

ACADEMIC ANALYTICS NA PRÁTICA: AGREGAR E ANALISAR MÚLTIPLAS FONTES DE DADOS

Sérgio André Ferreira*

António Manuel Valente de Andrade**

RESUMO: As instituições do ensino superior (IES) utilizam diversos subsistemas tecnológicos para gestão das várias dimensões da vida da organização, que recolhem um imenso volume de dados proveniente desses subsistemas, mas que não comunicam entre si. A ideia do *Analytics* na Educação alicerça-se no potencial da utilização agregada destes dados para o processo de tomada de decisão e no desenvolvimento de novos modelos na gestão. Partindo-se do caso da realidade da Universidade Católica Portuguesa – Porto, apresentam-se a arquitetura e os resultados alcançados no desenvolvimento de um protótipo de *Academic Analytics*, para a gestão da atividade de ensino na instituição, que agrega dados de três subsistemas tecnológicos. Deste trabalho resultam três contributos principais: *i*) Teórico – O *Analytics* na Educação é um tema novo e em rápida ascensão, sendo, por isso, difícil manter um discurso estável na definição de conceitos e descrição de processos. Neste quadro, apresenta-se uma reflexão sobre as principais questões discutidas e são colocados em relação os conceitos mais relevantes; *ii*) Metodológico – A fundamentação e aplicação do modelo *Design Science Research Process* (DSRP), que, apesar da sua versatilidade no campo da apresentação de artefactos, ainda tem pouca expressão no campo dos Sistemas da Informação (SI) aplicados às Ciências Sociais e Humanas; *iii*) Prático – Apresentação do protótipo de um sistema de *Academic Analytics*.

PALAVRAS-CHAVE: *Academic Analytics*, *design science*, Ensino Superior, gestão.

* CEDH – Centro de Estudos em Desenvolvimento Humano, Faculdade de Educação e Psicologia, Centro Regional do Porto da Universidade Católica Portuguesa (sergioandreferreira@gmail.com).

** CEGE – Centro de Estudos em Gestão e Economia, Faculdade de Economia e de Gestão, Centro Regional do Porto da Universidade Católica Portuguesa (aandrade@porto.ucp.pt).

1. INTRODUÇÃO

Ao longo do tempo existiram eventos que marcaram de forma profunda a história da humanidade, que revolucionam por completo as estruturas sociais, dando origem a novos modelos de sociedade ou mesmo civilizacionais. Alvin Toffler, no seu livro *The Third Wave*, que teve a sua primeira edição em 1980, identifica três grandes vagas com impactos à escala global: a primeira vaga, ocorrida há mais de dez mil anos, com a revolução neolítica e a invenção da agricultura; a segunda vaga, que aconteceu com a revolução industrial, em finais do século XVII; e a terceira vaga, que começou a desenhar-se após a II Guerra Mundial e que é fundamentalmente baseada na informação e na tecnologia. Toffler utilizou vários termos para definir esta terceira vaga: “Idade da Informação”, “Idade do Computador”, “Idade Digital” ou “Idade dos Novos *Media*” (Toffler, 1999).

Apesar da multiplicidade de termos existentes para designar a sociedade atual, a expressão “Sociedade de Informação” parece ser a que está mais instituída, mesmo no plano de organizações governamentais, como é o caso da Organização das Nações Unidas, que em Assembleia Geral aprovou, através da resolução 56/183, de 21 de dezembro de 2001 (ONU, 2001), a realização da “Cimeira das Nações Unidas para a Sociedade da Informação” (ITU, 2013). Na perspetiva de Castells e Himanen: “O fundamento desta sociedade é o informacionalismo, o que significa que as atividades decisivas da *praxis* humana se baseiam em tecnologias de informação, organizada (...) em redes de informação e centrada no processamento da informação (símbolos)” (2007: 1).

A Sociedade da Informação é resultado das alterações na produção e edição e da escala em que a informação pode ser difundida e recebida. Calcula-se que, no início do século XXI, apenas cerca de 25% do total da informação estavam arquivados em suporte digital, os restantes estariam preservados em papel, filmes, fotos e outros meios físicos. Atualmente, o meio digital representa já mais de 90% (Santos, 2013).

Esta explosão do volume de dados e dos fluxos de informação, suportados pela maior capacidade de armazenamento, aumento da memória dos computadores, processadores mais eficientes, *software* mais inteligente e algoritmos mais eficazes, é um fator que eleva para outra dimensão o armazenamento, o processamento e a análise de dados (Santos, 2013).

Caracterizado o contexto da sociedade global, fazendo um *zoom in* de escala e centrando a análise no plano das médias e pequenas organizações, em concreto as IES, interessa perceber como é feito o aproveitamento dos

dados digitais que estas instituições também têm disponíveis e com interesse para a sua atividade.

De facto, também nas IES existiu uma forte migração da informação do suporte físico para o formato digital e estas organizações têm hoje vários subsistemas tecnológicos, responsáveis, por exemplo, pela gestão administrativa, da qualidade e da atividade formativa, que disponibilizam grandes volumes de dados. A ideia do *Analytics* na Educação é simples e tem associado um potencial transformativo muito elevado: a mobilização e a leitura cruzada dos dados de diversas fontes começa a ser possível do ponto de vista tecnológico, criando condições para um processo de tomada de decisão mais informada, abrindo as portas a novos modelos na gestão das IES nos campos da eficiência organizacional e pedagógico.

O termo *Analytics* é um neologismo em Educação e é importado de outras áreas, principalmente do campo da Gestão:

“Analytics’ is a term used in business and science to refer to computational support for capturing digital data to help inform decision-making. With the growth of huge data sets and computational power, this extends to designing infrastructures that exploit rapid feedback, to inform more timely interventions, whose impact can in turn be monitored. Organisations have increasingly sensitive ‘digital nervous system’ providing real time feedback on the external environment and the effects of actions” (Shum, 2012: 1).

Neste trabalho, partindo da realidade da Universidade Católica – Porto e utilizando os dados dos subsistemas tecnológicos em funcionamento nesta instituição com interesse para gestão da atividade de ensino, apresentam-se os resultados alcançados na construção de um protótipo de um *Academic Analytics*. Em concreto, é arquitetado um modelo que operacionaliza o cruzamento de dados de três subsistemas tecnológicos: *i*) Ensino (*Campus online*); *ii*) Gestão Administrativa (*Sophia*); *iii*) Gestão da Qualidade (*SIGIQ*). Nesta linha, este protótipo tem dois objetivos principais: *i*) demonstra, à administração e aos serviços de gestão de qualidade da IES, o potencial do uso integrado dos dados provenientes dos vários subsistemas tecnológicos para a gestão da atividade de ensino; *ii*) definir as características que o futuro artefacto tecnológico deve reunir para que equipas com *know how* no campo do desenvolvimento de *software* e sistemas as possam concretizar.

A base metodológica seguida para o desenvolvimento deste protótipo de *Academic Analytics* foi o modelo *Design Science Research Process* (DSRP),

proposto por Pefferes *et al.* (2006), que prevê uma sequência de seis etapas para condução e controlo de todo o processo, desde a identificação do problema/motivação para a realização do protótipo até à avaliação e comunicação de resultados.

Considerando esta secção introdutória, este artigo está estruturado em seis secções. A secção dois centra-se na temática do *Analytics* na Educação e divide-se em três subsecções: a primeira tem um cariz marcadamente conceptual e explora os neologismos próprios de uma área ainda emergente com o objetivo de consolidar um discurso ainda não estabilizado; nas outras duas subsecções são equacionados os usos, as potencialidades e as dificuldades do *Analytics* na Educação. Na secção três é realizado e justificado o enquadramento metodológico da investigação à luz dos objetivos do trabalho. Na secção quatro é apresentado o protótipo construído: uma base de dados que lê e agrega registos de várias fontes e permite a realização de relatórios e análises. Na secção cinco são apresentadas as principais conclusões do trabalho e na secção seis são discutidos os trabalhos futuros.

2. ANALYTICS NA EDUCAÇÃO

2.1. A emergência do tema e a necessidade de clarificação do quadro conceptual

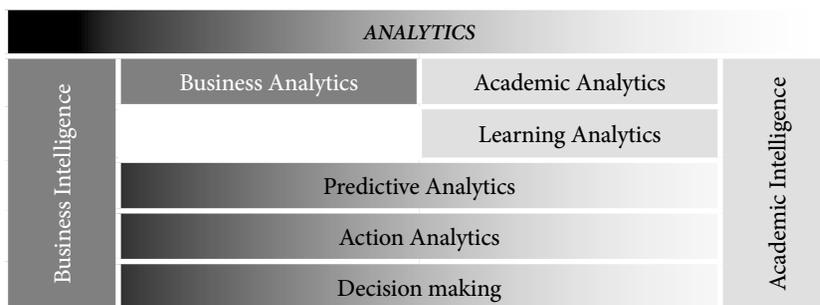
Há já várias décadas que o mundo empresarial implementa sistemas tecnológicos para análise avançada do enorme volume de dados de que dispõe. Estes sistemas são cada vez mais indispensáveis na informação durante o processo de tomada de decisão, requisito fundamental para uma gestão eficaz. O interesse das IES para esta nova vaga na forma de utilização dos dados só agora começa a despertar e estas tentam implementar, ainda timidamente, os seus sistemas de *Analytics*: “More recently, institutions of higher education are starting to adapt these methods to target fund raising, inform enrollment decisions, target marketing efforts, improve student support processes, and to better understand retention/persistence patterns” (Bach, 2010: 1).

De facto, o *Analytics* na Educação é uma temática emergente, prenunciando-se a sua adoção pelas IES num futuro próximo – segundo o “Horizon Reports”, num período de dois ou três anos (Johnson *et al.*, 2013). O interesse manifestado por organizações nacionais e internacionais na área do ensino superior, a organização de congressos para discussão específica

deste assunto – *e.g.*, três edições da Conferência “Learning Analytics and Knowledge”, em 2011, 2012 e 2013 (SoLAR, 2013) – e o aumento do número de publicações científicas são indicadores da emergência do tema e da sua relevância. Como em qualquer nova área de pesquisa, o *Analytics* na Educação tem importado de outras áreas do saber uma variedade de termos e expressões para descrever conceitos e processos, muitas vezes com diferenças significativas no plano conceptual e funcional, ocorrendo também a utilização de termos diferentes para as mesmas definições conceptuais e funções.

Neste contexto, por forma a garantir o rigor da comunicação na temática, impõe-se a clarificação do quadro conceptual. Na Figura 1 está representado esquematicamente o trabalho de sistematização em torno dos conceitos usados no *Analytics* em Educação, apresentado num *White paper*, da Educause, publicado em 2012 (Barneveld, Arnold, & Campbell, 2012), onde se estabelecem paralelismos entre o quadro conceptual do campo da Gestão e da Educação. Utilizando uma árvore como imagem, pode-se considerar o conceito de *Analytics* como o tronco comum e imaginar dois grandes ramos – o *Business Intelligence* e o *Academic Intelligence* – que apresentam uma simetria imperfeita na forma como os sub-ramos se organizam.

Figura 1. *Analytics*: paralelismos dos quadros conceptuais na Gestão e na Educação



No Quadro 1, sintetizam-se as principais conclusões do *White paper*, da Educause (Barneveld *et al.*, 2012), na desambiguação de conceitos e no estabelecimento de uma linguagem comum para o uso da *Analytics* na Educação. Este quadro está organizado em duas colunas: na primeira é identificado o termo em consideração; na segunda, partindo da leitura da informação presente no trabalho em análise (Barneveld *et al.*, 2012), apresentam-se a variação do significado do conceito na literatura e o quadro conceptual proposto para a uniformização da interpretação do conceito.

Quadro 1. Estabelecimento de um quadro conceptual comum para o uso do *Analytics* na Educação (adaptado de Barneveld *et al.*, 2012)

CONCEITO	DESAMBIGUAÇÃO
<i>Analytics</i>	<p>Varição do significado do conceito: Autores concordantes relativamente ao seu significado.</p> <p>“[The] <i>processes of data assessment and analysis that enable us to measure, improve, and compare the performance of individuals, programs, departments, institutions or enterprises, groups of organizations, and/or entire industries.</i>” (Norris, Baer, & Offerman, 2009: 1)</p> <p>Quadro conceptual proposto: Conceito abrangente que tem associada a ideia da tomada de decisão informada com base em dados. Tanto no campo da Gestão como no da Educação, a definição do conceito é consistente.</p>
<i>Academic Analytics</i>	<p>Varição do significado do conceito: A expressão <i>Academic Analytics</i> em Educação é um conceito polissémico. Na literatura este conceito adquire diferente abrangência no plano das matérias e das escalas da análise. No seu significado mais amplo, o <i>Academic Analytics</i> é o termo equivalente ao <i>Business Analytics</i>, que se baseia no uso dos dados para suportar a gestão da atividade da IES, nas suas diversas vertentes (<i>e.g.</i>, financeira, ensino, marketing). No sentido mais estrito, centra-se no acompanhamento de questões relacionadas com o sucesso académico do estudante individual (<i>e.g.</i>, estudantes em risco de retenção).</p> <p>“[The] <i>imperfect equivalent term for Business Intelligence, which [essentially describes the use] (...) of information to support decision making in the financial and business of the academy.</i>” (Goldstein, 2005)</p> <p>“<i>Mining data from systems that support teaching and learning to provide customization, tutoring, or intervention within the learning environment.</i>” (J. P. Campbell, 2007: 2)</p> <p>“[Focused on academic issues, primarily] <i>student access, affordability, and success.</i>” (Norris <i>et al.</i>, 2009: 2)</p>

CONCEITO	DESAMBIGUAÇÃO
<p><i>Academic Analytics</i> (cont.)</p>	<p>Quadro conceptual proposto: <i>Academic Analytics</i> como conceito paralelo ao <i>Business Analytics</i>. Nesta linha, o <i>Analytics</i> permite à administração da IES aceder a indicadores históricos ou em tempo real sobre vários aspetos da instituição e das suas unidades orgânicas (faculdade/escolas e departamentos).</p> <p><i>“A process for providing higher education institutions with the data necessary to support operational and financial decision making.”</i> (Adaptado de Goldstein, 2005)</p>
<p><i>Learning Analytics</i></p>	<p>Variação do significado do conceito: Centra-se nas questões da atividade de ensino e aprendizagem. A escala de análise varia entre o plano institucional (e.g., avaliação de currículo e instituições) e o acompanhamento dos estudantes individuais, confundindo-se aqui com o <i>Academic Analytics</i> no sentido mais estrito.</p> <p><i>“[The] interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues.”</i> (Johnson, Smith, Willis, Levine, & Haywood, 2011: 28)</p> <p><i>“It might be used as well to assess curricula, programs, and institutions.”</i> (Johnson et al., 2011: 28)</p> <p><i>“[The] collection and analysis of usage data associated with student learning; [to] observe and understand learning behaviors in order to enable appropriate intervention.”</i> (Brown, 2011: 1)</p> <p>Quadro conceptual proposto: O LA é o ponto de divergência no estabelecimento do paralelismo conceptual entre os mundos da Gestão/Negócio e o da Educação. O LA pode ser encarado sob duas perspetivas: uma centrada em métricas que fornecem informação aos departamentos sobre aspetos relacionados com a organização e funcionamento dos cursos; outra voltada especificamente para os estudantes e suas atividades de aprendizagem.</p> <p><i>“The use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals.”</i> (Adaptado de Bach, 2010)</p>

CONCEITO	DESAMBIGUAÇÃO
<p><i>Predictive Analytics</i></p>	<p>Varição do significado do conceito: Autores concordantes relativamente ao significado do conceito.</p> <p>“[A] set of [business intelligence] technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events (...) predictive analytics is forward-looking, using past events to anticipate the future.” (Eckerson, 2007: 5)</p> <p>Quadro conceptual proposto: Depois de recolhidos os dados, elaborados os relatórios e analisada a informação, o passo seguinte é utilizar esses dados para antecipar cenários (<i>Predictive Analytics</i>) e tomar decisões com base na informação (<i>Action Analytics</i>). Esta análise preditiva serve todos os níveis do ensino superior.</p>
<p><i>Action Analytics</i></p>	<p>Varição do significado do conceito: Autores concordantes relativamente ao significado do conceito.</p> <p>“action analytics utility proposal have the potential to make sense of information from across systems and shape future policies, practices, interventions, and actions.” (Norris, Baer, Leonard, Pugliese, & Lefrere, 2008: 64)</p> <p>Quadro conceptual proposto: Ações, intervenções, definição de políticas e práticas suportadas nos dados recolhidos pelos sistemas de <i>Analytics</i>.</p>

2.2. Usos e potencialidades do *Analytics* na Educação

Na perspetiva de Filipe Duarte Santos (2013), a análise de grandes volumes de dados em sistemas de *Analytics* tem associadas profundas alterações no modo como é usada a informação, que são particularmente significativas a três níveis:

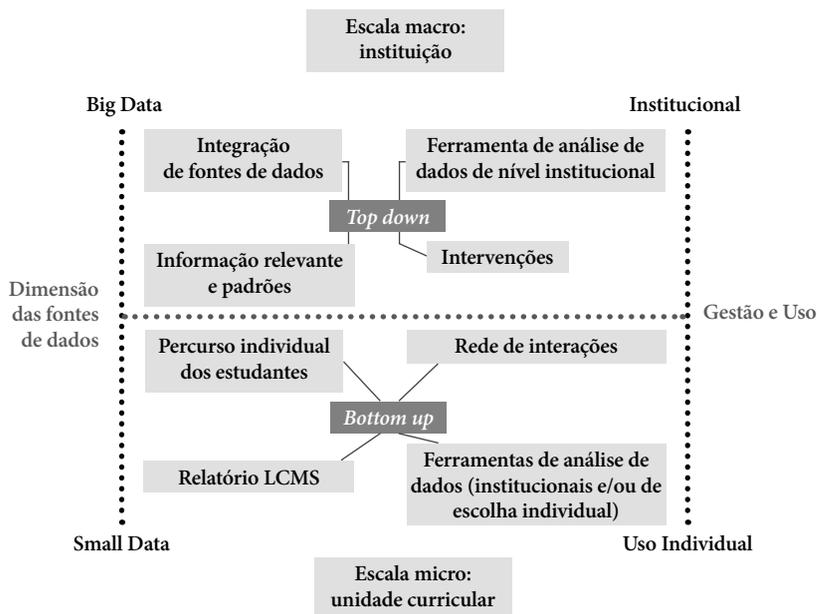
- i) *Uso da totalidade dos dados dos sistemas em detrimento da amostragem* – A evolução tecnológica tem permitido aumentar a capacidade de armazenamento, processamento e análise de dados. Assim, na realização de estudos estatísticos complexos sobre vários elementos do sistema, a possibilidade de trabalhar com a totalidade dos dados tem implicações no plano metodológico, tornando-se dispensável construir amostras representativas do sistema.
- ii) *Maior volume de dados permite lidar com a incerteza e menor qualidade de alguns dados* – Tendencialmente haverá condições mais favoráveis

para lidar com a incerteza associada ao facto de alguns dados não terem a qualidade necessária. O enorme volume de dados atenuará os efeitos da menor fidedignidade que possa existir em alguns deles.

- iii) *Conhecimento mais profundo do sistema e descoberta de correlações significativas* – A multiplicidade de correlações que podem ser feitas entre os vários elementos que constituem o sistema é importante para o conhecimento do seu funcionamento e o ponto de partida na procura de respostas sobre o significado dessas correlações e na definição de ações de intervenção.

O potencial do imenso volume de dados dos subsistemas tecnológicos para alimentar os subsistemas de *Analytics* das IES tem interesse e aplicabilidade a vários níveis e dimensões. As IES estão organizadas segundo uma estrutura hierarquizada; uma organização tipo poderá ser: Universidade/Faculdade/Departamento/Curso/UC. Estas estruturas e os vários atores das organizações têm necessidade de sistemas de *Analytics* de escalas e granularidades diferentes. A Figura 2, adaptada do trabalho de George Siemens (Siemens, 2012), sintetiza o uso do *Analytics* às escalas macro e micro, evidenciando a relação de complementaridade entre elas.

Figura 2. Escalas de uso do *Analytics* nas IES



A parte superior da Figura 2 reflete o papel do *Analytics* a uma escala macro – a IES como um todo ou suas unidades orgânicas –, estando associada ao conceito de *Academic Analytics* (Quadro 1). A esta escala, o *Analytics* reflete uma visão tipo *top down*, no sentido em que a análise é feita a uma escala global através da integração de fontes de dados dos vários subsistemas tecnológicos da IES numa ferramenta de análise, que deve permitir o acesso a indicadores, históricos ou em tempo real, sobre as várias dimensões da IES e suas unidades orgânicas.

As fontes de dados utilizadas para alimentar o *Academic Analytics* podem limitar-se a aspetos mais relacionados com a vertente pedagógica, neste caso, o *Academic Analytics* está mais próximo do conceito de LA: “Learning analytics is more specific than academic analytics: the focus of the former is exclusively on the learning process. Academic analytics reflects the role of data analysis at an institutional level, whereas learning analytics centers on the learning process (which includes analyzing the relationship between learner, content, institution, and educator)” (Long & Siemens, 2011; cf. Quadro 1); ou integrar dados de diversos tipos de fontes (*e.g.*, dados da atividade pedagógica, mas também fontes administrativas, financeiras entre outras, permitindo à IES gerir as várias dimensões subjacentes ao seu funcionamento), assumindo-se como um conceito paralelo ao *Business Analytics*: “Data is the foundation of all analytics efforts. Academic analytics can be based on data from multiple sources (such as an SIS, a CMS, or financial systems) and in multiple formats (such as spreadsheets, enterprise financial system reports, or paper records)” (John Campbell & Oblinger, 2007).

No plano institucional, estes sistemas são utilizados por administradores, serviços de controlo de qualidade, marketing e financiadores da IES para suporte à tomada de decisão ou guia para ação (J. Campbell, DeBlois, & Oblinger, 2007; John Campbell & Oblinger, 2007; Long & Siemens, 2011), sendo possível identificar cinco etapas no seu uso: captura de dados > produção de relatórios com base nos dados recolhidos > identificação de informação relevante/padrões/previsões > intervenção > redefinição de ações (John Campbell & Oblinger, 2007).

A parte inferior da Figura 2 centra-se numa escala micro e a abordagem é do tipo *bottom up*, em que os elementos mais pequenos do sistema são especificados com grande detalhe. Os dados (*small data*) são utilizados por um utilizador ou pequeno grupo de utilizadores (*e.g.*, coordenador de curso, professores que lecionam determinada UC), que procuram informação muito filtrada à escala do curso, UC ou aluno.

2.3. Desafios do *Analytics* na Educação

O uso dos LA pelas IES levará alguns anos a amadurecer, contudo, a sua presença já se faz sentir e não deve ser ignorada (Picciano, 2012). Segundo o relatório “The State of Learning Analytics in 2012: A Review and Future Challenges” (Ferguson, 2012), o desenvolvimento e uso dos sistemas de *Analytics* pelas IES enfrenta dois desafios principais: o desafio tecnológico e o desafio educacional.

O desafio tecnológico relaciona-se com a capacidade de as IES desenvolverem aplicações informáticas que permitam aprofundar modos de extração, agregação e visualização de dados e de elaboração de relatórios, constituindo-os em *Analytics*. Este desafio é potenciado quando as fontes de dados estão dispersas por múltiplas plataformas.

O desafio educacional refere-se à capacidade de as IES conseguirem otimizar o uso da informação na gestão dos ambientes de aprendizagem ricos em tecnologia e potenciarem a eficácia do processo de ensino e aprendizagem (Ferguson, 2012: 4).

Relacionadas com estes dois desafios, estão as seis barreiras que as instituições têm de ter em atenção no desenvolvimento dos seus sistemas de *Analytics*, que aparecem referidas no Relatório “Predictive Analytics – Extending the Value of Your Data Warehousing Investment”, de The Warehousing Institute (Eckerson, 2007):

- i) *Complexidade* – Desenvolver modelos de *Analytics* em Educação é um processo lento e complexo e de trabalho intensivo que exige o envolvimento de diferentes pessoas do quadro da universidade e equipas multidisciplinares.
- ii) *Dados* – A maior parte das IES tem bases de dados repletas de erros e inconsistências, que têm de ser limpas e normalizadas para que possam ser usadas.
- iii) *Processamento* – Análises complexas de dados exigem processadores eficientes e *softwares* inteligentes, requisitos que ainda não são cumpridos em muitas IES.
- iv) *Especialistas* – Há dificuldade em contratar especialistas informáticos no desenvolvimento de sistemas de *Analytics*.
- v) *Interoperabilidade* – Os sistemas de *Analytics* envolvem o acesso e integração de fontes de dados de múltiplas plataformas e aplicações que requerem *software* interoperável.
- vi) *Preço* – O desenvolvimento de sistemas de *Analytics* pode ter custos importantes para as organizações de média dimensão, como são a maior parte das IES.

Na sequência destas barreiras, no mesmo relatório pode ler-se: “Fortunately, these barriers are beginning to fall, thanks to advances in software, computing, and database technology.”

3. OBJETIVOS E OPÇÕES METODOLÓGICAS

A área dos Sistemas de Informação (SI) é fundamentalmente uma área aplicada que tem como objetivo otimizar a tecnologia enquanto veículo de informação no seio das organizações (Varajão, 1998). Nesta linha, os SI são aplicados em inúmeras outras disciplinas, nomeadamente no campo das Ciências Sociais e Humanas, para resolver problemas que intersejam as Tecnologias da Informação e as organizações (Hevner, March, Park, & Ram, 2004; McKay, Marshall, & Hirschheim, 2012; Peffers *et al.*, 2006; Varajão, 1998).

Um dos procedimentos mais importantes e frequentes nos SI consiste em arquitetar e desenvolver uma solução aplicável à resolução de um problema concreto. Enquanto desenho metodológico de pesquisa, este paradigma está instituído em disciplinas como a engenharia, mas muito pouco presente em trabalhos de investigação publicados em outras áreas, nomeadamente as Ciências Sociais e Humanas (Peffers *et al.*, 2006), onde os paradigmas dominantes continuam a ser os tradicionais estudos do tipo descritivo, explicativo e interpretativo (Peffers *et al.*, 2006; Sampiere, Collado, & Lucio, 2006). Neste contexto, o ato de desenvolvimento de novos sistemas tecnológicos não é visto como metodologia científica, como ilustra a visão de um editor-chefe de uma das mais prestigiadas revistas científicas na área dos SI: “the editor-in-chief of one of the highest ranked IS research journals, he was told that the journal didn’t entertain papers about new systems development methods, because they involved neither theory development nor theory testing” (Peffers *et al.*, 2006: 84).

A autoria do termo *design science* é atribuída a Fuller (1965), no já longínquo ano de 1963, que o define como uma forma sistemática de desenho (Wikipédia, 2013). Portanto, tanto na designação, como na sua definição, é evidente a tentativa de colocar o *design science* dentro dos limites da metodologia científica. Decorrido mais de meio século da cunhagem do termo, a aceitação do *design science* enquanto paradigma de pesquisa continua a ter grandes dificuldades de afirmação em áreas de conhecimento fora da engenharia. A reduzida publicação de artigos com arbitragem científica com esta abordagem metodológica é um indicador deste facto: “Engineering disciplines accept design as a valid and valuable research

methodology, but for the most part, major IS journals still seem to find it a questionable model for quality research” (Peppers *et al.*, 2006: 84).

A problemática da afirmação do *design science* tem sido amplamente discutida na literatura nos últimos anos, tendo sido publicados vários trabalhos que refletem sobre as causas da dificuldade de afirmação e se traduzem em esforços de conceptualização valiosos (Hevner *et al.*, 2004; McKay *et al.*, 2012; Peppers *et al.*, 2006). De facto, estes trabalhos estão alinhados entre si nas principais conclusões, ao apontarem a falta de um modelo conceptual que funcione como linhas-guia para investigadores e como modelo mental para os leitores e revisores compreenderem e avaliarem as investigações em *design science*. Na perspetiva de Hevner *et al.*, “Our purpose for establishing these seven guidelines is to assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research” (Hevner *et al.*, 2004: 82) Segundo Peppers *et al.*, “What’s missing may be a conceptual model for how researchers can carry out *design science* research in IS and a mental model or template for readers and reviewers to recognize and evaluate it” (Peppers *et al.*, 2006: 85).

O trabalho de Hevner *et al.* (2004) é um marco no plano da conceptualização e definição de linhas-guia para compreender, executar e avaliar o *design science* aplicado aos SI, sendo referenciado em muitos trabalhos realizados posteriormente sobre a temática (e.g. McKay *et al.*, 2012; Peppers *et al.*, 2006; Wikipédia, 2013). Neste artigo Hevner *et al.* (2004) estabelecem as bases do *design science*, concretizadas através de uma proposta de sete linhas-guia para aplicação deste paradigma, que estão sintetizadas no Quadro 2. Estas linhas-guia foram definidas com base numa revisão bibliográfica exaustiva sobre a temática do *design science*, sendo, também, demonstrada a sua aplicabilidade a três exemplos concretos de artefactos na área dos SI presentes na literatura (Hevner *et al.*, 2004).

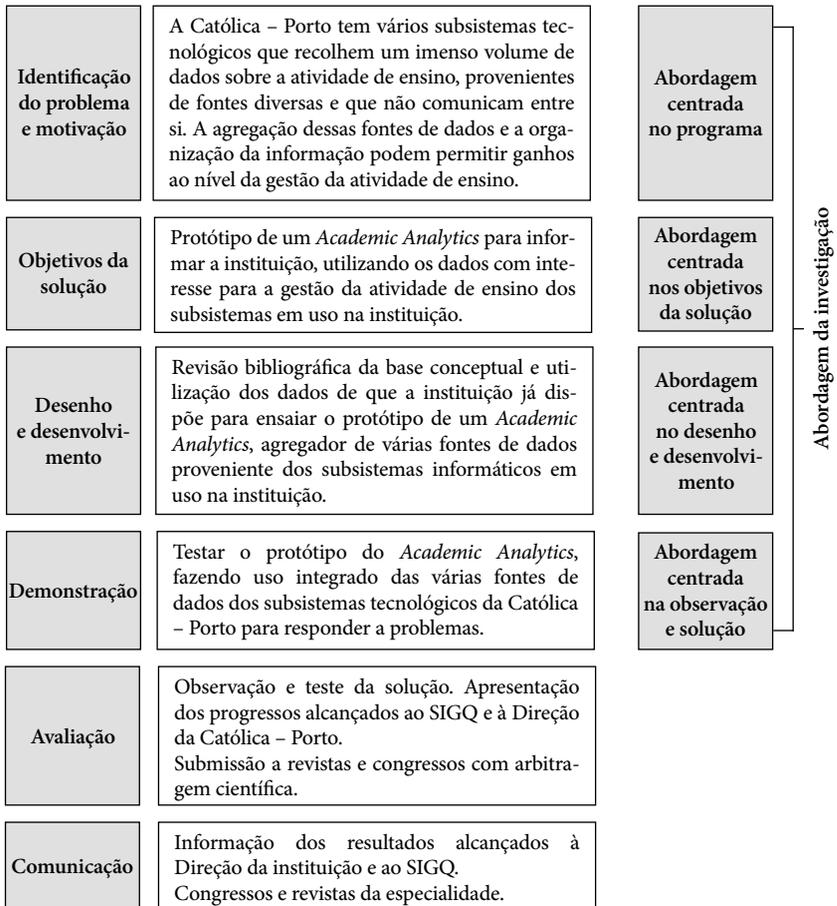
O modelo *Design Science Research Process* (DSRP), proposto por Peppers *et al.* (2006), resulta de um esforço de sistematização do conhecimento teórico e prático na área do *design science*. O modelo proposto preenche três requisitos: *i*) é consistente com o conhecimento teórico e prático produzido no âmbito do paradigma do *design science*; *ii*) apresenta um esquema com várias fases para condução do processo *design science*; *iii*) define um esquema mental com as características do *output* da pesquisa. A Figura 3 representa de forma esquemática o modelo DSRP, onde podem ser identificadas seis etapas na condução da investigação, que se articulam de forma bastante próxima com as sete linhas-guias da pesquisa *design science* de Hevner *et al.* (2004).

Quadro 2. Linhas-guias para a pesquisa *design science* (Hevner *et al.*, 2004: 83)

LINHA-GUIA	DESCRIÇÃO
Linha-guia 1: O <i>design</i> como um artefacto	A pesquisa <i>design science</i> deve produzir um artefacto viável na forma de construção, modelo, método ou representação de uma ideia.
Linha-guia 2: Relevância do problema	O objetivo do <i>design science</i> é desenvolver soluções para dar respostas a problemas relevantes na área da gestão.
Linha-guia 3: Avaliação do <i>design</i>	A utilidade, a qualidade, a eficiência do <i>design</i> do artefacto têm de ser rigorosamente demonstradas através de métodos de avaliação rigorosos.
Linha-guia 4: Contribuições da pesquisa	A pesquisa efetiva em <i>design science</i> deve dar contribuições claras e verificáveis no plano dos SI (capacidade de resolução de problemas até aí sem solução), da construção do conhecimento (desenvolvimento criativo de modelos, métodos, ideias, concorrendo para o alargamento da área do conhecimento onde se insere) e da inovação metodológica (<i>e.g.</i> , experimentação, análise, observação, teste e descrição).
Linha-guia 5: Rigor na pesquisa	O <i>design science</i> deve aplicar métodos rigorosos na construção e avaliação do artefacto.
Linha-guia 6: O <i>design</i> como a procura de um processo	O <i>design</i> de um artefacto deve adequar os meios disponíveis aos fins desejados, tendo em observância o contexto da intervenção (<i>e.g.</i> , elementos tecnológicos e organizacionais).
Linha-guia 7: Comunicação da pesquisa	Os resultados devem ser comunicados a audiências com conhecimento mais tecnológicos e a públicos com perfis mais orientados para a gestão.

Partindo deste estado da arte no paradigma do *design science*, adotou-se o modelo DSRP (Peppers *et al.*, 2006) para guiar a nossa pesquisa. Considerou-se que os requisitos do modelo são sólidos do ponto de vista teórico e facilmente operacionalizáveis como guia para ação da vertente prática, permitindo o controlo do processo e do *output*. Na Figura 3 estão esquematizadas as seis atividades de forma sequencial do processo DSRP aplicado ao *Academic Analytics* da Católica – Porto.

Figura 3. Processo de *Design Science* para o *Academic Analytics* da Católica – Porto



- Identificação do problema e motivação* – O *Academic Analytics* é uma temática emergente na literatura e de reconhecido potencial para a gestão dos diversos aspetos da vida das IES. A Católica – Porto tem vários subsistemas tecnológicos que armazenam dados de interesse para a gestão, nomeadamente no campo da atividade de ensino. Na situação atual, esses subsistemas tecnológicos não comunicam entre si, não sendo possível cruzar os dados das diferentes fontes. A abordagem é centrada neste problema e tem como motivação agregar dados de diversas fontes e cruzá-los entre si, de modo a extrair informação útil que se traduza em ganhos na gestão da atividade de ensino.

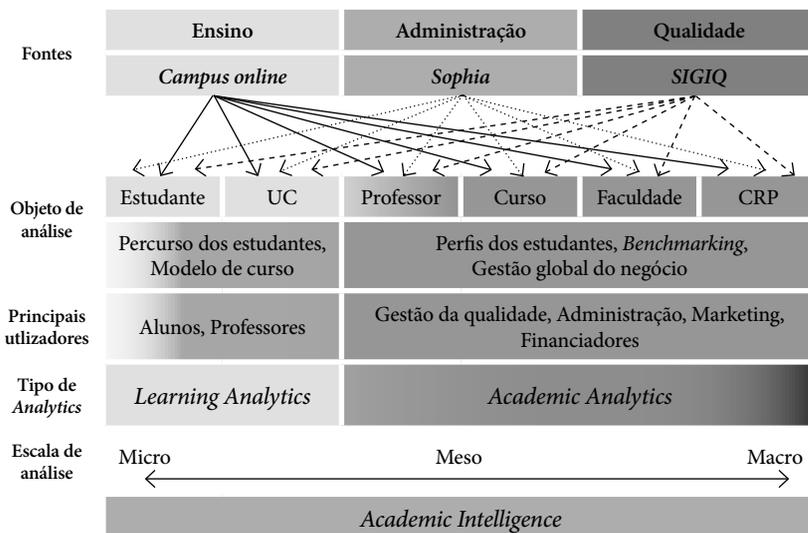
- *Objetivos da solução* – Utilizando os dados que a instituição já dispõe, é objetivo apresentar uma arquitetura de um sistema de *Academic Analytics* e desenvolver um protótipo que demonstre o modo de operacionalização e a sua utilidade para a gestão da atividade de ensino.
- *Desenho e desenvolvimento* – Mobilização da revisão da literatura realizada na área do *Analytics* para definição da base conceptual da arquitetura do sistema do *Academic Analytics*. Concretização de um protótipo de um sistema *Academic Analytics* que agrega três tipos de fontes de dados com origem em subsistemas tecnológicos em uso na instituição – *i*) Ensino (*Campus online*), *ii*) Administração (*Sophia*); *iii*) Qualidade (*SIGIQ*).
- *Demonstração* – A demonstração da eficácia da solução realizou-se através da experimentação e de simulações de casos, centrando-se na verificação da eficácia no cruzamento entre as três fontes de dados e na relevância e interesse para a gestão do *output* de informação.
- *Avaliação* – Observação e realização de testes sucessivos; apresentação dos progressos alcançados ao SIGQ e à direção da Católica – Porto, com o objetivo de ajustar o modelo às necessidades da informação da instituição; submissão do trabalho realizado a revistas e congressos com arbitragem para validação pela comunidade científica.
- *Comunicação* – Informação dos resultados alcançados ao SIGQ e à Direção da Católica – Porto e submissão a revistas e congressos com arbitragem.

4. RESULTADOS

A Figura 4 reflete a arquitetura do trabalho em curso no campo do *Analytics* na Católica – Porto. Nesta fase, circunscreveu-se a análise a três subsistemas tecnológicos que armazenam dados com interesse para a gestão da atividade de ensino: *Campus* (LCMS), Gestão Administrativa (*Sophia*) e Gestão da Qualidade (*SIGIQ*).

Na linha do enquadramento teórico apresentado na secção 2, o *Analytics* em Educação pode ter várias escalas de análise, que variam entre o plano micro (centrado em aspetos relacionados com o processo de ensino e aprendizagem como, por exemplo, o percurso dos alunos ou o modelo de curso. Esta informação tem potencialmente mais interesse para os professores e alunos mais diretamente envolvidos) e o plano macro (que contempla aspetos mais gerais como, por exemplo, questões administrativas ou gestão da qualidade). A escala micro está mais associada ao LA e a escala macro ao campo de ação do *Academic Analytics*, se bem que a fronteira entre os dois não seja fácil de delimitar (cf. Quadro 1).

Figura 4. Arquitetura do *Academic Analytics* com fontes de dados da atividade de ensino



O presente trabalho está alinhado com a perspetiva do *Academic Analytics* (análise macro) e nele é ensaiada a organização e agregação de dados existentes de diferentes subsistemas tecnológicos em uso na instituição para informar, em primeiro lugar, a administração e os serviços de gestão da qualidade da universidade sobre aspetos relevantes da atividade de ensino. Na Figura 5, sob a forma de esquema de uma base de dados, estão representados os resultados alcançados.

O desenvolvimento desta base de dados, que lê e agrega registos de várias fontes e que permite a realização de relatórios e análises, pressupõe a realização de um conjunto de passos que englobam a agregação de fontes, a limpeza e normalização de dados. Na Figura 5 são agregados três subsistemas tecnológicos:

- *C (Campus)* – A fonte de dados é o LCMS. Os dados disponibilizados são o *output* de um sistema de LA desenvolvido no *backoffice* do LCMS, numa fase anterior deste processo. Este sistema de LA permite fazer o posicionamento de cada UC numa matriz com cinco níveis de integração desta plataforma tecnológica no processo de ensino e aprendizagem – introdução, adoção, adaptação, imersão, transformação (Ferreira & Andrade, 2012a, 2012b).

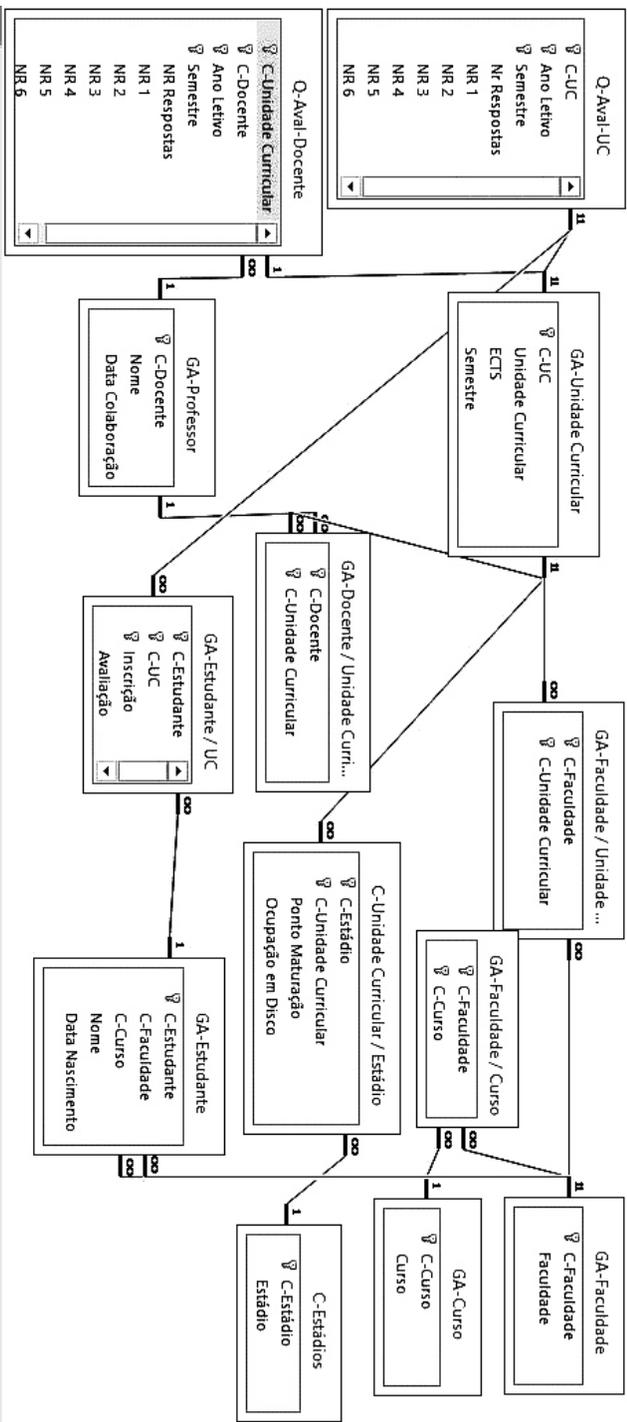


Figura 5. Esquema da base de dados com a agregação de fontes de subsistemas tecnológicos com interesse para a gestão da atividade de ensino

- *GA (Gestão Administrativa)* – A fonte de dados é o *Sophia*. Os dados disponibilizados possibilitam associar dimensões de cariz administrativo (faculdade, UC, professores, alunos, número de inscritos, ETCS...).
- *Q (Gestão da Qualidade)* – A fonte de dados é o *SIGIQ*. Os dados facultados por este subsistema resultam de questionários aplicados aos alunos no final de cada UC, onde se pode extrair informação sobre a avaliação global da UC e do docente.

No Quadro 3 é operacionalizada a agregação de dados dos três subsistemas tecnológicos: GA – foram extraídos dados com a designação da UC, o número de alunos e a média das classificações académicas dos estudantes inscritos; C – forneceu dados com o estágio e o ponto de maturação de cada UC relativamente ao uso e integração do LCMS no processo de ensino e aprendizagem (Ferreira & Andrade, 2012a, 2012b); Q – facultou dados sobre a apreciação global da UC feita pelos alunos no final das atividades através do preenchimento de um questionário (as seis colunas que aparecem no quadro referem-se ao seis níveis de posicionamento, que variam entre “Quest_1” – apreciação menos favorável – e “Quest_6” – apreciação mais favorável.

Quadro 3. Relatório com análise de dados com origem nos três subsistemas tecnológicos

UC	C Estágio	C Ponto de maturação (%)	Q Quest 1	Q Quest 2	Q Quest 3	Q Quest 4	Q Quest 5	Q Quest 6	GA N.º de alunos	GA Média
A	Introdução	15	0	1	3	3	1	1	10	14
B	Introdução	3	0	0	0	0	0	2	5	17
C	Introdução	3	0	0	0	0	1	14	20	14
D	Introdução	10	0	3	5	9	6	2	31	10

Na Figura 6 é ilustrada a potencialidade que este sistema de *Academic Analytics* pode ter numa fase de plena implementação no plano da: *i*) integração de dados com origem em diferentes fontes (subsistemas tecnológicos); *ii*) organização e análise multidimensional dos dados por níveis de detalhe (universidade, faculdade, UC, professor, aluno), com possibilidade de seleção e relação das dimensões que se querem efetivamente considerar num dado momento. Esta apresentação inteligente e versátil dos dados, inspirada na filosofia OLAP – *Online Analytical Processing* (OLAP. COM, 2010), tem um enorme potencial na apresentação e leitura dos resultados, pois diminui a entropia associada a elevados volumes de informação e garante o acesso condicionado a essa mesma informação, permitindo a elaboração de relatórios personalizados.

Os principais contributos práticos deste protótipo:

- *Ilustração da relevância, potencial e desafios da implementação dos sistemas de Academic Analytics para a gestão das IES.* O desenvolvimento de modelos de *Academic Analytics* são processos complexos que pressupõem a idealização do *output* de informação que dê resposta às necessidades da IES e sirva de suporte a uma ação informada, devendo estar alinhado com as condições tecnológicas e com a capacidade dos recursos humanos da IES. Este trabalho enquadra em termos teóricos os sistemas de *Analytics* na Educação, e o protótipo coloca em equação os desafios organizacionais e tecnológicos no seu desenvolvimento.
- *Compreensão dos dados: coleta, descrição, exploração e verificação da qualidade dos dados.* O protótipo revelou o tipo de dados que a instituição já possui nos vários subsistemas tecnológicos; esta informação é importante no desenvolvimento do futuro modelo de *Academic Analytics* e na adaptação do *output* de informação dos vários subsistemas às necessidades de informação da IES.
- *Preparação dos dados.* Os dados dos três subsistemas apresentaram erros, redundâncias e diferentes códigos para designar o mesmo objeto de análise (*e.g.*, o código das UC é diferente nos subsistemas *Sophia* ou *Blackboard*), o que exigiu um grande esforço de normalização de dados. No protótipo já foi dada resposta ao problema de agregação de diferentes fontes de dados relativos a ocorrências no mesmo período temporal. Os dados foram devidamente normalizados e identificados, através de uma codificação apropriada. Foi com base neste trabalho que se partiu para a análise realizada no passo seguinte.

- *Integração das fontes de dados de diferentes subsistemas tecnológicos que não comunicavam: Campus, Sophia e SIGIQ.*
- *Demonstração do potencial do uso integrado dos dados já disponíveis nos subsistemas tecnológicos para informar a universidade.*

5. CONCLUSÕES

A terceira vaga de Toffler, que se manifesta na atual Sociedade da Informação, fundamenta-se na importância da tecnologia e do processamento da informação como elemento decisivo nas várias atividades humanas. Em vários setores de atividade (e.g., Marketing e Gestão) já há várias décadas que estão implementados sistemas tecnológicos que permitem fazer uma análise do enorme volume de dados existentes e com interesse para a sua atividade, sendo indispensáveis para a tomada de decisão informada, requisito fundamental de uma gestão eficaz.

No setor da Educação, em concreto no ensino superior, os sistemas de *Analytics* tecnologicamente robustos que permitam agregar e processar dados que as IES já dispõem, para informar a universidade e servir de guia para uma ação mais informada, são ainda básicos ou mesmo inexistentes: “Although (...) generic software features, the depth of extraction and aggregation, reporting and visualisation functionality of these built-in analytics has often been basic or non-existent” (Ferguson, 2012: 4).

O produto deste trabalho, que se insere neste contexto, concretiza-se na realização do enquadramento teórico da temática *Analytics* na Educação e na apresentação de um protótipo de um sistema de *Academic Analytics*.

Na perspetiva de Lucas, H. C. (1987: 10), um sistema é “um conjunto organizado de procedimentos, que, quando executados, produzem informação para apoio à tomada de decisão e ao controlo das organizações”. Foi este objetivo que se perseguiu no desenvolvimento do protótipo do sistema de *Academic Analytics*. A concretização deste protótipo implicou um conjunto de procedimentos que englobou a agregação, a limpeza e a normalização de fontes de dados e resultou numa base de dados que agrega registos de várias fontes e possibilita a realização de relatórios e análises.

6. TRABALHO FUTURO

O desenvolvimento de um modelo de um sistema de *Analytics* mais efetivo e complexo, que dê resposta às reais necessidades de informação da instituição e que permita organizar e apresentar dados de maneira versátil, de

modo a fornecer informação pertinente para a tomada de decisão relativamente aos vários aspetos da vida da IES, deve ser discutido pelos decisores e pelos diferentes órgãos da IES. Neste sentido, impõe-se a definição das dimensões e dos dados que devem integrar os relatórios do *Academic Analytics*.

Este protótipo terá, também, de evoluir para um sistema com uma base tecnológica mais robusta, que permita ler dados de várias fontes e alimentar a base de dados com outros sistemas tecnológicos em uso na instituição. A filtragem de dados por dimensão, período temporal e escala de análise é requisito fundamental para a eficácia do *Academic Analytics* na gestão das várias dimensões e a vários níveis da vida da instituição.

A solução poderá passar pela construção no servidor de uma aplicação de agregação das fontes de dados e pelo desenvolvimento de um sistema de análises (aplicação clássica na *web*) ou aplicação de uma ferramenta de cubos OLAP que replique o modelo e permita análises pré-formatadas (relatórios predefinidos) e deixe flexibilidade para se definirem novos relatórios, consoante as necessidades de informação da IES, acrescentando versatilidade.

Referências

- Bach, C. (2010). Learning Analytics: Targeting Instruction, Curricula and Student Support. *Office of the Provost*. Retrieved from http://www.iiis.org/CDs2010/CD2010SCI/EISTA_2010/PapersPdf/EA655ES.pdf.
- Barneveld, A., Arnold, K., & Campbell, J. (2012). Analytics in Higher Education: Establishing a Common Language. *EDUCAUSE Learning Initiative (ELI) White Paper*.
- Brown, M. (2011). Learning Analytics: The Coming Third Wave. *Learning Initiative Brief*, 1-4.
- Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic Analytics: A New Tool for a New Era. *Educause Review Online* (Vol. 42, pp. 40-57).
- Campbell, J., & Oblinger, D. (2007). *Academic analytics*. Retrieved 2013-08-31, from <http://connect.educause.edu/library/abstract/AcademicAnalytics/45275>.
- Campbell, J. P. (2007). *The Grand Challenge: Using Analytics to Predict Student Success*. Paper presented at the EDUCAUSE Midwest Regional Conference, Chicago, Illinois.
- Castells, M. & Himanen, P. (2007). *A Sociedade da Informação e o Estado-Providência: O modelo finlandês*. Lisboa: Fundação Calouste Gulbenkian.
- Eckerson, W. W. (2007). Predictive Analytics: Extending the Value of Your Data Warehousing Investment, *TDWI Best Practices Report, First Quarter 2007*. Renton, WA, USA: The Data Warehousing Institute.

- Ferguson, R. (2012). The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report KMI-12-01. Milton Keynes, UK: Knowledge Media Institute, The Open University.
- Ferreira, S. A. & Andrade, A. (2012a). Ambientes de aprendizagem ricos em tecnologia – Arquitetura e contributos para a gestão. *Revista Portuguesa de Investigação Educacional*, 12, 241-272.
- Ferreira, S. A. & Andrade, A. (2012b). Conception of a management tool of Technology Enhanced Learning Environments. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 3(2), 42-47.
- Fuller, R. B. & McHale, J. (1965). *World Design Science Decade, 1965-1975: Five Two-year Phases of a World Retooling Design Proposed to the International Union of Architects for Adoption by World Architectural Schools*: World Resources Inventory.
- Goldstein, P. J. (2005). Academic Analytics: The Use of Management Information and Technology in Higher Education (Vol. Educuse Center for Applied Research). ECAR Key Findings.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems. *MIS Quarterly Vol. 28, No. 1*, 28(1), 75-105.
- ITU. (2013). World Summit on the Information Society. Retrieved 07-24, 2013, from <http://www.itu.int/wsis/index.html>.
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). *NMC Horizon Report: 2013. Higher Education Edition*. Austin, Texas.
- Johnson, L., Smith, L. R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas.
- Long, P. & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. *Educause Review Online*, (46), 31-40. Retrieved from <http://net.educause.edu/ir/library/pdf/erm1151.pdf>.
- Lucas, H. C. (1987). *Information Systems, Concepts for Managements* (3rd ed.). Singapura: McGraw Hill, International Editions.
- McKay, J., Marshall, P., & Hirschheim, R. (2012). The design construct in information systems design science. *Journal of Information Technology*, 27, 125-139.
- Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action Analytics. Measuring and Improving Performance that Matters in Higher Education. *Educause Review Online*, 28(1), 42-67. Retrieved from <http://net.educause.edu/ir/library/pdf/ERM0813.pdf>
- Norris, D., Baer, L., & Offerman, M. (2009). *A National Agenda for Action Analytics – White paper*. Paper presented at the National Symposium on Action Analytics, St. Paul, Minnesota. <http://lindabaer.efoliomn.com/uploads/settinganationalagendaforactionanalytics101509.pdf>.

- OLAP.COM. (2010). OLAP software and education wiki Retrieved Web Page, 2010, from http://olap.com/w/index.php/OLAP_Education_Wiki .
- World Summit on the Information Society (2001).
- Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). *The Design Science Research Process: A model for producing and presenting information systems research*. Paper presented at the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006) Claremont, California.
- Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20.
- Sampiere, R., Collado, C., & Lucio, P. (2006). *Metodologia de Pesquisa* (Vol. 3). São Paulo: McGrawHill.
- Santos, F. D. (2013). Big data, big brother e Snowden. *Público*. Retrieved from <http://www.publico.pt/ecosfera/noticia/big-data-big-brother-e-snowden-1601186#> .
- Shum, S. B. (2012). *Policy Brief – Learning Analytics*. Moscow, Russian Federation: UNESCO.
- Siemens, G. (2012). *The Data-intensive University*. Paper presented at the American Association of State Colleges and Universities Conference, San Francisco, California.
- SoLAR. (2013). *Third Conference on Learning Analytics and Knowledge*. Retrieved 06-24, 2013, from <http://lakconference2013.wordpress.com/> .
- Toffler, A. (1999). *A terceira vaga*. Lisboa: Livros do Brasil.
- Varajão, J. (1998). *A arquitetura da gestão de sistemas de informação* (2.ª ed.). Lisboa: FCA Editora Informática.
- Wikipédia. (2013). Design science. *Wikipédia*. Retrieved from http://en.wikipedia.org/wiki/Design_science#Design_as_science_in_information_systems .

ABSTRACT: Higher education institutions (HEIs) use several technological subsystems for the management of multiple aspects of the organization, which collect a huge amount of data from these subsystems, but which do not communicate with each other. The idea of the Analytics in Education is based on the possibility of using this data in an aggregated way in the process of decision making and the development of new models in management. Using the reality of the Universidade Católica Portuguesa – Porto as a starting point, we present the architecture and the results achieved in the development of a prototype of Academic Analytics, for the management of teaching activity at the institution, which aggregates data from three technological subsystems. Three main contributions resulted from this study: i) Theoretical – Analytics in Education is a new theme with rapid rising interest, and therefore it is difficult to maintain a stable discourse when defining concepts and describing processes. In this context, we present a reflection on the main issues discussed and list the most relevant concepts; ii) Methodology – The

rationale and application of the Model Design Science Research Process (DSRP), which, despite its versatility in the field of presentation of artifacts, still has little importance in the field of Information Systems (IS) applied to Social and Human Sciences; iii) Practical – Presentation of a prototype of an Academic Analytics system.

KEYWORDS: Academic Analytics, design science, high education, management.