

# Value optimisation in construction, from Building Information Models to Big Data

Benghi, C.<sup>1</sup> and Williamson, L.<sup>2</sup>

<sup>1</sup> Department of Mathematics and Information Sciences Northumbria University, Newcastle Upon Tyne, UK.  
claudio.benghi@northumbria.ac.uk, Tel: +44(0)788 249 0525

<sup>2</sup> MSc Project Management candidate, Northumbria University, Newcastle Upon Tyne, UK.

## ABSTRACT

### Purpose of this paper

The paper starts from the analysis of the benefits achievable and expected from the use of Building Information Management (BIM) in construction and aims at identifying and prioritising expected trends in the area.

### Design/methodology/approach

The methodology adopted starts from a systematic review of current academic literature, for the identification of the state of the art, and it then moves to the identification of trends by the application of two further biblio-metric techniques; the first one provides reference to government and institutional policies in the area of construction and discusses their correlation with the benefits identified in the first analysis and the second one analyses the corpus of the US National Science Foundation grant awarded in the area of data-driven science to identify efficient principles to be adopted by players interested in the development of data-driven exploitation of Integrated Project Delivery and BIM for construction value.

### Findings and value

The paper lists strategic criteria to drive the development for the next generation of integrated speciality IT components in consideration of readiness factors for artificial intelligence supported value extraction.

### Originality/value of paper

The paper is original in the identifications of extremely recent trends and technologies and does so on quantitative grounds.

**Keywords (no more than 5):** multi-dimensional BIM, Integrated Project Delivery, Data driven business, Integration, Big Data

## 1 INTRODUCTION

A humorous quote disputably attributed to Nils Bohr states that "Prediction is very difficult, especially about the future", yet prediction is central to most human endeavours, and it is the centrepiece of all areas at the core of total cost management: whether we are performing a feasibility study or a whole-life cost appraisal we need to rely on the accuracy our predictions.

This challenge is even further complicated by the ultimate desire of clients and consultants to transcend mere cost metrics and adopt value metrics such as benefits and scope for the evaluation of return of investment, thus extending the problem domain to the intricate and surprising dynamics identified in the work of Tversky and Kahneman (1999) on behavioural economics eventually leading to the 2002 Nobel prize in Economic Sciences.

In years of financial expansion, the market as a whole self-organises to accept large uncertainty margins of error in cost predictions, but - with the reduced business margins resulting from the current financial crisis - it immediately emerged that the path to restore businesses sustainability in the construction industry passed from the need for an holistic approach capable of comparing design, construction, operation and business costs with the value delivered by construction benefits

(Wolstenholme 2009); the simplified heuristic solutions that the industry adopted for many years to cope with the inherent uncertainty of consolidated processes have suddenly become insufficient.

Building Information Modelling (BIM) and Integrated Project Delivery (IPD), after many years of incubation in small niches, have been identified as enablers to fill the knowledge gaps between all the stakeholders of a facility, forming the basis for objective decision making during its life cycle from inception onwards.

Rich 3D models incorporating cost, planning and resource management domains provide one obvious advantage: they offer the infrastructure to collect the contribution of all relevant stakeholders to the benefit of uncertainty reduction in all deterministic cost estimation processes; however more radical approaches can be envisioned in the medium and long term that could further alter the discipline of TCM when the information domains will extend to the almost uncharted territories of Carbon Costs and Operational Facility Management.

This paper will start from the analysis of the state of the art in multi-dimensional project control and attempt to predict the evolution of TCM dynamics in a scenario characterised by (1) BIM-enabled operation analysis of smart-buildings supported by (2) current development in the areas of Big Data and Machine Learning (ML) algorithms; particular attention will be paid to the identification of enabling technologies in order to support the readers in the definition of roadmaps towards the goal of cross-domain value optimisation.

## 2 BENEFITS OF BIM IN CONSTRUCTION

To setup for the assessment of recognised benefits that BIM and IPD bring we have adopted a quantitative biblio-metric method. A range of peer reviewed and commercial publications has been identified by merging subsets from different sources provided in Table 1.

**Table 1** Identification of suitable publication groups.

Group name	Process adopted	Number of items
Automation in construction	The online repository of "Automation in Construction" has been parsed for papers referring to explicit the benefits of information management for construction processes.	22
Benefits	Google scholar selection of full-paper articles with the keywords "BIM IPD Benefits"	21
Value	Google scholar selection of full-paper articles with the keywords "BIM IPD Value"	9
Commercial/Policy	A small number of selected commercial and governmental policy sources have been identified starting from Google search results for "BIM benefits"	7
Books	A selection of monographic publications available in the university library in hardcopy	5

The selection has been produced with the intent of representing both the academic state of the art and the information that the general public would get if they approached the problem through different means in the July of 2014.

Throughout the selection process care has been placed in filtering out contents that would be considered surpassed as reflected in the time classification of items shown in Table 2.

**Table 2** Time distribution of selected sources.

Year	Count
2014	17
2013	12
2012	11
2011	9
2010	5
2009	5
2008	3
2007	1
2005	1

The identified sources were analysed and coded against a set of 25 benefits and 7 areas of benefit through excerpts of the text; the final database contains 519 mappings that can be summarised in 463 associations between concepts and sources, then counted as follows:

**Table 3** Areas of benefit and their relevance through reference count.

Area of benefit	Reference count
Collaboration	34
Project Management	24
Communication	17
Facilities Management	16
Design	15
Integration	13
Construction Management	12

**Table 4** Benefits and their relevance through reference count.

Benefit	Reference count
Visualization	27
Cost reduction	26
Efficiency improvement	22
Process Coordination	21
Time reduction	20
Accuracy improvement	19
Sustainability improvement	17
Quality improvement	15
Errors reduction	14
Clash Detection	12
Planning improvement	11
Value discovery	11
Change control	11
Improved Decision Making	11
Risk reduction	10
On-site Productivity improvements	9
Life-cycle management	8
Improved Safety	8
Waste reduction	7
Improved Understanding	5
Rework reduction	5
Analytic benefits	4
Prefabrication improvement	3
Increased competitiveness	2

An analysis of the presented rankings reveals a number of considerations that might resound with intuitive perceptions of the industry:

- The substantial primacy of Collaboration and Project Management in the ranking for Areas is yet another indicator of the fragmentation of the industry along with the awareness of its relevance;

- The identification of “Facilities Management” before “Design” and “Construction Management” reflects a growing awareness of the importance of better integration across all stages of procurement, and possibly the completed achievement of the latter two benefits, now given for granted.
- When it comes to the ranking of specific benefits, “Visualisation” and “cost reduction” significantly distance all competitors; it is possible to argue that their relevance stands for quite different reasons: while the first points to an effective method of communication that recent technological improvements have made mainstream and current in the industry, the second seems to point to the irrational prevalence of stand-alone cost factors in the determination of project success, as opposed to the more rational perception of cost dynamics in the balance with time and quality criteria, that have been relegated to much lower ranks in the chart.
- “Life-cycle management” is proportionally lagging behind the ranking of “Facilities Management” in the area chart; this might signal a skewed perception to isolated benefits in the area of FM achieved by better IT support, but unrelated to a more integrated process of data-retention and feedback loop that the idea of “Life-cycle management” seems to suggest.

### 3 THE EDGE OF DATA-DRIVEN INNOVATION

To provide an insight into the future of information science for data driven process improvements in the medium term the database of grants provided by the US National Science Foundation under the terms of the Data Infrastructure Building Blocks (DIBBs) programme, as published by online on NSF (2014a), has been analysed with similar criteria of the previous literature groups.

The database selection appears appropriate because of the strategic aims of the programme as summarised in a recent press release perfectly aligns with the expectations captured by the data driven processes that BIM and IPD promise to enable:

“The ability to collect and analyze massive amounts of data is rapidly transforming science, industry and everyday life, but what we have seen so far is likely just the tip of the iceberg. Many of the benefits of “Big Data” have yet to surface because of a lack of interoperability, missing tools and hardware that is still evolving to meet the diverse needs of scientific communities.”

(US National Science Foundation 2014b)

and looks at enabling its vision through the funding of projects for the total amount of 191,553,617 USD this far with the following scope:

“The DIBBs program encourages development of robust and shared data-centric cyberinfrastructure capabilities to accelerate interdisciplinary and collaborative research in areas of inquiry stimulated by data.

Effective solutions will bring together cyberinfrastructure expertise and domain researchers, to ensure that the resulting cyberinfrastructure components address the researchers' data needs. The activities should address the data challenges arising in a disciplinary or cross-disciplinary context.”

(US National Science Foundation 2014)

The classification process on the documents downloaded from the side resulted in the identification of 77 codes directly attached to the content of project abstracts. These were subsequently grouped into 3 representative categories on commonality: data, infrastructure and intelligence. Their relevance, as indicated by the count of associations with projects is presented in the following tables 5, 6 and 7.

**Table 5** Infrastructural development areas for data-driven science and relative relevance from reference counts.

Infrastructural development area	Reference count
community	52
infrastructure	46
sharing	29
performance	19
preservation	14
long-term	14
web	13

scalable	12
workflow	12
dissemination	11
cloud computing	11
expansion	9
publishing	9
massive	8
interactive	8
cost	8
privacy	6
interface	6
security	4
provenance	4
long tail	4
hardware infrastructure	4
data preservation	2
citizen-science	2
network infrastructure	1
social networking	1

A few considerations could be useful to support the interpretation of the infrastructure rankings:

- The opposite standings of “community” and “social networking” targets highlight how the IT industry is still substantially concentrated on the exploitation of the stream of societal and business changes sparked by the pervasive spread of internet access boosted in the 90s by the popularity of web, and that, however, social networking is not perceived a crucial element of further development, possibly in favour of more structured communities of practice.
- The combined standing of “infrastructure”, “preservation” and “long-term” is also indicative of the growing recognition that legacy code and legacy data have become a substantial burden for any IT development project, from which the push to develop structural patterns and infrastructures capable of dealing with the obsolescence of technologies while retaining business intelligence solidity.
- Projects connected with the keywords “performance”, “scalable”, “expansion”, “massive” and “hardware infrastructure” recognise the potential of adopting progressively more sophisticated analytical tools that leverage big data for the provision of artificial intelligence behaviours; they therefore push on the development of infrastructural solution for available and affordable computational power.
- At lower rank standings finally appear areas connected with human computer interaction patterns, that used to be prevalent less than a decade ago; in this area while there is still room for improvement, the keyword “workflow” ranking higher than the combined value of “interactive” and “interfaces” signals focus on automated workflow negotiation between human and computer processes that appears to be the route for a progressive marginalisation of human intelligence in the mix, as suggested by the notes on the Intelligence set that follow table 6.

**Table 6** Intelligence development areas for data-driven science and relative relevance from reference counts.

Intelligence development area	Reference count
analysis	35
model	26
big-data	17
integration	16
complex	16
decision	13
simulation	11
quality	11
patterns	11
scope	10
crowdsourcing	10
analytics	9
evidence	8
curation	7
value	7
interoperability	6
reliability	6

testing hypotheses	5
machine learning	5
reproducibility	5
data mining	5
cognition	4
validation	4
automated data sensing	4
semantic	3
correlation	3
intelligence	3
verification	2
uncertainty	2
statistical	2
multidimensional	1
digital library	1
estimation	1
network dynamics	1
confidence	1
granularity	1
persistence	1
data fusion	1
accountability	1

With regards to the ranking of the “intelligence” area:

- The first seven higher ranking keywords, perhaps not surprisingly, almost read like a complete sentence; in the area of intelligence the aim is to develop “analytical models [that leverage] big-data integration [for] complex decision [making through virtual] simulations”. This is a radical shift of approach from only a few years ago, where business intelligence was mostly about the provision of appropriate reports for human decision making. Recent commercial and technical successes leveraging big data for Artificial Intelligence purposes have restored faith in the area after the disillusion of the long “AI winter” and the continued investment of the IT global firms in the area is likely to produce the scenario that will avoid a second one. (Hendler 2008)
- Indeed many of the second tier priorities in the list seem to address the reduction of human control that the suggested adoption of AI tools would entail. Projects associated with the keywords “analytics”, “evidence”, “curation”, “reliability”, “reproducibility”, “data mining”, “cognition” and “validation” concern themselves with the epistemological problem of relinquishing decision making and automation processes to algorithms while retaining some level of confidence on their reliability.
- Finally, the presence of the “crowdsourcing” keyword complements the picture by allowing the delocalised integration of on-demand outsourced human intelligence components amidst the proposed workflow orchestration services.

**Table 7** Data development areas for data-driven science and relative relevance from reference counts.

Data development area	Reference count
time series data	27
spatial data	20
sector specific data	12
generic data	6
data accuracy	5
civil engineering	3
life-cycle data	3
3D data	2
social sciences	2

In the area of data development only a short list of keywords have been identified, however it still allows the production of, perhaps interesting, considerations:

- The prevalence of time-series data can be directly connected to the centrality of time-series information for any prediction and decision making mechanism, but reading through the proposals we have also detected a clear concern for the ability to preserve the enormous wealth of data of legacy system and retain it, with adequate time tagging, for online analytical processes.

- “Spatial”, “3D” and “civil engineering” ranks, if read together, also assume a relevant role in the ranks that is connected with the acknowledgement that even in a globalised market efficient management of geospatial data allows for substantial business opportunities in the area of cost optimisation on one side, while on the other, they can provide an essential framework for the support of strategic investment decision making at governmental level in the areas of civil networks and infrastructures.
- Finally it has to be noted that sector-specific data frameworks proposals, in spite of the obvious limitations they have, constitute a greater proportion than general purpose data infrastructures; this clearly signals that there is still substantial road ahead in the conceptualisation of an all-encompassing general-purpose data management framework; artificial intelligence processes of the coming generation are going to be narrow in scope, but integrated in nature.

#### 4 CONCLUSIONS

The time of Data-driven decision making in the construction industry has arrived; the maturity of data schemas, the availability of specialised analytical tools along with the relative low cost of computational power for required algorithms and the diffused availability of sufficient network bandwidth for communication and collaboration exchanges have provided the fertile environment on which authoritative players such as government agencies and professional bodies have been producing supporting guidelines to integrate the workflow of the still fragmented supply chain; amongst these it's worth mentioning:

- The work of RIBA and other chartered bodies in the revision of guidance for the implementation of efficient coordinated workflows (RIBA 2012; Sinclair 2012; RICS 2011)
- The effort of international bodies in the production of effective, albeit limited, standards for interoperability (Liebich 2010; USACE/NASA and East 2012; UK BIM Task Group 2012), coordination and validation (buildingSMART International 2011)
- The commitment of large clients such as the UK government and professional organisations to put a stop to separation of activities and information exchange patterns across the client handover stage with a set of coordinated standards and requirements (BSI 2014; CIC 2012; British Institute of Facilities Management 2012; BSRIA 2009)
- The funding of the UK government of free-to-use information tools for the coordination of the project management aspects in the Digital Plan of Works; that will be bridging and integrating client requirement capture, with the responsibilities of project team members.

The review of trends in the field of data-driven automation allows us to speculate that many of the low-level priorities of the delivery of value from BIM, including “Value discovery”, “Life-cycle management”, “improved decision making”, “improved understanding” and “analytic benefits” will found a fruitful IT environment when they come to the forefront of priorities in a few years; however to prepare for this every player in the industry needs to understand the core principles of efficient IT data-driven automation which could be summarised with the development of highly specialised artificial intelligence expertise areas, fed by automated data collection processes, and integrated in orchestrated services overviewed by human agents.

Revised copies of the present paper will be available at the web address <http://www.overarching.it/wordpress/2014/10/08/iccc-2014/> along with the data sets used in the work and full bibliography of the bibliographic review.

#### 6 REFERENCES

- British Institute of Facilities Management. 2012. “BIM and FM: Bridging the Gap for Success”. BIFM, Bishop's Stortford, Hertfordshire, UK.
- BSI. 2014. “PAS 1192-3:2014: Specification for Information Management for the Operational Phase of Assets Using Building Information Modelling”. BSI Standards Limited.
- BSRIA. 2009. “GSL: The Soft Landings Framework”. BSRIA.
- buildingSMART International. 2011. “MVD Process — buildingSMART.” Accessed May 7. <http://buildingsmart.com/standards/idm/mvd/mvd-process>.
- CIC. 2012. “Building Information Modeling Planning Guide for Facility Owners.”
- Hendler, James. 2008. “Avoiding Another AI Winter.” *IEEE Intelligent Systems*.
- Liebich, T. 2010. “Unveiling IFC2x4-the next Generation of openBIM.” In *Proceedings of the CIB W78 2010: 27th International Conference*. Cairo, Egypt.

- RIBA. 2012. "RIBA Plan of Work 2013: Consultation Document."  
<http://www.architecture.com/Files/RIBAProfessionalServices/Practice/FrontlineLetters/RIBAPlanofWork2013ConsultationDocument.pdf>.
- RICS. 2011. *RICS 2011 Building Information Modelling Survey Report*. Building Cost Information Service. London: BICS. [http://www.scan2bim.info/files/rics\\_2011\\_BIM\\_Survey\\_Report.pdf](http://www.scan2bim.info/files/rics_2011_BIM_Survey_Report.pdf).
- Sinclair, Dale. 2012. "BIM Overlay to the RIBA Outline Plan of Work". RIBA.  
<http://www.ribabookshops.com/item/bim-overlay-to-the-riba-outline-plan-of-work/78038/>.
- Tversky, Amos, and Daniel Kahneman. 1999. "Judgement under Uncertainty: Heuristics and Biases." In *Judgment and Decision Making: An Interdisciplinary Reader*, 35–52. Cambridge: Cambridge University Press.
- UK BIM Task Group. 2012. "COBie UK 2012." <http://www.bimtaskgroup.org/cobie-uk-2012/>.
- US National Science Foundation. 2014a. *Active DIBB Program Grants*.  
<http://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=7726&BooleanElement=ANY&BooleanRef=ANY&ActiveAwards=true&#results>.
- . 2014b. "Laying the Groundwork for Data-Driven Science." Press Release 14-132.  
[http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=132880&org=NSF&from=news](http://www.nsf.gov/news/news_summ.jsp?cntn_id=132880&org=NSF&from=news).
- . 2014. "Data Infrastructure Building Blocks (DIBBs)." Accessed August 10.  
[http://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=504776](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504776).
- USACE/NASA, and Bill East. 2012. "Construction Operations Building Information Exchange (COBIE)." Accessed August 20.  
<http://www.buildingsmartalliance.org/index.php/projects/activeprojects/25>.
- Wolstenholme, Andrew. 2009. "Never Waste a Good Crisis". Constructing Excellence.  
[http://www.constructingexcellence.org.uk/pdf/Wolstenholme\\_Report\\_Oct\\_2009.pdf](http://www.constructingexcellence.org.uk/pdf/Wolstenholme_Report_Oct_2009.pdf).