Interview

Power of search: Fundamentals of DAM expanded to semantic analysis of digital workflows – An interview with Chris Glynne of Bold Visions, Ltd.

Chris Glynne

has a vast breadth of knowledge across the sciences in general. Having studied at the University of Sussex and gained an Honours degree in electronic and electrical engineering, he moved into the print and publishing world, delivering core and wide-reaching change. His proudest achievements have come through huge improvements in publishing technology for IPC Media (a UK-based Time Warner division).

ABSTRACT All of the benefits you expect to get out of your asset management solution are rendered null and void if you can't perform appropriate searches and find those assets. Michael Moon interviews Chris Glynne of Bold Visions, Ltd. on the future of semantic analysis for workflowed digital asset management.

Journal of Digital Asset Management (2009) 5, 90–97. doi:10.1057/dam.2008.55

Keywords: workflow; asset management; publishing; metadata; XML; technology infrastructure

MM: We're here with Chris Glynne of Bold Visions.

Chris, let's start off with a little bit of a background in terms of what you've been doing professionally.

CG: I've spent the last 15 years or so in print and publishing in the UK. Having previously gotten a degree in engineering, I moved into the print world. I was fortunate enough to work on behalf of some of the more blue-chip companies in the UK, helping them through the early days of desktop publishing and bringing their design teams in-house.

For the last 10 years or so, I have worked for International Publishing Company (IPC) Media – a large UK publisher that's part of the Time Warner Group. I've worked there predominantly as their technical strategy planner, laying foundations and structures of technology that provide more efficient and effective working practices. I also seeded thoughts and ideas about how that company could move forward and exploit technology in areas that they weren't even aware were available or possible.

MM: Excellent.

Would you give us a quick overview of your new company, Bold Visions, in terms of its mission and maybe a couple of key values and client engagements?

CG: Sure. I established Bold Visions about 2 months ago, in May of 2008. Its goal is to help organizations understand that moving through life at this ever-increasing pace – putting out fires and making short-term decisions – is *not* the way to grow your business.

In fact, what we see more and more are businesses that think profit can only be achieved by reducing overhead – by making things more cost-efficient. Sadly, most of them ultimately discover that there is a finite limit to this 'cost-cutting only' approach. There are two parts to the profit equation: decreased overheads or at the other end of the scale increased revenues.

If you really want to grow your company, you need to have long-term plans – strategic plans that will allow you to progress incrementally to where you want the company

Correspondence:

Chris Glynne
Bold Visions Limited
E-mails: enquiry@
boldvisions.co.uk, www.
boldvisions.co.uk

to be in 5 or 10 years' time – and to grow your revenues accordingly.

The services that Bold Visions offer are *vision and strategic planning to align technology infrastructure in organizations with the business goals* of the management team. Most organizations do not have specific resources or skills to focus on this long-term planning. We are able to bring the expertise in this domain to bear in a prompt, pragmatic and personalized way. The value we bring is to leave clients with a congruent and sensible strategy for them to maximize their business potential. We also aim to leave those clients with a greater awareness and ability to continue molding and growing that technology strategy as the years move on.

MM: Would you take us through some of the major digital asset management (DAM)-centric projects that you accomplished at IPC Media?

CG: Well, back in 2000, one of my first major appointments was to head up the complete rethinking of the company's technical vision and strategy.

There were a number of different threads that came out of that piece of work, but two of the key threads were asset management and workflow management – which at the time, we thought, existed simply to make life easier for the users. We had been thinking in terms of alleviating pressure on the editorial users who had been dumped with more and more workload over the years. For example, having them perform many of the Desk Top Publishing (DTP) tasks historically sub-contracted out to reprographic studios.

At the time, we had thought of 'Asset Management' and 'Workflow Management' as two separate threads, and we were going to think about buying two separate and different solutions. We elected to start working on the asset management program, first.

MM: In retrospect, is that what you would do now?

CG: Absolutely not. We looked at the various different asset management providers and came across one particular provider – a company called Picdar with a product called Media Mogul. Through looking at that particular solution, we realized very quickly that DAM should not be considered as a silo solution. It *can* be put in as a silo solution, and may – in certain circumstances – provide useful benefits. However, the essence of asset management is

actually to have the assets be an end product of the business practices or the processes that that company employs to achieve their daily activity.

Particularly, at IPC – where their end goal is predominantly magazine production – the processes and the workflows that we're talking about are the life cycle of how a magazine is put together. Asset management is just one part. It's a fairly large part, but it's one part of that entire workflow process.

In hindsight, we realized that asset management and workflow management need to be part of the same solution. We then coined the phrase *workflowed asset management*.

MM: Why don't you just take us through what you mean by 'workflowed asset management.'

CG: I should perhaps explain upfront that when I talk about publishing I'm not strictly limiting it to magazine publishing – this could apply just as easily to newspapers, books and any kind of corporate publishing, as well. When you embark on that first seed of thought for a publication, there are usually a number of pieces of information – whether it be design briefs or instructions to various teams that will go off and start putting together the textual content or a photo shoot that needs to be commissioned. There are a number of pieces of descriptive information that need to be produced at the very outset.

Those pieces of information can actually be extraordinarily useful for their *context*, by way of metadata that could be applied to the actual assets that are subsequently produced.

That descriptive information usually grows from different sources of content. For example, the photographer who shoots any images can actually provide a lot of contextual description about the pictures that ultimately get uploaded from their photo studio.

With asset management solutions in general, one of the biggest problems is that whatever you ingest in has to have decent tagging, or else you can't search for it meaningfully at the end. All of those benefits you expect to get out of asset management by way of reuse and potential resale value to both internal and external clients ... All of that is a bit null and void if you can't find those assets in the first place.

Now, workflow management – at its heart – should cover the entire life cycle. It should include those first instructions of a particular editorial feature, let's say, all the way through as

the individual teams now start to create and pull together the various pieces of content.

One of those areas of creation is at the stage where all the individual pieces start to come together as a more collective piece – a page or web layout of the media.

As you then move on further still, you now start to add an element of almost metadata inheritance. Because those pictures are now on the same page as the text that's been written by the features team, that text now provides a rich descriptive metadata tagging for the images that sit on that page.

At a high level above it are the 'job-bag' style instructions for the actual feature name and the publication it sits in – even the administrative publication details. All of that starts to provide a better context upon which searches can be based at a later point.

MM: Could you define and put into a larger context the notion of 'inheritance' as it originally derived from object drawings and programming, applied to the notion of digital assets and metadata.

CG: In this instance, the words 'inheritance/association' could be used. We're really talking about assets that exist in one collective section of a publication. That section could be a page, it could be a feature that's a multiple number of pages long. However you describe that grouping of individual assets, there is an association in the fact that the picture sits on the same page as the text that describes that story. It therefore describes that picture.

In addition, it can work the other way around. If you have a certain number of keywords that are associated with a picture, and that picture is then placed on a particular page, then those picture keywords can then also provide a level of context to the page itself.

This centers on the fact that the descriptive metadata tags are applied to elements that sit within a larger collective group – a page or web design or whatever. That association between them is what I've described as 'inheritance.' It's such that as soon as you drop a new picture onto a page, any textual description on that page is now also going to be inherited by that picture that you've just dropped on the page. Does that explain it better?

MM: Perfect.

Earlier in this discussion, you talked about how when you look at DAM, you can look at it in terms of storage capacity. Or you can look at it in terms of a workflow. The correct way to approach this is really to look at it first and foremost as a workflow. Is that a fair characterization?

CG: Yes. I wouldn't necessarily say that the DAM part of it needs to be thought of in a particularly *different* way. But the emphasis needs to be much more about managing the workflow. If you manage the actual business processes – the workflow that takes place in producing your publishing – then you inherently need to store the assets as part of that workflow.

Asset management becomes a byproduct of the thinking of workflow management. It's not that you don't *need* asset management any more ... It's not that workflow management magically conjures it up out of nowhere. Basically, my belief is that the people that implement asset management systems as pure DAM solutions often end up creating more work for themselves – requiring more resources – particularly to meta-tag and structure the storage of all of those assets.

Part of the reason they suffer that problem is because they haven't actually thought about how that asset management system needs to sit cleanly and fluidly within their business practices.

MM: Something we've generally found in our interviews and research of the users of pure repository-oriented DAM is that inevitably – once they get their DAM up and running – they run into two major issues.

One involves the upstream creation process – which you've alluded to. Usually the toughest part of the upstream process entails the review and approval of assets that are in process. Sometimes the content creator has to go about creating media and copy and content in a new way.

Once we start creating workflows that reflect the prevailing practices of an organization, often workflows become locked in a certain way of doing things that ultimately makes the business even more difficult to adjust or realign to external market conditions.

CG: Yes.

MM: Could you speak to that? How do you build in the DNA for change – understanding that the future is nothing but change?

CG: You could talk about the movement of the business and how it might need to evolve as the months and years go by. But the other bit that we mentioned right at the beginning is about users and change-resistance and how they potentially can be the downfall of a project.

So, over-regimenting workflows by making them too prescriptive can also completely annihilate any chance of a project being successful.

On the other end of the scale, users – being human beings – tend to work in very different ways, if they don't have some degree of 'structure.' So the grey area between structure and regimentation – that grey area needs to be navigated very, very carefully.

I don't think there's any one prescriptive rule for everybody. It would depend very much on the type of business. This then comes back to what I'm trying to do as a consultant. Companies need to think about wider context and the longer term plan for what they intend their company to achieve over 5 or even 10 years, rather than 1 or 2 years – which is what a lot of people initially try to do when they put in a workflow solution.

First of all, that might seem like a little bit of a contradiction to what you're talking about. That is, that companies need to be flexible and move with business times – and to move with emerging technologies and emerging business and social trends. But what I'm really getting at for workflow management is providing a platform upon which the business can efficiently operate – whether that be how they operate now or how they need to change their operation for a slightly different media outlet in the next week, month or year.

The platform on which you build that workflow needs to be flexible in how it's implemented for your user. You need to be able to give users things they want to buy into, and in my opinion there are actually two sets of users. You have the people who are actually being creative and using it day-by-day for its intended purpose. You also then have the support people – usually IT, but not *necessarily* IT. It needs to be intuitive to both of those sets of users.

When we come back to flexibility and trying to avoid systems becoming inflexible, one of my key beliefs is that any solution put into a

company needs to be built in such a way that the supporting staff can make quick, simple and intuitive changes to the workflow.

If today we're publishing to a printed page, tomorrow we may switch over and we start using smart phones, for example, and there might be a couple of little tweaks that are needed to the overall workflow. Those should be quick and simple to implement at the solution level, because someone previous to that has invested good time and thought into building that entire solution on a strong foundation.

MM: Chris, would you speak to some of the changes underway in the magazine and publishing business that might then correlate back to this ability to make quick, bottom-up changes?

CG: There was some interesting stuff that Dave Cushman was talking about at the London DAM symposium. He was talking about social networking being the really big next media outlet. I think he calls it the eighth media outlet.

Publishers will buy into that – whether or not it is actually a new media outlet. It's certainly a way of getting opinion from much wider demographic areas. As magazine publishers, it's also potentially a way of getting our readers to actually start doing some of our work for us. That's quite an interesting concept. Being able to actually get them to bear some of the burden of creative overhead, at the same time as getting a much wider group of feedback about any particular brand or field that you might be covering in your publication.

The whole area of social involvement in media is very interesting. Certainly, workflow solutions that we talk about now are predominantly based on how we as a company would operate, and how we would deal with content that has been created usually in-house, but not necessarily only in-house.

Some of the wider parameters of a workflow management system over the next few changing years, for example, might need to include social networks as methods of receiving different pieces of content. Obviously, when you start opening up social networks, you're talking about millions, even billions, of people, potentially providing bits and pieces of content. I'm not entirely sure how solutions such as asset management then start containing all of that. That whole social environment – the movement

towards everyone being connected and everyone being a publisher in some shape or form – that starts to pose the very harsh questions of rights management and rights as a whole subject.

MM: It seems to me, Chris, that you have two distinct and separate loosely coupled ecosystems. You have one ecosystem that we could characterize as the ‘walled garden’ of policy-managed content and assets. That’s the DAM system.

Then you have the open frontier. That really has very individualistic and parochial, idiosyncratic policies, managing contents and assets. Those include websites as well as blogs and forums and social networks.

It seems to me that there’s a bridge between the two. The notion of extending the policy of your walled garden to the frontier is probably a hopeless cause, in that the frontier is beyond the ability of any one system to manage.

It seems to me that the open frontier comprised of millions of chiefdoms would then entail some sort of social media analytics capacity. Whereby you go off and catalog and profile and develop metadata portraits of what’s out there. You really point to them, as opposed to ingesting them and pulling them into an asset management system.

Therefore, the editorial operations of publishing firms really have two wells of assets and contents from which to draw. One, from the walled garden of policy-managed assets and content, and then the open frontier, where basically you have a thumbnail, a quick little profile of what constitutes or what’s relevant in that thumbnail.

You’ll leave the management to spiders and analytics suites – as well as the people who actually own and control the content. Does that make sense?

CG: Yes. It does. I agree, actually, with that whole scenario. I think that if publishers want to stay around, publishers have to get back to some of the core aspects of what people are willing to invest in – whether it be investment of time or of money.

In my mind, it breaks down into four things. You have ‘Products,’ which are going to be physical – tangible things. Whether it is a phone or a magazine or a book. ‘Services,’ where people provide you something that you could do yourself, but they might be able to provide it quicker or at a much vaster reach than you’re

capable of doing. ‘Experience,’ which provides a rich engaging encounter for the user. Lastly ‘Knowledge,’ which speaks for itself.

I think that with the scenario you’ve presented, publishers do have a vast array of their own content. But it’s fairly silo-created and insulated. In theory, if you’re a great journalist, you’re going to use as much information to gather as wide an opinion as possible. But there’s no greater breadth of opinion than the whole new ecosystem that’s spread out from some of these social scenes. That provides publishers with a much larger wealth of information.

The problem is, when you have huge quantities of content, how do you get through all of it to sort the wheat from the chaff? When you talk about analytics I think that’s going to be a crucial thing. Most people will want to consume consolidated content ... publishers have an opportunity to provide that service and present it as something that is artistically designed and told as a story. That’s going to require solutions at corporate level that can sort that wheat from the chaff and automate as much of the packaging as possible.

Human beings are becoming more inclined towards convenience, and also more inclined towards personalization. We want exactly *what* we want *when* we want it and via whichever consumption channel is the most convenient to us at the time. I actually see publishing drifting more towards personalization – individual demographics. So as a consumer, on whatever service it is that I’d choose to go to, I would be able to log in and have individualized content served to me.

The service and the experience it’s offering me is something I could do myself if I’d bothered to troll the Internet and all the various blogs and whatnot. I could do it, but it would take me so long that it wouldn’t be a decent investment of my time. As a result, I’m prepared to invest an amount of my money to have that service done for me.

MM: That would suggest, from an infrastructure and process perspective, a digital supply chain of several loosely coupled systems.

But it seems to me that individualizing content to users requires some of the following capabilities:

- A more granular multi-dimensional database profile of what a user wants and will consume.

- A method of tagging content and assets in a far more granular way, such that it probably entails some sort of text-mining, semantic tagging and creating very granular topic maps beyond just topics, places and products – to extend to things such as context, social context, meaning, concepts and related concepts that would only develop through some sort of thesaurus-like taxonomy.
- The editorial and content and assets that you have must then live in some sort of dynamic publishing system such that the system then has the ability to dynamically push granular pieces of information or content to me within a framework.

CG: Yes.

MM: That also suggests that as the iPhone is transforming, it is leading the way in terms of a radical simplification of user interfaces.

Basically, it's forcing publishers to strip away a lot of extraneous stuff – navigational stuff or simply advertising that's cluttering up the page.

This then requires rethinking how we assemble content, which puts greater agility on the content management system and the systems that are feeding it. Then ultimately, we're talking about specialized databases. XML databases for editorial text, and multimedia asset repositories for multimedia assets.

Then, how do we put things into those repositories, so that it doesn't entail a lot of manual reconditioning or reprocessing in order to make them more fluid for various presentation contexts?

CG: When I first came up with this concept some 8 years ago ... in my own mind, at least, that's not to say I was any kind of pioneer. I'm sure lots of other people were thinking the same.

MM: Well, you were. Let's call it what it is.

CG: The thinking at the time was that if we have this bloody great store of digital assets, and we're populating it using all of the rich context that we can glean from the way the workflow actually progresses, then it's not a great leap of faith to be able to put some kind of intelligent semantic search and text-mining type of technology in place – the likes of which have already existed with people like Autonomy for a good 10 years. I'm sure there are other tools like theirs as well.

It's not a great leap to say, 'Well, we can get at least the start of a personalization agent' –

which I think is what Autonomy used to be referred to as. That will then allow me to log in and start building a profile about myself.

Then specifically, the important bit is to have this engage with a user. We all know now that users around the world are prepared to put some effort into these kinds of social scenes. People do leave comments on blogs. People do invest time participating in forums. Part of that is basically just about feedback. 'What have I seen? Do I agree with it? Have I got anything more to add?' The Wikipedia-type world is also based on communities trying to come together to turn small numbers of facts into large, concise and well-structured end stories.

When a user interacts with a personalization engine, the key is making sure that it learns properly about them by understanding some of the feedback cues that consumers would be giving.

That has a lot of different springboard offshoots from it. One of them is that the user gets more benefit, because that profile is now better tuned to them in the future. But the providers of the content also get some more advantageous feedback about what stuff they should concentrate more time and effort on.

Everyone else in the community starts to get some advantage from the simple types of techniques that even Amazon used over a decade ago. If you buy a particular book, you might also be interested in this book, because everybody else that bought the one you're buying also bought that one.

MM: Now you're talking about a kind of new, expanded level of managing metadata frameworks or metadata schemas.

You have metadata that describes not just the user, but who the user is as an economic actor within various theaters – theater in terms of who this person is in a demographic context.

That also then requires that you have a much more granular set of metadata – specifically – about the inventory of content that you have.

CG: Yes.

MM: So this becomes a much more granular, detailed, nuanced and faceted taxonomy of what's in the content.

The third set of metadata you then have is a far more granular, complete set of metadata about your inventory of ads. Both advertising that's secured through your advertising reps, as

well as ads that are then syndicated in through ad-publishing networks.

You have these three sets of metadata.

Metadata about the user in terms of information and consumption, metadata about the content and its ways of consuming and metadata about the advertising inventory. Somehow, those three sets of metadata needed to get synchronized in terms of a seamless user experience.

How do you do that?

CG: Today, I'd probably say that it's impossible in anything other than a small portal type solution. I think some of the larger publishers could probably meaningfully do it for a growing subscription market. I would suspect that it would take a reasonable amount of manual effort for a lot of that categorization that we've talked about there.

Having said that, we can look at some of the technology that is available even today, using semantic logic and Bayesian probability theory. All of the stuff that sounds a little bit like rocket science, but actually is basically the way that the human brain is thought to work ... There are certain elements – and to a certain degree of complexity, it can probably be done today. In the next 10 years, inevitably, the capability will have risen by another order of magnitude, at least.

The point that I try to make with publishers isn't necessarily that they can have nirvana right now. It's that in order to get to nirvana in the future – if that's where they would ideally like to be – they have to start taking steps towards it.

Some of those steps are small baby steps, but they have to start getting on that learning curve. That is what will allow them to witness some of the pitfalls that they'll see along the way, and will allow them to test and feed back to the technology providers and developers of this world, such that things can be evolved to the point where the solutions can be as granular or as complex and as wide-reaching as that kind of scenario that you've just painted.

Actually, there's one other bit that we haven't really mentioned, here ... particularly on the advertising side. When you potentially have consumers interacting through portals, and if the portal is completely aware of what that particular consumer will buy or not buy into, you've got the perfect opportunity for direct advertising at a much higher yield level than

current advertising models enable. Current advertising has a scattergun approach, blasting it out to a demographic that generally has been responsive to this kind of product – and you'll get x -percentage yield from that approach. If you now know that a consumer loves Apple, for example, and you have more specific detail about *why* they love it, then you know you can target that consumer with a much greater expectancy of yield from it.

MM: Yes. This sets up what I've begun calling a 'contextual consumption' of content and advertising.

CG: Yes. Absolutely.

MM: Specifically, contextual consumption requires a growing level of metadata about users, advertising inventory and content. It also requires the ability then to semantically tag content so as to match the consumption requirements of the consumer, as well as the relevance of ads to that particular consumer. Not specific to the content being consumed, but specific to the larger picture. In other words, 'Who are you as an economic actor?'

We can no longer afford to think of the web as simply the 'online channel.' We have to adopt what I call an 'ecosystem' strategy. An ecosystem strategy dictates that an increasing amount of content and services that I'll provision into my engagement theater will come from third parties, and will live some place in the cloud.

So I need to have a sourcing framework for how I find and provision third-party content and services.

That requires that I have a much more open, agile, technical infrastructure by which I can quickly integrate content and services to my engagement theater.

Can you just take us through that a little bit, as a summarizing point?

CG: Well I think the concept of 'the cloud' is something that most people have now finally latched on to. Wikipedia is a fairly good example of how multiple people from all over the place are putting together content-rich, factually correct – for the majority of the time – summarizations of bits and pieces of stuff that are all over the 'cloud.'

That's not to say that the bits and pieces can no longer be gotten that way, because there is already a linking structure in place.

We already have certain frameworks in place, as well. If we look at the markup-type language – XML – that’s already fluid enough and scalable enough to potentially provide some of that mash-up linking. There are already interfaces that allow people to look at all of the elemental parts of a particular subject that they might be looking at or researching. To get all of that content aggregated into one place.

Even if you use things such as Really Simple Syndication (RSS) feeds, it’s a very naïve way of thinking about that, but nonetheless, they provide a facility for you to aggregate content very quickly and easily as a user.

The seeds to all of these kinds of things are already here. Like everything in the world, the first iteration often never quite hits the mark. It never quite achieves the goals it originally set out to do, because it was maybe a pioneering concept.

But initially, you also don’t understand all of the potential outcomes – all of the potential ideal end results that you could achieve – because you just haven’t played with it enough to start with.

When people started working, for example, on XML, I wonder whether they realized that it could become a completely universal translation language. When people started working on the Internet, they didn’t realize it was going to become what it is today. A lot of these technologies – these formats, languages and technologies – are available today. They are the seed points for another couple of iterations and another couple of sets of people aiming for the stars and maybe only reaching the moon – figuratively speaking.

In that process, those technology ‘seeds’ will grow into things that are more attuned to what we’re potentially talking about today.

I have no doubt that in the next 10 years, if not 5, we will see those technologies get to the point where semantic analysis does become a feasible prospect in real-time. We will start to see the cloud develop to the point where it does provide a multidimensional and multifaceted asset management repository in the sky.

MM: That sounds like a great place to leave this.

Chris, thank you so much.