# The road to data maturity - Why publishing has to stay relevant in the digital age and why it needs to start now

Sam Herbert[*]

*Co-Founder, 67 Bricks, Attlee House, St. Aldates Courtyard, 38 St. Aldates, Oxford, OX1 1BN, UK*

**Abstract.** The world is rapidly changing driven by the rise of data and Artificial Intelligence (AI). Many industries are facing wholesale change. Customers now expect tailored, digital solutions and personalized insights delivered at the point of need - and structural transformation is simmering as a result.

This paper looks at the sea change and opportunities in the academic publishing sector. Driven by technology and changes in user experience, publishers need to harness technologies such as AI to help them evolve from being traditional content providers to data-driven product companies. Data is at the heart of everything. Increasing their data maturity will fuel innovation and power the transformed products and services that customers demand going forward.

As part of this feature 67 Bricks will also introduce the Product Development Data Maturity model for scholarly publishers to help plot a course from where they are now to where they need to be to survive.

Keywords: Data maturity, software, development, information, product, 67 Bricks, artificial intelligence, data mining, digital data management, publishing industry

## 1. The rise of data and AI

- **Here and now:** *We have supercomputers that have more processing power than the human brain and hold ten times as much data*
- **Accessible:** *In the next five years, we will have a $1,000 computer that will exceed the processing power of the human brain*
- **Capable:** *In the year 2050 we expect to have a $1,000 computer that will exceed the processing power of the entire human race*

Welcome to the brave new world. The advancement of algorithms, cloud computing, and the sheer expanse of data have reached a tipping point and now is the time to innovate in your sector - or risk being left behind. Technologies such as AI and machine learning are no longer emerging, they are established, and pioneering industries such as science, healthcare, and aerospace are among those who have changed business strategies to become more data-driven - with great results.

---

[*]E-mail: sam.herbert@67bricks.com.

## 2. Pioneers leading the way

Manufacturers General Electric wanted to understand why some of it's jet engines developed problems before others. A single jet engine - generates 20 TB of data per engine every hour, considerably more than Twitter generates over the course of the year. General Electric used its data to identify and solve its initial problem - and in the process discovered a whole new business, driven by data. The new enterprise is more profitable than the original manufacturing business [1]. As a result of capturing, analyzing, and acting on this data the jet engine industry has been able to change rapidly.

In the healthcare sector, a test was carried out [2] by the Google Brain team [3] to see how an AI analysis would compare to a human expert when diagnosing breast cancer. It turns out that the AI was substantially better at correctly identifying breast cancer from scans. What is so interesting is that it was the data work that was the challenge in this activity, not the writing of AI algorithms. That was relatively straightforward. The hard work was to turn the scans into data that the computer could use. In more recent studies Google claims 99% accuracy in metastatic breast cancer detection.

## 3. Resisting change

But not all sectors want to embrace change. We work in the academic publishing industry where many remain staunch - continuing to see themselves as traditional 'content providers' rather than as data-driven product companies. Risk averse leaders argue: *"Scholarly publishing is different to other industries..."* *"We have survived so far without big change..."* But the advancement of cloud computing, the power of algorithms and data, alongside increasing user expectations has made change unavoidable.

Some players are embracing it. SAGE, Springer Nature, and Elsevier are among those reinventing themselves by moving from a document-led approach which centers on providing books and journals as 'units of content', to a data and knowledge-led approach. The mergers and acquisitions (M&A) market is busy as publishers invest in specialist Information Technology skills and platforms to reposition themselves. These efforts are encouraging. However, to begin this journey to modernization on a sure footing, publishers need to pay heed to the *essential fuel* behind the apps, hacks, and shiny re-brands.

- How best do they store, manage, enrich and utilize that data at the heart of their business?
- How best do they optimize that data and serve it up to a new type of customer, one that expects digital, personalized products as standard?

Master these key elements and survival during the next period of change is more likely.

## 4. Where are you on the road to data maturity?

At 67 Bricks we have developed the Product Development Data Maturity Model so organizations can measure their data maturity and get a baseline view on their raw content assets (Step A). What shape are they in? Where are they housed? How are they managed and utilized? Once companies have got a true measure of that, they can move on to scoping out the steps they need to take to having rich, connected data that can drive a valuable information product (Step E).
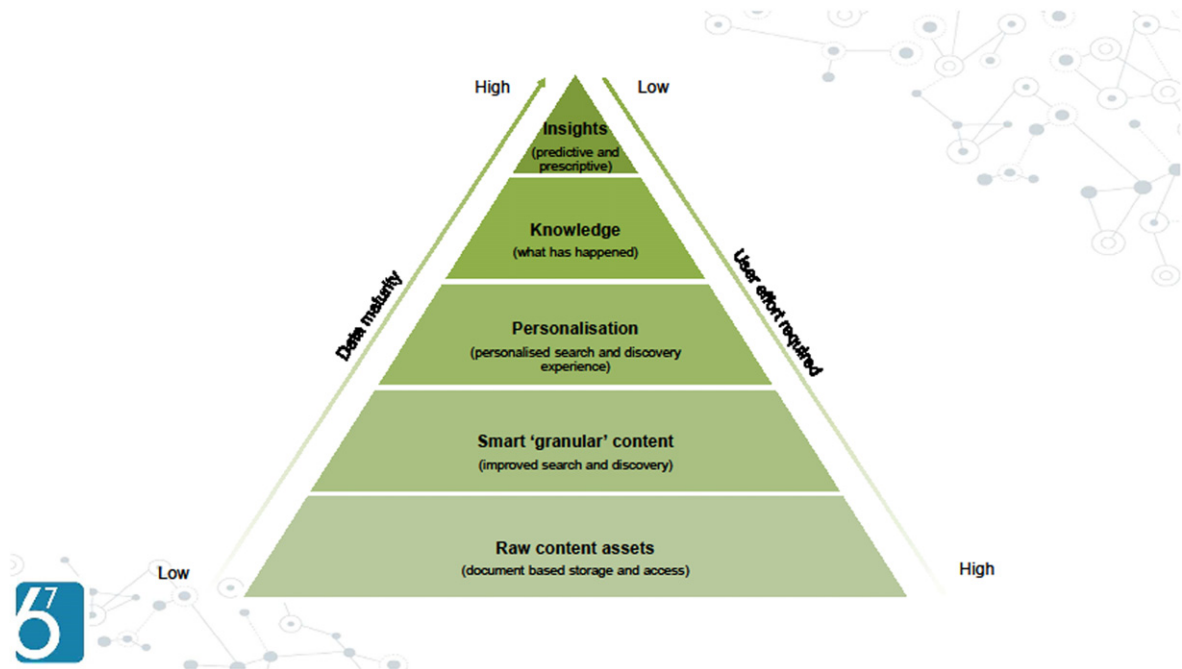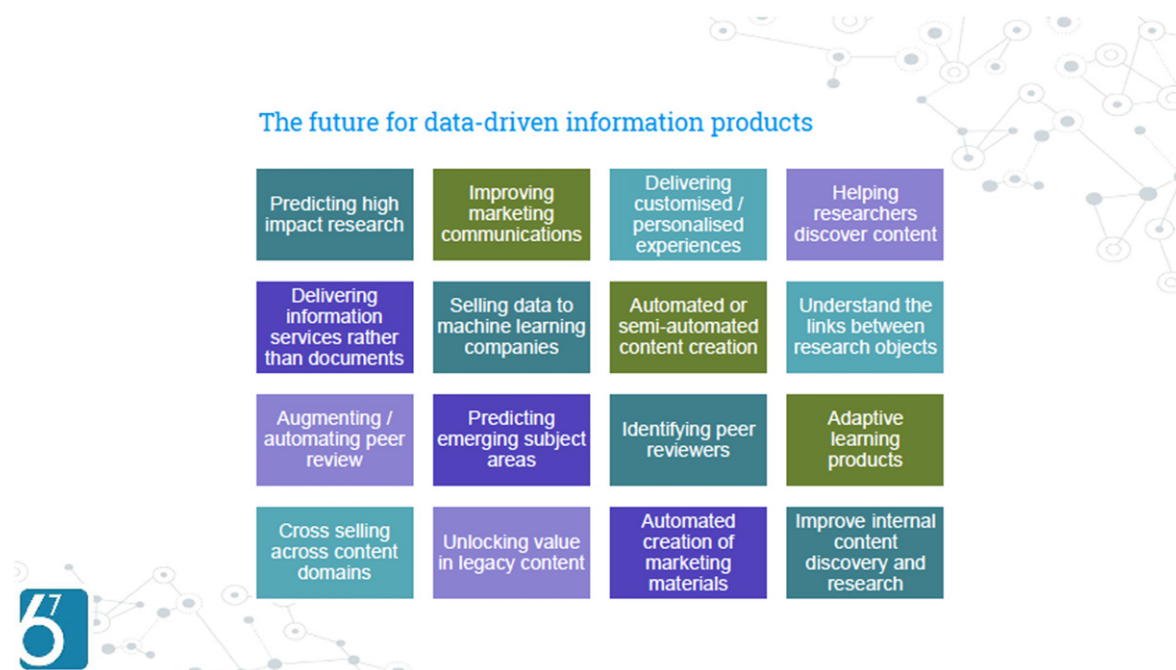
## Product development data maturity model

| Level | 1: Raw content assets | 2: Smart 'granular' content | 3: Personalisation | 4: Knowledge | 5: Insights |
|---|---|---|---|---|---|
| Store and manage | Document level metadata<br>User access rights data | Granular content 'chunk' level metadata | Semantic fingerprints for users<br>'Integrated' usage data | Data relationships (e.g. as triples) | Large primary and supplementary data sets |
| Extract and create | Manual metadata creation during editorial processes | Semantic fingerprints for content items | User interest metadata<br>User 'cohort' analysis | Knowledge extraction - as relationships | Predictions and recommendations |
| Data use | Access control<br>Document collections | Granular search<br>Faceted search<br>Proximity matching<br>Usage trending | Personalised proximity matching<br>User type analytics | Powerful knowledge query capabilities | Predictive and prescriptive analytics |
| Product value | Find and access documents online | Enhanced search and discovery<br>Slice and dice content products | Personalised discovery experience | Answer questions<br>Explain what has happened | Explain what will happen and what to do about it |

This may feel like a daunting task, but not only will a new focus on better utilized data help retain current customers, it will also open up opportunities for new partnerships, products, and opportunities.

## 5.  User experiences driving change

Traditionally, academic publishers have not focused well on user experiences; understanding what the researcher is trying to achieve, or working out better ways to help them get it. They have seen their value as being providers of long-form content rather than as providers of knowledge or personalized product experiences. There are a few reasons for this; first, budget holders have typically been librarians rather than authors and researchers (the end users of the content and data) which means that publishers have always been one step away from their users. Second, the traditional consumer, the author, has been tied into the traditional model of working with academic publishers for their career progression - so there was little need or demand to change it.

However, this situation is changing rapidly. The rise of Open Access, where authors pay an article processing charge (APC) to publish, essentially becoming both a customer and a budget holder, has changed the playing field. Plus, the next wave of authors and researchers are the smartphone generation. Regular users of Google, Uber, and Amazon, they are a group of users who are used to a personalized, fast, seamless user experience, and digital products which provide rich knowledge, tailored to their specific interests and needs. Their end goal is the same - *getting published* - but they want a different user experience - delivered better, smarter, and online. If they don't get this from their current academic publisher they have other, viable options. They can move to a more digitally-minded publisher or they can find any piece of content they want on SciHub in a matter of seconds. Open Access, Plan S, funder power, and questions about the peer review process all mean users are demanding to be heard and put "front and center" in a different way, and publishers are being forced to change the way they do things. Companies need to ask themselves; have I done the groundwork on this? Am I moving fast enough to give users what they want?

## 6. Same goal, new user experience



© 2019 67 Bricks



© 2019 67 Bricks

It is important to understand that in a period of change it is the development of new user experiences where the battle lines are drawn. Amazon, Uber, and Google did not change the outcomes for users, they simply changed how users *achieved* those outcomes. For example, with Uber you still get from A to B, but it is *the experience* of achieving that outcome that has created a world-dominating transportation company that has massively disrupted an existing industry. This is now being replicated in other sectors. And those that don't keep up, risk their survival.

## 7. No going back

The data and technology is ready to go. Customers too. But what about the publishers themselves? There is little doubt that maturation and dramatic change of this kind in a traditional industry can feel brutal. People don't want to leave behind the uniqueness of what they know and love. Earlier this year, the Chairman of Elsevier, Ys Chi, presented at the ITHAKA's Next Wave Conference in New York and admitted, in a keynote speech [4], that the initial internal push back to the new data-driven and product-first mindset was "violent and robust" at times. But Elsevier embarked on a cultural and business turnaround and have breathed new life into their business. It is an exciting and challenging time for the scholarly publishing industry and we are hugely motivated to play a role in its continued evolution. We hope to meet many of you who embark on similar journeys.

## About the Author

Sam Herbert is Co-Founder of 67 Bricks. Sam has the rare ability to recognize each client's needs and help them unravel the complexities of today's technology, and build products and platforms that deliver compelling user experiences. He has worked with some of academic publishing's most respected firms including SAGE, Emerald Publishing and Wiley. Tel: 07734138274. Email: sam.herbert@67bricks.com.

## References

[1] J.M. Immelt, How I remade GE, *Harvard Business Review* (2017), https://hbr.org/2017/09/inside-ges-transformation, access September 8, 2019.

[2] Y. Liu et al., Detecting Cancer Metastases on Gigapixel pathology Images, March 8, 2017 (version 2), https://arxiv.org/abs/1703.02442, accessed September 8, 2019.

[3] Google Brain, Wikipedia, https://en.wikipedia.org/wiki/Google_Brain, accessed September 8, 2019.

[4] https://scholarlykitchen.sspnet.org/2018/12/10/elsevier-chairman-ys-chi-interview/.