for extended reuse of data not only by its creators but potentially by the wider scientific community, outside organisations and the public. Within research this opens up opportunities for meta-studies and complex longitudinal studies with potential to revolutionize public health and social sciences research. An example is translational science, a new scientific paradigm, often labelled „bench to bedside”, which emphasises the interface between basic sciences and clinical sciences.
and which has been made a research priority by the National Institutes of Health\(^1\) in the United States (Woolf, 2008). Computing based technologies make possible new methods of data collection, particularly in social sciences where research can be carried out using the Internet and social media as tools for recruiting, engaging with and communicating with participants. The analytical power provided by computing hardware and software allows for analysis of large quantities of data, investigation of patterns and trends, data mining, and collaborative work between otherwise disconnected researchers.

For researchers and those who train researchers, this generates challenges for the management of information which did not apply with pre-digital forms because their concrete physical presence (in the form of paper documents, books, tapes and so on) forces an inevitable process of natural selection for the finding, appraisal and disposal of resources, “It is increasingly easy and cheap to copy and store information, which means that we have largely lost the process of „natural selection” that existed when information was paper-based” (Stevenson, 2005).

A consequence of the critical mass of digital information is the open access movement, which has brought many academic publications within reach of students and academics via e-Journals particularly databases such as PubMed\(^2\) and this has included digitisation programmes for pre-digital materials. For research data itself specialist subject repositories, such as the European Bioinformatics Institute databases\(^3\) have been developed which aim to provide access to certain generic subsets of research data such as gene sequences and protein formulae. These in turn are driving the trend towards open access data and standardisation of data formats, which will increase interoperability, and reusability of data across geographic and disciplinary boundaries to create community resources. The UK MaDAM\(^4\) project is building a pilot e-infrastructure for research data management for biomedical scientists by identifying user requirements for an e-infrastructure which will support the day to day management

---

\(^1\) [http://www.nih.gov/](http://www.nih.gov/)


\(^3\) [http://www.ebi.ac.uk/](http://www.ebi.ac.uk/)

\(^4\) [http://www.library.manchester.ac.uk/aboutus/projects/madam/](http://www.library.manchester.ac.uk/aboutus/projects/madam/)
of research data as well as its long term preservation (Poschen et al. 2010). Data storage is becoming cheaper but storing and managing vast quantities of data for the long term is both financially costly in terms of hardware and also potentially labour intensive. Such labour can be minimized with up front user-driven data management planning and access to suitable tools which will enable researchers to store and preserve data so it is identifiable, searchable and can be annotated, tagged and linked flexibly so it can be accessed and reused from the short to the long term. A key objective now, of many research funding bodies is that “maximum and optimum use be made of expensively generated and expensively maintained data” (Sinnott et al. 2005) particularly in the context of research funded by public money and the need for the contribution of that research to the public good to be made clear. There is therefore increasing pressure on universities and researchers to protect preserve and maximize the usefulness of research outputs and there is an effort to ensure that the products of such research are publically available, as recommended by the OECD (2007). Because of the labour-intensiveness of the processes to make data reusable and to curate for long term accessibility it is important for institutional and subject repositories and data centres to be selective and flexible in what they preserve. Much research data falls into a grey area where it may or may not be of future value, for example, a considerable proportion of bio-science data becomes useless with advances in the instrumentation used to generate the data which makes the older data, by comparison lower quality and less manipulable. Only researchers themselves are truly qualified to judge the value and potential future value of their data and it is apparent that a highly tacit process of evaluation is going on when researchers make a decision whether to invest time to „develop” data in order to increase its value for future use. Initial findings, of the MaDam project, indicate that researchers were keen to retain all data for potential reuse and reanalysis, even data without immediate obvious re-use value. For example, data from failed experiments may be kept in order to benefit from a „lessons learned” process and to potentially uncover patterns in the future that might lead to new research questions and projects (Goff et al.2010). This has significant implications for storage, preservation and management of knowledge generated from science research in terms of infrastructure and most importantly the effective information skills needed to be taught.

In the teaching sphere, open access educational resources are making teaching materials freely available and access to all these information resources is being facilitated by federated access management across research and education institutions globally. Higher education, and indeed education in general is being transformed by the Internet as geography, and where, to some extent age, background and socio-economic status
cease to be barriers to access to education, distance learning, self-learning and alternative modes of teaching delivery. However, the major problem to be confronted is when teachers, researchers and students are faced with an unprecedented volume and choice of information resources. All need to be able to effectively find, select and assess the appropriate resources. This information needs to be well structured, labelled and annotated for discoverability and if it is to be accessible in the long term consideration needs to be given to appropriate formats and media allowing that the pace of technological change will quickly render some obsolete (for example, data held on 5 ¼ “floppy disks or written in WordPress would now be difficult to access). The Oxford Internet Survey 2009 (Dutton et al. 2009) bears out the notion of information overload for internet users as a recognised problem, with 31% of users and 43% of ex-users stating that they waste too much time going through irrelevant information on the Internet.

In the knowledge intensive educational context the combined effects of new computing based tools and abundant information have created a situation where it is increasingly difficult to find appropriate, good quality information from what could be considered as background „noise”’. The implications are a demand for skills within education which we will refer to in this paper as information science skills. Information science skills are needed both to help manage information and to educate others in information science skills in order to meet the challenges outlined above.

The Internet acts as both the gateway to information resources and as the mechanism by which those resources are increasing in volume and diversity. It acts as a powerful leveller of organisations (whether public or private) by removing any competitive advantage through having special access to a particular technological platform (Ciborra 2002). Similarly, in the educational context institutions and individuals are less and less the privileged holders of information and this are redefining the teacher-student relationship. Students are no longer only passive recipients of content, they are acknowledged as potential sources of knowledge and intermediaries in learning with their peers, teachers and institutions. As the content of learning has decreased in importance because it is widely available, the processes for successfully discovering knowledge is now of primary concern in the information society (Hernández-Serrano 2009).
Developing information science skills

A definition for information science skills is offered by Eisenberg & Berkowitz (1990), who specify a „big six“ of skills:

- task definition;
- information seeking strategies;
- location and access;
- use of information;
- synthesis and finally
- evaluation.

In the context of information seeking via the Internet foremost within these skills are information seeking and evaluation components as described in Wilson and Krikelas” (1983) Information Seeking Model and Kuhlthau’s (1992) Search Process Model which enables individuals to use effective search strategies to find the information they need and to assess information for quality and provenance. These are seen as important transferable skills both for education and for future life (Brophy, 1993). The EU, for example, affirms the wide importance of such skills as key competencies for social inclusion, personal fulfilment, employability and active citizenship (European Commission, 2010). Such skills are also necessary for enhancing the mobility of researchers and for job mobility for individuals within a wider context, “e-skills become a requirement not only to enter the workplace, but also to benefit from upward and horizontal mobility once employed” (Lanvin & Passman, 2008).

The important dual implications of learner-centred, collaborative approaches to education (CEDEFOP, 2009) and the wide availability of information resources via the internet are elucidating new sets of skills to enable students to negotiate their way through the sea of information to find what is pertinent to their learning objectives. These have been framed variously as „media literacy“ (European Commission, 2009), e-skills (WEF, 2008) and information literacy (Bertot, 2003). Here we are using the term „information science“ skills as being a subset of more general ICT skills. It is important to distinguish information skills from wider ICT skills, because it is can be easy to underestimate the skill involved in online information seeking. This is because of the apparent but deceptive ease of use, low barriers to use and because of a general assumption that because students use ICT in their personal lives as well as for education that they already possess these skills (Brophy, 1993). It is, however, clear that the requirement for information science skills must also encompass the understanding of how to use information resources for learning (Hernández-Serrano, 2009).
Information science skills have previously been the territory of libraries and librarians. The skill sets of academic library communities now requires greater IT related knowledge and overlap with IT services. In light of the decreasing importance of physical library resources, libraries and institutional repositories are redefining their role in providing information services to students, teachers and researchers alike (Akeroyd, 2001). This suggests them as ideal candidates as the „information science” educators that institutions need in developing curriculum. Libraries can "reach a position where the acquisition of information skills is acknowledged as one of the key learning objectives for every student entering a university, so that no student leaves without being fully equipped to cope with the information intensive world - the information society - as an end-user" (Brophy, 1993: 55). In the UK and Ireland the Society of College, National and University Libraries” (SCONUL) recognise a requirement for basic ICT skills as a pre-requisite for information science skills but are clear about the need for a distinction in order to avoid the fallacy of equating computer literacy with information literacy (SCONUL, 2003). The Harnessing Technology Review (BECTA, 2009) agreed that general ICT skills are a prerequisite for using technology with learners in the classroom. They also acknowledge that teaching these skills will require new curricula, “It is proposed that the development of the idea of „information literacy” requires a collaborative and integrated approach to curriculum design and delivery based on close co-operation between academic, library and staff development colleagues.” (SCONUL, 2003).

Information science skills for students need then to be taught as part of a cross discipline curriculum and this entails a concomitant requirement for information science educators and teachers with such skills to be produced (Prensky 2001).

The majority of students now in higher education have been defined as, „digital natives”, having grown up with Internet technologies. This presents a challenge for many educators who are largely from a non-internet generation („digital immigrants”) in having to teach skills which they themselves may find novel (Prensky 2001, Selwyn 2009). Training students in information science skills will therefore entail a change in the pedagogical approach and needs to be addressed within teacher training as a general policy response (Pedró, 2006).

There are particular challenges in teaching information science skills which centre on the experiential and intuitive nature of such skills. Intuition is framed as „learned experience” (Eisenhardt, 1989) and rapid processing of experienced information
(Crossan, 1998) which cannot be articulated, and can therefore also be categorised as having a large tacit component (Polyani, 1966), perhaps even „deep” tacit knowledge (Leonard-Barton & Sensiper, 1998) where holders do not know that they have certain knowledge. Teaching such skills therefore may be more suited to a „learning by doing” approach. In this regard, the Internet can be seen as a tool, or „probe”, of which skilled use relies on familiarity with the tool and expectations of the medium (Polyani – Personal Knowledge). The importance of social interaction as a source of learning in this regard (Hutchins 1991, Jones & Miller, 2007) is significant in the context of learning information skills in distributed environments. It has been raised as to whether formal education frameworks offer appropriate mechanisms for teaching ICT skills (UN, 2010) and that informal practice, self-learning and on the job training are the key modes for acquiring these skills: for individuals the main channels of acquiring skills are education, training (on and off the job), self-learning and practice at home (Demunter; Eurostat, 2006).

If information science skills have a large tacit component then in an educational context where use of virtual learning environments (VLEs) are reducing face time between teachers and students, teaching such skills through a „learning-by-doing” requires pedagogical innovation in teaching process. A „situated learning” model may be a more useful way to conceptualise the kind of process that is needed where students learn, for example, through the act of participating in communities of practice (Lave & Wenger, 1990). As a progression from this idea of learner autonomy, heutagogy as the concept of self-determined learning (Hase & Kenyon, 2000) which emphasises more informal modes of learning, embraces the use of social media and the internet as resources for self-directed learning and may be an appropriate model for learning information skills. The Centre for Education Research and Innovation (CERI OECD, 2008) has initiated a study across 15 countries in conjunction with the Swedish Knowledge Foundation, to investigate how ICT and ICT skills are being incorporated into initial teacher training. This is part of the OECD’s New Millennium Learners Programme. The aim of the project is to gain understanding of an increasing gap between the use of ICTs in schools and daily experiences of students with ICT outside of school. The project is based in interviews and surveys of teachers in 2 teacher training institutes in each participating country and is a 6 year collaboration, started in 2006, which aims to produce policy recommendations for teacher training institutions and governments in this domain. The IRIS project at the University of Cambridge aimed at capturing the level of information skills amongst undergraduates (Edwards-Waller, 2009) using an online survey, questionnaire and focus groups. The key findings were that many students had low
levels of awareness of electronic resources, combined with a high use of Google. Whilst some resources were registered as being frequently used, these were not necessarily regarded as being the most effective or comprehensive for that subject, although students from some subjects showed a much higher use of course-specific resources. This is also borne out by the CIBER programme (CIBER, 2007) which found that unsophisticated understanding of what the Internet is and how it works as a network of resources entails students” primary association of well known search engines such as Google and Yahoo as synonymous with the Internet. Furthermore the CIBER study found that this is also associated with a preference for Google or Yahoo for study needs over library-sponsored resources because students find them more intuitive and familiar, even if they are simplistic. This has implications for example, for libraries and raises the question of whether they need to integrate with or imitate commercial search engines (Brophy & Bawden, 2005). Other findings from the IRIS project are that undergraduate students expressed a preference for finding relevant information over finding information quickly or knowing who wrote or researched it and that because this reflected perceived time pressures associated with information seeking there was a clear difference between undergraduates and postgraduate students. In passing it might be noted that for user-generated content there is no standard framework for assessing the quality of content unless that material is formally reviewed and published, a process intended to ensure reliability and quality. This entails a challenge for educators, researchers and individuals who need to be able to assess content quality and reliability for themselves (Punie, 2008).

**Implications and conclusions**

The information age entails a need for developing information science skills, which are embedded in curriculum within the educational context. The production of students for whom information literacy skills are “second nature”, will become increasingly important as organisations respond to the challenges of producing and using knowledge on a vast scale. Policy responses which formalise the status of such skills and move them beyond the boundaries of libraries and IT support services are needed in order to disseminate such skills within education and into wider society. In the area of developing information science skills, the approach of policy makers is still very much focused on general ICT access and formal skills training. The European Commission’s „Digital Literacy Review“ for Europe (European Commission, 2008) focussed on recognising that previous programmes had not succeeded in fully addressing the information skills gaps and that there were reasons other than simply limited physical
access to the technology and networks. Although the Review made recommendations for less formal teaching and learning approaches for ICT skills in general, less attention was given specifically to information science skills. These are, although, partially addressed by the European Commission’s Study on the Current Trends and Approaches to Media Literacy in Europe.

Whilst information science skills and training exist in pockets, most often within libraries, uptake of such courses is not universal and the training itself needs to go beyond transmitting information such as use of recommended databases and how to cite resources. If students are to develop critical thinking habits then they need to not only search effectively but also interpret search results (Brabazon, 2007). Search engines are the primary tool used by individuals to begin an information search. Connaway & Dickey (2010) found that 84% of all users in their study responded that they began an information search with a search engine while only 1% indicated they began on an education/ library web site. Relying too heavily on well known search engines such as Google paradoxically risks losing the advantages that such unrestricted access to information, much of it free, has to offer. Focusing only on the information outputs from the internet also risks treating this as direct unmediated access to information without attending to the intermediaries who select, collect and present such information, and whose outputs are therefore very much social constructs (Seely Brown & Duguid 2000). Appropriate training has to enable students to learn how to learn using the knowledge resources available (Hernandez-Serrano 2009). Students, teachers and researchers producing their own content online also need to learn how to manage their personal data, their externally facing profile and their privacy in general when creating their own personal digital spaces (Punie, 2008). Satisfying ICT skills needs via education policy alone is insufficient because introduction of ICT equipment is not enough alone to incorporate ICTs into the learning process if educators themselves do not have the appropriate skills.

Science as a driver of knowledge creation within societies is intense now that technological advances have resulted in data production on the petabyte scale (Hey & Tansley, 2009). Educational institutions are at the forefront of such knowledge creation and of transmitting and disseminating such knowledge. This implies a requirement for a

--

5 http://ec.europa.eu/culture/media/literacy/docs/studies/study.pdf.
new generation of information scientists (European Union, 2010) and for information scientists who can feed into the education agenda to strategically adapt the educational processes and curricula to meet the needs of managing knowledge in such an environment.

Floridi (2007) has proposed a vision of the near future in the idea of “inforgs” - human agents who are informational organisms where the distinction between online and offline has disappeared. Such a prospect makes it important to equip both the digital natives and digital immigrants with the wherewithal to operate in digitally dominated environments. Educational institutions need to develop strategies which cover the spectrum of activities from managing the outputs of knowledge production to enabling effective knowledge consumption. This needs to include information skills training throughout compulsory and higher education and the pre-requisite teacher training.

BIBLIOGRÁFÍA.


Para citar el presente artículo puede utilizar la siguiente referencia:
