

1 of 2 DOCUMENTS

The Independent (London)

May 11, 2005, Wednesday

## TECHNOLOGY: THE **WORLD ON A CHIP;** ARCHIVISTS ARE IN A RACE AGAINST TIME TO PREVENT VITAL DATA BEING

**BYLINE:** DANNY BRADBURY

**SECTION:** First Edition; FEATURES; Pg. 38

**LENGTH:** 749 words

Joni Mitchell once sang, 'You don't know what you've got till it's gone.' She hit the nail on the head: IUCN, the World Conservation Union, published a report last November showing that the numbers of endangered species worldwide grew from 10,500 in 1996 to 15,500 in 2004 " and even that number is only scratching the surface.

'We don't know what's on our own planet,' explains Alexander Rose, director of the Long Now Foundation, a San Francisco-based charity that promotes long-term thinking. 'At present, we probably only know between 1 and 10 per cent of the species on Earth.'

Digital technology is vital for groups hoping to preserve what information we do have. It is enabling a level of documentation that would have been impossible 15 years ago, says Richard Edwards, director of the ARKive initiative, which is collecting images of the IUCN's 15,500 listed endangered species to record them for future generations. 'We are storing digitally,' says Edwards. 'It means our costs are vastly reduced.'

But while digital technology makes storage easier and cheaper, it brings its own problems. 'It's much easier to find material,' says David Thomas, who is a director at the UK's National Archives, an organisation that stores government documents. 'But in terms of ensuring the survival of material, it's much harder.'

This is partly because the amount of data that can be stored digitally is growing. At the National Archives, for example, hundreds of hours of underwater video taken as part of a shipwreck investigation sit alongside text documents, archived government web pages and virtual reality files. It must all be transferred between media every five years or so, because, Thomas says, even optical and magnetic media (CDs, floppy discs, tapes and so on) will degrade " and in the transfer, data can be lost.

But there's a more serious problem, too: how can we guarantee that computer systems and programs in 200 years' time will be able to read our spreadsheet files? One answer is to include metadata " that is, data about the data that you are storing " explaining how to read them. XML, a language designed to describe metadata and data together, is a good candidate for future storage, says Kevin Schurer, director of the UK Data Archive, which stores social science information.

But how can you digitally store an old computer that is necessary to run a program of historic value? Emulating old software and hardware might be one way to decode information tied to old systems, says Rose. He promotes the use of

open formats, which can be more easily modified, as opposed to proprietary ones. Meanwhile, he is building a repository of different data formats.

Foremost among these is the Rosetta project. It is predicted that 90 per cent of languages will die out in the next few decades, and Rosetta aims to document roughly 5,000 of them. The project is bypassing the problem of media degradation by etching images of pages describing the languages onto a large metal disk (above), in microscopic form.

Rosetta is also significant because it allows anyone to submit information about a language. 'Our method goes against a lot of academic methodologies for storing data, but it also provides a richer idea of what language is,' says Rose.

Brewster Kahle works along similar open source lines. He founded the Internet Archive, a non-profit-making digital library, in 1996. He says that the cost of storage has shrunk so much that it is now possible to preserve digitally all the published works of humankind. He invites the public to donate works for archiving on the web.

However, copyright law is a huge obstacle. Last year, Kahle launched a lawsuit (which has gone to appeal) against the US government protesting statutes passed in the Seventies that extend copyright protection to 70 years past the author's death and automatically impose copyright on orphaned works (works for which copyright holders cannot be found).

Kahle says that digital rights management, too, makes it hard for digital libraries to operate. 'Not only is it impossible to copy this material but they've made it illegal to write decoders,' he complains. And if you can't transfer the information on the discs, they will degrade and it will be lost. This is a big worry, for without digital preservation, much will vanish. We may not know how future generations will use what we're storing, says Rose, but we must preserve as many options as we can.

**LOAD-DATE:** May 11, 2005

**LANGUAGE:** ENGLISH

Copyright 2005 Newspaper Publishing PLC