

Uma métrica de tamanho de software como ferramenta para a governança de TI



Marcus Vinícius Borela de Castro

é servidor do Tribunal de Contas da União, graduado em Informática pela Universidade Federal de Viçosa, especialista em Governança de Tecnologia da Informação pela Universidade de Brasília e certificado na métrica pontos de função (CFPS).



Carlos Alberto Mamede Hernandes

é servidor do Tribunal de Contas da União, graduado em Processamento de Dados pela Universidade de Brasília e mestre em Gestão do Conhecimento e da Tecnologia da Informação pela Universidade Católica de Brasília.

RESUMO

No presente trabalho¹ propõe-se uma nova métrica de software de tamanho funcional derivada da Análise de Pontos de Função (APF), porém sem algumas de suas deficiências conhecidas. Os resultados estatísticos demonstram que a nova métrica, denominada Elementos Funcionais, e sua submétrica Elementos Funcionais de Transação possuem maior correlação com o esforço despendido no desenvolvimento de software do que a APF no contexto dos dados analisados. Ilustra-se a aplicação da nova métrica como ferramenta para aprimorar a governança de TI com seu uso em atividades de avaliação, monitoramento e direção da área de desenvolvimento e manutenção de sistemas de informação.

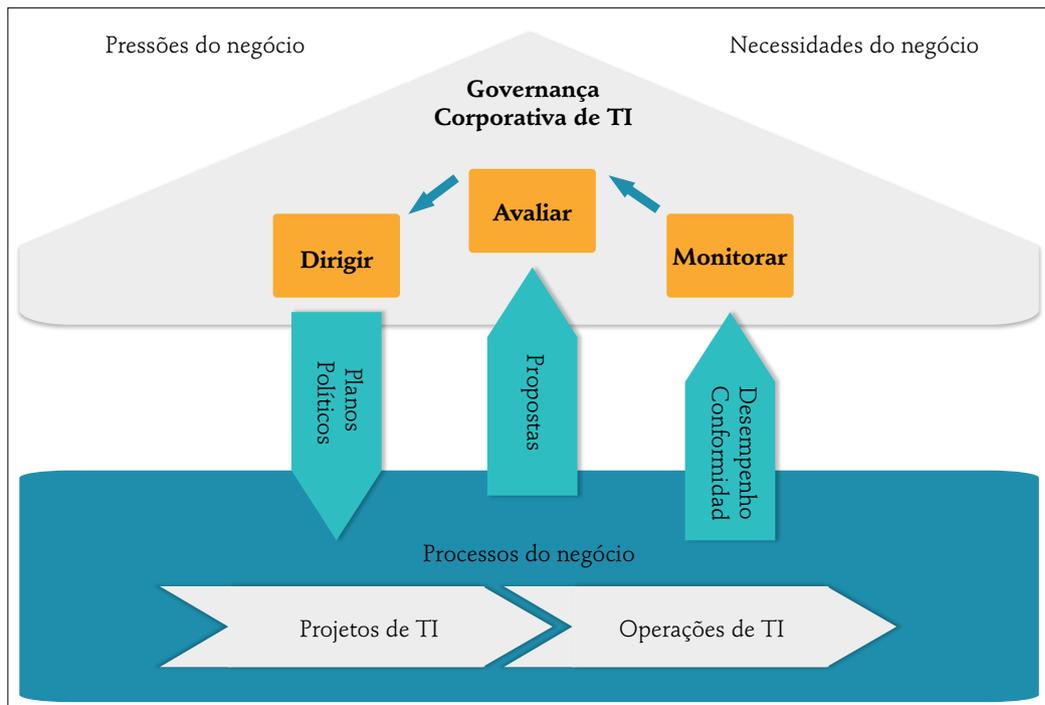
Palavras-chave: Engenharia de Software; Governança de TI; Indicadores de desempenho de TI; Métrica de Software; Pontos de Função.

1. INTRODUÇÃO

1.1 QUESTÃO DE PESQUISA

Rubin (1993, p. 473) afirma que as organizações precisam alavancar os investimentos em tecnologia para criar novas possibilidades e promover mudanças em suas capacidades. Segundo o ITGI (2007, p. 7), a tecnologia da informação (TI) tem se tornado parte integrante do negócio de muitas empresas, com papel fundamental

Figura 1:
Ciclo Avaliar-Dirigir-Monitorar da governança de TI



Fonte: ISO 38500 (ISO/IEC, 2008, p. 7)

para apoiar e promover o seu crescimento. Nesse contexto, a governança de TI cumpre papel importante de direcionar e impulsionar a TI para o atingimento de seus objetivos em consonância com a estratégia da empresa.

Para que a governança da TI propicie o sucesso da TI e da organização, a norma ISO 38500 (ISO; IEC, 2008, p. 7) propõe três atividades principais: avaliar o uso corrente e futuro da TI; dirigir a preparação e a implantação de planos e políticas para assegurar que a TI atinja os objetivos organizacionais; e monitorar o desempenho e a conformidade com essas políticas (Figura 1).

Uma métrica de tamanho de software pode compor diversos indicadores para ajudar a revelar a real situação da área de desenvolvimento de sistemas para a alta administração de uma organização, diretamente ou por intermédio de uma estrutura de governança de TI (e.g. comitê diretivo de TI). Medidas como a produção de software em um período (e.g. medida de tamanho de software/mês) e a produtividade de uma área (e.g. medida de tamanho de software/esforço) são exemplos de indicadores que podem apoiar as três atividades de governança propostas pela norma ISO 38500.

Para a formação desses indicadores, pode-se usar a técnica de dimensionamento de tamanho de software Análise de Pontos de Função (APF) para derivação do número de pontos de função (PF). Criada por Albrecht (1979), a APF tornou-se um padrão internacional de me-

dição funcional com a denominação ISO 20926 (ISO; IEC, 2009). Suas regras são mantidas e aprimoradas por um grupo internacional de usuários, sem fins lucrativos, denominado IFPUG (*International Function Point Users Group*), responsável por publicar o Manual de Práticas de Contagem (CPM – *Counting Practices Manual*), hoje na versão 4.3.1 (IFPUG, 2010).

A APF, por possuir correlação direta com o esforço despendido no desenvolvimento do software (ALBRECHT; GAFFNEY, 1983; KEMERER, 1987), tem sido usada como ferramenta para a gestão do processo de desenvolvimento e manutenção de software não só no Brasil como em todo o mundo. Conforme identificado no relatório da *Pesquisa de Qualidade no Setor de Software Brasileiro de 2009* (BRASIL, 2009, p. 93), a APF é a métrica mais usada para avaliação de tamanho de produto de software entre as empresas da área de software no Brasil, usada por 34,5%. Conforme levantamento de Bundschuh e Dekkers (2008, p. 393), 80% dos projetos cadastrados na base ISBSG (*International Software Benchmarking Standards Group, release 10*) que aplicaram métrica utilizaram a APF.

Considera-se a métrica APF um instrumento bastante eficaz para medição de contratos (VAZQUEZ et al., 2011, p. 191). Contudo, tem a limitação de não tratar requisitos não funcionais², como critérios de qualidade e restrições de tempo de resposta, por exemplo.

Constata-se o uso da APF no âmbito das instituições do governo federal para contratações de desenvolvimento e manutenção de sistemas. Além dos diversos acordos³ do Tribunal de Contas da União (TCU) que apontam seu uso em contratações, como uma opção de dimensionamento do objeto contratado, o *Roteiro de Métricas do SISF* (BRASIL, 2012) preconiza sua aplicação nos órgãos e entidades da administração pública federal.

Apesar do uso extensivo da métrica APF, um grande número de críticas à sua validade e à sua aplicabilidade, detalhadas na seção 2.2, põe em xeque a correção do seu uso em contratações e a confiabilidade de sua aplicação como ferramenta para a gestão e para a governança da TI.

Assim surge a questão para a pesquisa: é possível propor uma métrica para desenvolvimento de software, com a aceitação e a praticidade da APF, ou seja, baseada em seus conceitos já amplamente conhecidos, porém sem algumas falhas identificadas, de forma a potencializar seu uso como ferramenta para governança da TI, com foco na área de desenvolvimento e manutenção de sistemas?

1.2 JUSTIFICATIVA DA PESQUISA

As justificativas para a pesquisa podem ser analisadas a partir do interesse de diversos atores envolvidos no contexto de desenvolvimento e manutenção de software:

1. comitês de governança: a métrica pode derivar indicadores que permitem uma maior governabilidade da TI;
2. gestor da TI⁴: a métrica pode permitir um melhor controle em direção ao alcance dos objetivos traçados pela alta administração;
3. fornecedores do mercado privado e órgãos públicos: a métrica pode aumentar a objetividade do relacionamento, de modo a possibilitar contratos com menor probabilidade de problemas, com pagamento por resultado e a preços justos;
4. órgãos de controle, como o TCU: a métrica pode apoiar a avaliação, em bases mais objetivas, das contratações públicas de desenvolvimento de software (e.g. avaliações a respeito do planejamento da TI, do planejamento das contratações e da gestão dos contratos);

5. instituições de pesquisa: o estudo proposto pode servir de base para que novos estudos sejam realizados; afinal, a área de métricas encontra-se carente de boas pesquisas. Levantamento de Jörgensen e Shepperd (2007, p. 36) demonstra que a maioria das pesquisas em custo de software não leva em consideração artigos já publicados e critica a obsolescência dos dados usados.

1.3 CLASSIFICAÇÃO QUANTO À METODOLOGIA

O trabalho pode ser classificado como uma pesquisa prática, conforme classificação de Demo (apud ANDRADE, 2002, p. 4), pois objetiva resolver problemas de aplicação real, conforme citado na seção 1.2.

Conforme Andrade (2002, p. 5-6), quanto aos objetivos, o artigo é exploratório ao propor uma nova abordagem de métrica, e também descritivo ao apresentar conceitos, como métrica de software, e ilustrar sua aplicação em governança de TI. Quanto à abordagem para o alcance desses objetivos, o trabalho pode ser classificado como dedutivo (ANDRADE, 2002, p. 11), pois propõe uma nova métrica a partir de conceitos teóricos. Quanto aos procedimentos adotados, o trabalho utiliza o método estatístico (ANDRADE, 2002, p. 14) para construção e avaliação dos resultados.

1.4 OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

1. apresentar uma visão geral sobre métrica de software e sobre a APF;
2. apresentar as críticas à técnica APF que motivaram a proposição da nova métrica;
3. derivar uma nova métrica nos moldes da APF;
4. avaliar a nova métrica em comparação à APF quanto à correlação com o esforço;
5. ilustrar o uso da métrica proposta no âmbito da governança de TI, no contexto da área de desenvolvimento e manutenção de sistemas.

Aborda-se cada objetivo específico em subseção própria na parte de desenvolvimento, a seguir.

2. DESENVOLVIMENTO

2.1 MÉTRICAS DE SOFTWARE

2.1.1 Conceituação, categorização e aplicação

Bundschuh e Dekkers (2008, p. 180-181) descrevem diversas interpretações dadas na literatura aos termos “métrica”, “medida” e “indicador”. No contexto deste trabalho, não se faz distinção entre os três termos. Toma-se para esses termos a definição de Fenton e Pfleeger (1998, p. 5) para medida: um número ou símbolo que caracteriza um atributo de uma entidade do mundo real, objeto ou evento, a partir de regras formalmente definidas⁵.

Segundo Fenton e Pfleeger (1998, p. 74), métricas de software se aplicam a três tipos de entidades: processos, produtos e recursos. Eles também distinguem métricas diretas, quando envolvem apenas um atributo de uma entidade, de métricas indiretas, o caso contrário (FENTON; PFLEEGER, 1998, p. 39). As indiretas são derivadas por regras baseadas em outras métricas. A velocidade de entrega de uma equipe (tipo de entidade: recurso) é um exemplo de métrica indireta, pois é calculada a partir da razão entre duas medidas: tamanho produzido de software (produto) e tempo decorrido do desenvolvimento (processo). Esta última é um exemplo de métrica direta. Moser (1996, p. 32) diferencia métricas de tamanho de métricas de qualidade ao afirmar que métricas de tamanho distinguem entre o menor e o maior, enquanto métricas de qualidade diferenciam entre o bom e o mau. O Quadro 1 consolida os exemplos citados de categorias de métricas de software.

Moser (1996, p. 31) comenta que, dado o relacionamento entre um produto e o processo que o produziu, pode-se atribuir uma medida de processo para um produto e *vice-versa*. Por exemplo, o percentual do esforço em testes, atributo do processo de desenvolvimento, pode ser associado ao produto gerado, como indicador

de sua qualidade, e o número de erros em produção nos primeiros três meses, atributo de um sistema (produto), pode ser atribuído ao processo de desenvolvimento, como indicativo de sua qualidade.

Fenton e Pfleeger (1998, p. 12) colocam três objetivos para uma métrica de software: entender, controlar e melhorar a entidade alvo da medição. Eles alertam que a definição das métricas depende do nível de maturidade do processo a ser medido: quanto mais maduro, mais visível e, portanto, mais mensurável (FENTON; PFLEEGER, 1998, p. 83). Chikofsky e Rubin (1999, p. 76) ressaltam que um programa inicial de medição para as áreas de desenvolvimento e de manutenção de sistemas deve abranger cinco dimensões fundamentais que endereçam atributos centrais para o planejamento, controle e melhoria de produtos e processos: tamanho, esforço, prazo, qualidade e retrabalho. Refutando a possibilidade de se medir sem se antever o objetivo, lembram os autores que o objetivo não são as métricas em si, mas as decisões que serão tomadas a partir delas (CHIKOFSKY; RUBIN, 1999, p. 75).

Segundo Beyers (2002, p. 337), o uso de métricas em estimativas (e.g. de tamanho, prazo, custo, esforço, qualidade e alocação de pessoas) ajuda nas tomadas de decisão relacionadas ao desenvolvimento de software e ao planejamento de projetos de software.

2.1.2 Visão geral da APF

Conforme categorização da seção anterior, a APF é uma métrica indireta de tamanho de produto. Dimensiona o tamanho funcional de uma aplicação (sistema), como uma medida das funcionalidades solicitadas e entregues ao usuário. Trata-se de uma métrica compreensível pelo usuário do software, independente da tecnologia aplicada⁶.

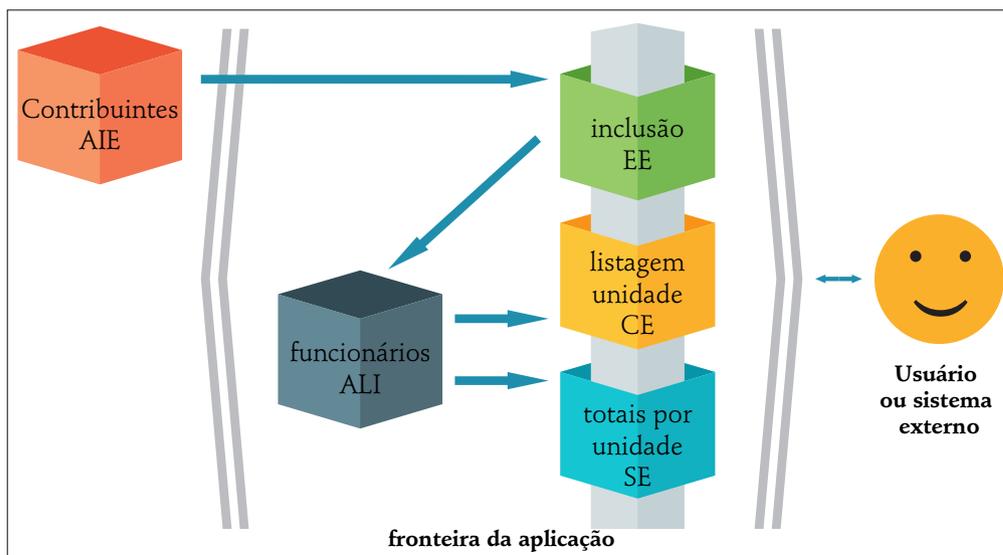
Cabe ressaltar que, além da APF, há outras quatro métricas funcionais consideradas padrão ISO de métrica funcional, por seguirem as regras definidas nas seis nor-

Quadro 1:
Exemplos de categorias de métricas de software

Critério	Categorias	Fonte
Entidade	De processo De produto De recurso	Fenton e Pfleeger, 1998, p. 74
Número de atributos envolvidos	Direta Indireta	Fenton e Pfleeger, 1998, p. 39
Diferenciação do alvo	De tamanho De qualidade	Moser, 1996, p. 32

Figura 2:

Visualização dos tipos de funcionalidade que compõem uma aplicação na APF



mas da série ISO 14143 (ISO; IEC, 2002a, 2003, 2004, 2006, 2007, 2011a): MKII FPA (ISO; IEC, 2002b), COSMIC-FFP (ISO; IEC, 2011b), FiSMA (ISO; IEC, 2010) e Nesma (ISO; IEC, 2005). Conforme Gencil e Demirors (2008, p. 4), as métricas funcionais padrão ISO estimam tamanho de software a partir da funcionalidade entregue aos usuários, com diferença nos objetos contados e na forma como são contados⁷.

As funcionalidades podem ser de dois tipos: transações, que implementam trocas de dados com usuários e outros sistemas, e arquivos de dados, que indicam a estrutura dos dados armazenados. São três tipos de transações: consultas externas (CE), saídas externas (SE) e entradas externas (EE), conforme a intenção primária da transação seja, respectivamente, uma simples consulta, uma consulta mais elaborada (e.g. com totalizações) ou uma atualização de dados. São dois tipos de arquivos lógicos de dados: arquivos lógicos internos (ALI) e arquivos de interface externa (AIE), conforme seus dados sejam atualizados ou apenas referenciados (acessados) no contexto da aplicação.

A Figura 2 ilustra graficamente os tipos de função. Para facilitar a compreensão, pode-se consi-

derar a EE como, por exemplo, uma funcionalidade de *incluir funcionário* que atualiza o arquivo de dados *funcionários* (ALI) e valida o CPF informado pelo usuário acessando o arquivo *contribuintes* (AIE), externo à aplicação, que contém os dados de CPF da Receita Federal. Também na aplicação temos, hipoteticamente, uma listagem, consulta simples, de funcionários lotados em uma unidade (CE) e um relatório mais complexo com o quantitativo de funcionários por unidade (SE).

Na regra de cálculo da APF, cada funcionalidade é avaliada quanto à sua complexidade e assume uma das três classificações: simples (de baixa complexidade), média (de média complexidade) ou complexa (de alta complexidade). A cada nível de complexidade associa-se um tamanho em pontos de função. A Tabela 1 ilustra a regra de derivação do tamanho de funcionalidades do tipo consulta externa a partir do número de arquivos acessados (ALR – arquivos lógicos referenciados) e do número de campos que cruzam a fronteira da aplicação (TD – Tipo de Dado).

Assim como para CE, há regras específicas para derivação da complexidade de cada tipo de funcionalidade (EE, SE, ALI e AIE), bem como para definição

Tabela 1:

Derivação da complexidade e do tamanho em PF de CE

ALR (arquivos)	TD (campos)	1 a 5	6 a 19	20 ou mais
1		baixa (3)	baixa (3)	média (4)
2 a 3		baixa (3)	média (4)	alta (6)
4 ou mais		média (4)	alta (6)	alta (6)

Quadro 2:

Categorias dos atributos funcionais por tipo de funcionalidade

Funcionalidades	Atributos Funcionais
Transações: EE, SE e CE	Arquivos referenciados (ou ALR – arquivos lógicos referenciados) e Campos (ou TD – tipos de dados)
Arquivos lógicos: ALI e AIE	Registros lógicos (ou RLR – registros lógicos referenciados) e Campos (ou TD – tipos de dados)

do tamanho associado à complexidade, de forma semelhante à Tabela 1. O Quadro 2 resume as categorias dos atributos usados no cálculo de pontos de função de cada tipo de funcionalidade.

O tamanho de um software é o resultado da soma dos tamanhos de suas funcionalidades. Foge ao escopo do trabalho a apresentação aprofundada dos conceitos associados à APF. Detalhes da métrica podem ser obtidos no *Manual de Práticas de Contagem de Pontos de Função, versão 4.3.1* (IFPUG, 2010).

2.2 CRÍTICAS À APF QUE MOTIVARAM A PROPOSIÇÃO DA NOVA MÉTRICA

Apesar do uso extensivo da métrica APF, citado na seção 1.1, há um grande número de críticas à sua validade e à sua aplicabilidade que põe em xeque a correção do seu uso em contratações e fragilizam a confiabilidade de sua aplicação como ferramenta para a gestão e para a governança da TI (ABRAN; ROBILLARD, 1994; FENTON; PFLEEGER, 1998; KITCHENHAM, 1997; KITCHENHAM; KÄNSÄLÄ, 1993; KITCHENHAM et al., 1995; KRALJ et al., 2005; PFLEEGER et al., 1997; TURETKEN et al., 2008; XIA et al., 2009).

Várias métricas foram propostas a partir da APF, seja para adequá-la a determinados modelos, seja para aperfeiçoá-la com a correção de algumas falhas conhecidas. Como ilustração, temos o trabalho de Antonioli et al. (2003), que propõe uma métrica para o modelo orientado a objetos, e o artigo de Kralj et al. (2005), que propõe uma alteração na APF para tornar mais precisa a medição de funcionalidades de alta complexidade (item 4, a seguir).

A métrica proposta neste trabalho não objetiva resolver todas as deficiências da métrica APF, mas contribui para reduzir os seguintes problemas relacionados à sua definição:

1. baixa representatividade: embora não haja limite no número de combinações possíveis dos elementos considerados no cálculo da complexidade de uma funcionalidade pela APF, a técnica

restringe o tamanho de uma função a apenas três valores, conforme sua complexidade seja baixa, média ou alta;

2. funcionalidades com complexidades diferentes dimensionadas com o mesmo tamanho: Pfleeger et al. (1997, p. 36) afirmam que, se H é uma medida de tamanho e se A é maior do que B , então H_A deve ser maior do que H_B . Xia et al. (2009, p. 3) descrevem exemplos de funções com complexidades diferentes que recebem indevidamente um mesmo valor em pontos de função por possuírem uma mesma classificação de complexidade;

3. transição abrupta entre faixas: Xia et al. (2009, p. 4) exemplificam a questão ao apresentar dois arquivos lógicos B e C , de complexidades aparentes bem próximas, com diferença apenas no número de campos: B com 19 campos e C com 20 campos. Ao se dimensionar os dois arquivos, classificam-se B e C como de baixa complexidade (7 PF) e de média complexidade (10 PF), respectivamente. A diferença nas medições efetuadas recai sobre a transição de duas faixas: até 19 campos, considera-se complexidade baixa; a partir de 20 campos, considera-se complexidade média. Como a variação de 1 até 19 campos não implica mudança de faixa e de tamanho em pontos de função, considera-se incoerente o acréscimo de apenas um campo levar ao acréscimo em 3 pontos de função. Resultado semelhante ocorre em outras transições de faixas;

4. dimensionamento limitado de funcionalidades com alta complexidade: APF impõe limites superiores ao tamanho de uma função. Kralj et al. (2005, p. 83) descrevem a situação de funções que são indevidamente classificadas como de alta complexidade. Alertam para a necessidade de haver números maiores para complexidades maiores e propõem uma alteração na forma de cálculo da APF como solução⁸;

5. operação em escala ordinal: como visto, a APF envolve a classificação da complexidade de funcionalidades em baixa, média ou alta, atribuindo-lhes, assim, uma escala ordinal. Em seguida, esses rótulos são substituídos por números. Um arquivo lógico interno, por exemplo, recebe 7, 10 ou 15 pontos de função, conforme sua complexidade seja baixa, média ou alta, respectivamente. Kitchenham (1997, p. 29) critica a APF pela impropriedade em se somar medidas em escala ordinal. Reforça não fazer sentido somar o rótulo *baixa complexidade* com o rótulo *alta complexidade*, ainda que se usem os rótulos 7 e 15, respectivamente, como sinônimos;

6. impossibilidade de se dimensionar alterações em partes de uma funcionalidade: essa característica impede o dimensionamento em pontos de função, por exemplo, do tamanho de parte de uma funcionalidade que precisa ser alterada em uma manutenção. Assim, uma função tratada em várias iterações em um método ágil ou outro processo iterativo será sempre medida com o tamanho completo, ainda que a alteração seja considerada pequena.

Dadas as deficiências identificadas, a correlação entre o tamanho em pontos de função de um software e o esforço necessário para o seu desenvolvimento tende a não ser adequada, uma vez que a APF apresenta deficiências de representação do real tamanho funcional do software. Se há inexatidão no dimensionamento do tamanho do que deve ser feito, não há como se esperar uma definição adequada do esforço necessário

e, por conseguinte, uma exatidão na definição do custo do desenvolvimento ou da manutenção. Os problemas citados motivaram o desenvolvimento deste trabalho, de modo a se propor uma métrica com escala quantitativa, com valores infinitos, denominada Elementos Funcionais (EF).

2.3 PROCESSO DE DERIVAÇÃO DA NOVA MÉTRICA

A nova métrica proposta neste trabalho, Elementos Funcionais, adota os mesmos conceitos da APF, com alteração somente na derivação do tamanho das funcionalidades⁹.

O processo dedutivo de derivação da nova métrica, detalhado nas próximas seções, implementa uma regressão linear parecida com a do Gráfico 1. Para cada tipo de funcionalidade, objetiva-se derivar uma fórmula para cálculo do número de EF (Tabela 5 da seção 2.3.4) a partir dos quantitativos de atributos funcionais¹⁰ considerados na derivação de sua complexidade, indicados no Quadro 2 da seção 2.1.2.

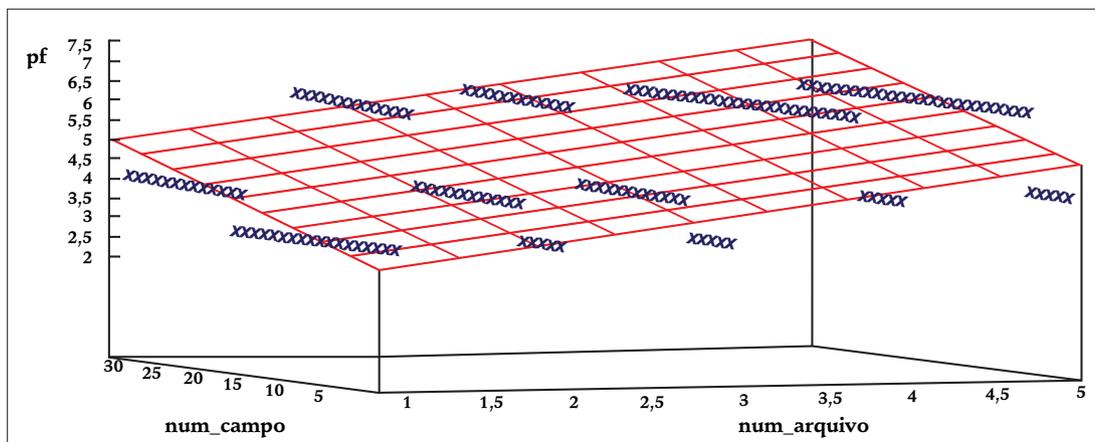
No Gráfico 1, os pontos X indicam o valor derivado em PF (eixo Z) de uma consulta externa a partir do número de arquivos (*num_arquivo*) e do número de campos (*num_campo*), atributos usados na derivação de sua complexidade (vide Tabela 1 da seção 2.1.2). O plano quadriculado simboliza o resultado de uma regressão linear a partir desses pontos e representa, de forma resumida, o valor da nova métrica.

2.3.1 Passo 1 – definição das constantes

Se os valores associados às duas categorias de atributos funcionais forem zero, a métrica assumirá o

Gráfico 1:

Derivação do número de PF (eixo Z) de uma consulta externa a partir dos atributos de cálculo



valor de uma *constante*. Os atributos podem ser dimensionados como zero, por exemplo, no caso de manutenção apenas no algoritmo das funcionalidades que não implique alteração no número de campos ou de arquivos referenciados. No contexto da nova métrica, o dimensionamento de uma operação de exclusão de uma funcionalidade assume o valor da constante, uma vez que não há atributos especificamente impactados por essa operação.

O valor atribuído a essas constantes segue proposição da métrica Nesma (*Netherlands Software Metrics Association*), padrão ISO de medição funcional, para esses casos que não envolvem atributos, conforme documentado em *Function Point Analysis For Software Enhancement* (NESMA, 2009). A própria APF (IFPUG, 2010, v. 4, p. 94) indica a Nesma como alternativa para medição de manutenções, por sua capacidade de tratar a 6ª crítica da seção 2.2. A Nesma dimensiona uma manutenção com a multiplicação do tamanho original da função por um fator de impacto da alteração. O fator impacto deriva da proporção entre o volume dos atributos (e.g. campos) incluídos, alterados ou excluídos e o volume original desses na função. O fator de ajuste assume valores múltiplos de 25%, até o teto de 150%.

Foi assumido, para cada tipo de funcionalidade, o menor valor possível pela aplicação da técnica Nesma, ou seja, o equivalente a 25% do número de PF de uma função de baixa complexidade de cada tipo, ou seja: AIE – 1,25 (25% de 5); ALI – 1,75 (25% de 7); CE – 0,75 (25% de 3); EE – 0,75 (25% de 3) e SE – 1 (25% de 4).

2.3.2 Passo 2 – tratamento das faixas com número ilimitado de elementos

Na APF, cada tipo de funcionalidade possui uma tabela própria para derivação da complexidade de uma função de modo similar à Tabela 1 da seção 2.1.2, que apresenta os valores das faixas dos atributos funcionais para a derivação da complexidade de consultas externas. A terceira e última faixa de valores de cada elemento funcional nas tabelas de cálculo de todos os tipos de funcionalidades são ilimitadas, como o caso da primeira célula da quarta coluna da Tabela 1 da seção 2.1.2, 20 ou mais TD, e da última célula da primeira coluna, 4 ou mais ALR.

Para se gerar um conjunto de dados finito para a regressão, estabeleceu-se um limite superior para essas faixas com o número de elementos equivalente ao da maior faixa precedente¹¹. No caso das faixas citadas para consultas externas, o número de campos ficou limita-

do em 33, resultado de se definir 14 como o número de elementos da terceira faixa (20 a 33), mesmo tamanho da maior faixa (6 a 19 – 14 elementos). O número de arquivos referenciados ficou limitado em 5, por igual raciocínio. A limitação da faixa é um artifício matemático para se evitar um limite superior para a nova métrica (4ª crítica na seção 2.2).

2.3.3 Passo 3 – geração de pontos para regressão

O objetivo dessa etapa foi gerar, para cada tipo de funcionalidade, um conjunto de registros de dados com três valores: os valores dos atributos funcionais e o número de PF derivado desses, decrescido do valor da constante indicado no passo 1. A Tabela 2 ilustra alguns pontos gerados para consulta externa.

Foi desenvolvido um aplicativo em MS-Access para geração de um conjunto de dados com todos os pontos possíveis para todas as faixas das tabelas de complexidade dos cinco tipos de função, limitadas em número de elementos, conforme passo anterior. A Tabela 3 ilustra todas as combinações das faixas consideradas para CE.

2.3.4 Passo 4 – regressão linear

Os diversos pontos obtidos na seção anterior foram importados para o software Excel 2007 para a realização da regressão linear entre o tamanho em PF e os atributos funcionais, usando-se o método dos Mínimos Quadrados Ordinários, sem constante, uma vez que elas já foram definidas no passo 1 e decrescidas do valor esperado no passo 3.

Os resultados estatísticos da regressão são mostrados por tipo de função na Tabela 4.

Para cada tipo de funcionalidade foi derivada uma fórmula com os valores dos coeficientes arredondados em duas casas decimais (Tabela 5). Cada fórmula deriva o número de elementos funcionais, variável dependente, métrica proposta, a partir das dimensões impactantes no cálculo, variáveis independentes, atributos funcionais, já com as constantes indicadas no passo

ALR	TD	PF (menos constante do passo 1)
1	1	2,25
1	2	2,25 (...)
1	33	3,25
2	1	2,25 (...)

Tabela 2:

Extrato do resultado com alguns dados associados à consulta externa

1. As siglas Eft e EFd representam respectivamente os elementos funcionais associados a transações (CE, EE e SE) e a dados (AIE e ALI).

A métrica elementos funcionais, EF, resulta da soma dos elementos funcionais de transação, Eft, com os elementos funcionais de dados, EFd, conforme fórmulas explicitadas na Tabela 5. Subdivide-se, assim, a métrica proposta: $EF = Eft + EFd$.

A submétrica Eft não considera os arquivos lógicos (ALI e AIE) de forma independente, mas sim enquanto arquivos referenciados no contexto das transações. A não contagem dos arquivos lógicos de forma independente condiz com outras duas métricas padrão ISO de tamanho funcional que também não os consideram (BUNDSCHUH; DEKKERS, 2008, p. 388): MKII FPA (ISO; IEC, 2002b) e COSMIC-FFP (ISO; IEC, 2011b).

Na avaliação da métrica, próxima seção, foram testadas as duas métricas, EF e Eft, contando-se e não se contando os arquivos lógicos, e os resultados mos-

tram que a Eft apresenta uma melhor correlação com o esforço¹². Embora não avaliada, a submétrica EFd tem seu valor ao refletir a complexidade estrutural dos dados de uma aplicação.

2.4 AVALIAÇÃO DA NOVA MÉTRICA

A nova métrica EF, bem como a submétrica Eft, foram avaliadas quanto à correlação com o esforço comparativamente à métrica APF. Não foi objetivo avaliar a qualidade dessas correlações, mas sim a comparação delas quanto à capacidade de explicação (R2) da variação do esforço¹³.

Obteve-se em um órgão público federal uma planilha com registros referentes a Ordens de Serviços (OS) contratadas de empresas privadas para atividades de codificação¹⁴ e de teste. Uma OS envolvia a implementação de um ou mais comandos de manutenção ou desenvolvimento de funcionalidades de um único sistema,

Tabela 3:
Combinações de faixas de valores para cálculo de PF de consultas externas

Tipo funcao	ALR inicial	ALR final	TD inicial	TD final	PF	PF menos constante
CE	1	1	1	5	3	2,25
CE	1	1	6	19	3	2,25
CE	1	1	20	33	4	3,25
CE	2	3	1	5	3	2,25
CE	2	3	6	19	4	3,25
CE	2	3	20	33	6	5,25
CE	4	5	1	5	4	3,25
CE	4	5	6	19	6	5,25
CE	4	5	20	33	6	5,25

Tabela 4:
Resultados estatísticos da regressão – comparativo entre tipos de funcionalidades

	ALI	AIE	SE	EE	CE
R-quadrado	0,96363	0,96261	0,95171	0,95664	0,96849
Observações	729	729	198	130	165
P-value do coeficiente (ALR ou RLR)	3,00E-212	1,17E-211	7,65E-57	1,70E-43	4,30E-60
P-value do coeficiente (TD)	2,28E-231	2,71E-225	1,44E-59	2,76E-39	2,95E-45

Tabela 5:
Fórmulas de cálculo de elementos funcionais por tipo de funcionalidade

Funcionalidade	Fórmula derivada
ALI	$EFd = 1,75 + 0,96 * RLR + 0,12 * TD$
AIE	$EFd = 1,25 + 0,65 * RLR + 0,08 * TD$
SE	$Eft = 1,00 + 0,81 * ALR + 0,13 * TD$
EE	$Eft = 0,75 + 0,91 * ALR + 0,13 * TD$
CE	$Eft = 0,75 + 0,76 * ALR + 0,10 * TD$

como a criação de um relatório ou a alteração de uma transação. A planilha continha para cada OS o tempo gasto na sua execução em horas, bem como o tamanho da equipe em número de pessoas alocadas. Os únicos dados fictícios foram as identificações das OS, dos sistemas e de suas funcionalidades, pois não eram relevantes para o escopo deste trabalho. A partir do produto do tempo pelo tamanho da equipe derivou-se o esforço por OS em homem-hora. No Quadro 3, apresenta-se a estrutura dos dados recebidos.

Foram recebidos dados de 183 OS, porém, 12 foram descartadas por estar com informações duvidosas: campos, tipo de função e operação indefinidos, por exemplo. As 171 OS restantes referenciavam 14 sistemas com 505 comandos relativos a 358 funcionalidades. Para se ter uma maior qualidade na correlação com o esforço, optou-se por considerar apenas os dados de sistemas associados a pelo menos quinze OS, ou seja, quatro

sistemas¹⁵: H, B, C e D. A Tabela 6 indica o quantitativo de OS e de comandos de cada sistema selecionado.

Após o cálculo para cada comando da medida em EF e EFt em um aplicativo MS-Access desenvolvido pelo autor, os dados foram importados para o software Excel 2007 para a realização da regressão linear¹⁶, usando-se o método dos Mínimos Quadrados Ordinários, entre o esforço e o tamanho, calculado nas métricas PF, EF e EFt. Realizou-se a regressão linear com constante nula, uma vez que não existe esforço se não houver tamanho¹⁷. A operação foi feita por sistema, uma vez que a variabilidade dos fatores que influenciam o esforço diminui no contexto de um mesmo sistema¹⁸. O Gráfico 2 ilustra a dispersão dos pontos (OS) na correlação entre o tamanho (EFt) e o esforço (homem-hora) e a reta derivada pela regressão linear no contexto do sistema H.

O coeficiente de explicação R² foi usado para representar o grau da correlação entre o esforço e os

Quadro 3:

Estrutura dos dados recebidos para avaliação da métrica

Sigla	Descrição	Domínio
OS	Número de identificação de uma ordem de serviço	até 10 números
Função	Número de identificação de uma funcionalidade	até 10 números
Tipo	Tipo da funcionalidade conforme categorização da APF	ALI, AIE, EE, SE ou CE
Operação	Indica a operação efetuada. Pode se tratar de uma Inclusão de uma nova função ou Alteração de uma função em um sistema (manutenção)	I ou A
ALR RLR final	Se a funcionalidade for uma transação, indica o número de arquivos lógicos referenciados (ALR). Se for um arquivo lógico, indica o número de registros lógicos (RLR). Indica o quantitativo ao final da implementação do comando	até 3 números
ALR RLR operação	Indica o número de ALR ou RLR que foram incluídos, alterados ou excluídos do escopo da manutenção da funcionalidade (operação de alteração)	até 3 números
ALR RLR original	Indica o número de ALR_RLR que originalmente compunham a função (operação de alteração)	até 3 números
TD final	Indica o quantitativo de campos ao final da implementação do comando	até 3 números
TD operação	Indica o número de TD que foram incluídos, alterados ou excluídos do escopo da manutenção da funcionalidade (operação de alteração)	até 3 números
TD original	Indica o número de TD que originalmente compunham a funcionalidade (operação de alteração)	até 3 números
PF	Número de pontos de função da funcionalidade ao final da operação	até 2 números
%Impacto	Percentual da função original impactado na manutenção, dimensionado pela técnica Nesma (2009)	25, 50, 75, 100, 125, 150
PM	Número de pontos de manutenção, dimensionados pela técnica Nesma (2009) da funcionalidade comandada	até 4 números
Sistema	Identificação única de um sistema	uma letra
Horas	Horas dedicadas pela equipe à execução da ordem de serviço	até 5 números
Equipe	Número de pessoas da equipe responsável pela implementação da OS	até 2 números

diferentes tamanhos calculados em cada métrica avaliada. Segundo Sartoris (2008, p. 244), o R2 indica em uma regressão linear o percentual da variação de uma variável dependente Y, no caso o esforço, que é explicada pela variação de uma segunda variável independente X, no caso a métrica de tamanho avaliada. A Tabela 6 apresenta o resultado das regressões lineares realizadas.

A partir dos resultados da Tabela 6, comparando-se a correlação das métricas com o esforço, observa-se que:

1. as correlações das novas métricas foram consideradas significantes a um nível de confiança de 95% em todos os sistemas (*p-value teste-f* da correlação inferior a 0,05¹⁹). A correlação da APF, contudo, não foi significativa para o sistema B (*p-value* = 0,088 > 0,05).
2. as correlações das novas métricas foram superiores nos dois sistemas com maior número de OS (H e B). Um resultado melhor em amostras maiores é um ponto positivo, pois quanto maior o tamanho da amostra, maior a confiabilidade de seus resultados, haja vista o *p-value* ter alcançado os menores valores para esses sistemas;
3. nenhuma métrica conseguiu alto coeficiente de explicação ($R^2 > 0,8$). Contudo, as novas métricas conseguiram uma correlação considerada média ($0,8 > R^2 > 0,5$) nos 4 sistemas avaliados, enquanto a APF obteve uma correlação fraca ($R^2 < 0,2$) no sistema B, isso se fosse aceito o nível de confiança 91,2% (*p-value* = 0,088) da correlação;

4. as correlações das novas métricas foram superiores²⁰ em 3 dos 4 sistemas (H, B e D), ou seja, em 75% dos sistemas;

Diante desses fatos, no contexto dos dados analisados, pode-se concluir que as métricas propostas, EF e EFt²¹, apresentam uma melhor correlação com o esforço do que a APF²².

O Quadro 4 contém as justificativas de como as métricas propostas, EF e EFt, resolvem as críticas apresentadas na seção 2.2.

2.5 ILUSTRAÇÃO DO USO DA NOVA MÉTRICA NA GOVERNANÇA DE TI

Kaplan e Norton (1992, p. 71) afirmam que o que se mede é o que se obtém (*what you measure is what you get*). Segundo o Cobit 5 (ISACA, 2012b, p. 13), a governança objetiva a criação de valor com a obtenção de benefícios a custos e riscos otimizados. Em relação à governança de TI, a métrica proposta neste trabalho não só ajuda a dimensionar a capacidade da TI como também viabiliza a otimização de seus processos para o alcance dos resultados.

Ao permitir a tradução de objetivos e resultados em números, métricas suportam a comunicação entre os diversos atores da governança de TI (vide Figura 3). Ao se traçar objetivos e se medir resultados por métricas, promove-se o aumento da qualidade de um processo (MOSER, 1996, p. 19). Ao se pensar a função de desenvolvimento de sistemas de informação como um processo a ser governado, pode-se potencializar a capacidade de produção de sistemas para o alcance dos objetivos

Gráfico 2:

Dispersão de pontos (OS) no sistema H: esforço (hh) x tamanho (EFt)

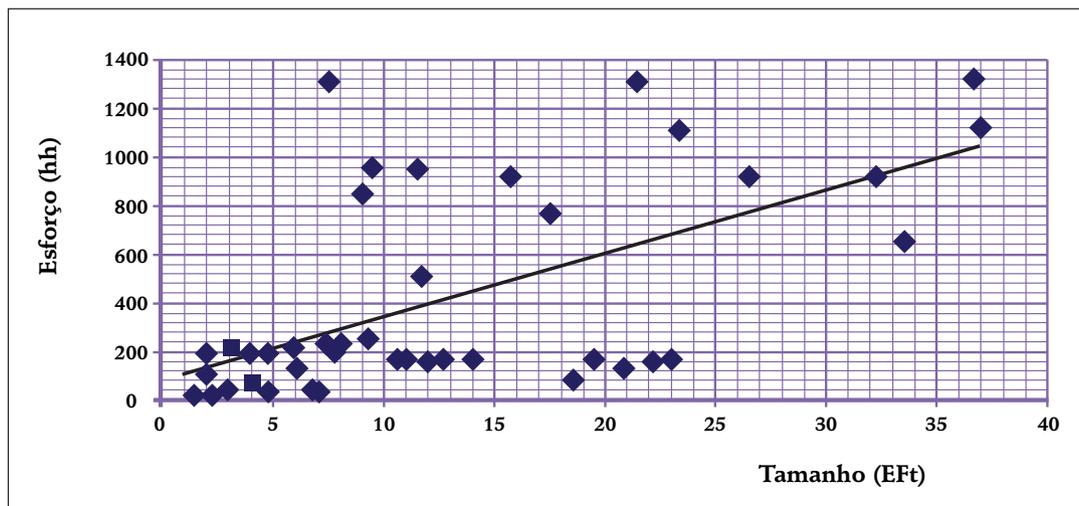


Tabela 6:

Resultados das regressões lineares correlacionando o esforço e as métricas de tamanho

Sistema		H	B	C	D
Qtd_OS		45	25	21	15
Qtd_comandos		245	44	60	20
PF	R2	59,3%	11,2%	67,7%	51,8%
	p-value (teste-f)	4,6E-10	8,8E-02	3,3E-06	1,9E-03
EF	R2	65,1%	60,3%	53,0%	54,7%
	p-value (teste-f)	1,5E-11	2,3E-06	1,4E-04	1,2E-03
	comparação R2 PF	+10%	+438%	-22%	+5%
EFt	R2	66,1%	60,3%	53,0%	54,7%
	p-value (teste-f)	8,5E-12	2,3E-06	1,4E-04	1,2E-03
	comparação R2 PF	+11%	+438%	-22%	+5%

estratégicos traçados com o uso adequado de métricas e estimativas.

Métricas de software contribuem para as três atividades de governança propostas pela norma ISO 38500, citadas na seção 1.1: avaliação, monitoramento e direcionamento da TI. Essas atividades correspondem, respectivamente, aos objetivos de métricas de software citados na seção 2.1.1: entendimento, controle e melhoria da entidade alvo da medição.

Quanto ao direcionamento da TI, Weill e Ross (2006, p. 188) colocam que a criação de métricas para a formalização de escolhas estratégicas é um dos quatro princípios administrativos que resumem como a governança de TI ajuda as empresas a atingir seus objetivos estratégicos. Métricas devem capturar o progresso em

direção às metas estratégicas e indicam assim o funcionamento ou não da governança de TI.

Kaplan e Norton (1996, p. 75-76) afirmam que as estratégias precisam ser traduzidas em um conjunto de objetivos e métricas para que todos se comprometam. Afirmam ser o *Balanced Scorecard* (BSC) uma ferramenta que propicia a todos os níveis da organização o conhecimento das estratégias de longo prazo e que promove o alinhamento dos objetivos departamentais e individuais com essas estratégias. Conforme o ITGI (2007, p. 29), além de ser uma visão holística das operações de negócio, o BSC também contribui para conectar objetivos estratégicos de longo prazo com ações de curto prazo.

Para adaptar os conceitos do BSC para a função de TI, as perspectivas de um BSC foram redefinidas

Quadro 4:

Justificativas de como as novas métricas resolvem as críticas apresentadas na seção 2.2

Crítica	Solução
Baixa representatividade	Cada combinação possível dos elementos considerados na derivação da complexidade da APF está associada a um valor distinto.
Funcionalidades com complexidades diferentes dimensionadas com o mesmo tamanho	Funcionalidades com complexidades diferentes, conforme determinado pelo quantitativo de atributos funcionais, assumem um tamanho diferenciado.
Transição abrupta entre faixas	Com a aplicação das fórmulas de cálculo descritas na seção 2.3.4, a variação do tamanho fica uniforme em relação a cada variação do quantitativo de atributos funcionais, conforme seus coeficientes.
Dimensionamento limitado de funcionalidades com alta complexidade	Não há limite superior no tamanho atribuído a uma função ao se aplicar as fórmulas de cálculo descritas na seção 2.3.4.
Operação em escala ordinal	As métricas não possuem uma escala ordinal com valores finitos, mas sim uma escala quantitativa com infinitos valores discretos, o que permite maior confiabilidade nas operações com os valores.
Impossibilidade de se dimensionar alterações em partes de uma funcionalidade	Permite o dimensionamento de alterações em parte de uma funcionalidade ao se considerar no cálculo apenas os atributos funcionais impactados na alteração.

(VAN GREMBERGEN; VAN BRUGGEN, 1997, p. 3). O Quadro 5 apresenta as perspectivas de um BSC-TI e suas perguntas base.

Segundo o ITGI (2007, p. 30), o BSC-TI ajuda de forma efetiva o corpo governante a alcançar o alinhamento entre a TI e o negócio. Trata-se de uma das melhores práticas para medição de desempenho (ITGI, 2007, p. 46). O BSC-TI é um veículo que organiza as informações para o comitê de governança, cria consenso entre os *stakeholders* sobre os objetivos estratégicos da TI, demonstra a efetividade e o valor adicionado pela TI e comunica informações sobre capacidade, riscos e desempenho (ITGI, 2007, p. 30).

Van Grembergen (2000, p. 2) afirma que o relacionamento entre TI e o negócio pode ser mais explicitamente expresso através de uma cascata de *scorecards*, e divide o BSC-TI em dois: BSC-TI Desenvolvimento e BSC-TI Operação. Rohm e Malinoski (2010), membros do *Balanced Scorecard Institute*, apresentam um processo com nove passos para se construir e implementar estratégias baseadas em *scorecard*. Becker e Bostelmann (1999) apresentam um método para se derivar objetivos e métricas a partir da combinação da técnica BSC com a técnica *Goal Question Metric* (GQM), proposta por Basili e Weiss (1984). Essa associação do BSC ao método GQM condiz com o que afirma a ISACA (2010, p. 74): boas estratégias começam com as perguntas certas. A métrica proposta neste trabalho pode compor vários indicadores que podem ser usadas no BSC-TI Desenvolvimento.

Quanto às atividades de monitoramento e avaliação da TI (ISO; IEC, 2008, p. 7), métricas propiciam o monitoramento da taxa de melhoria das organizações em direção a um processo maduro e melhorado (RUBIN, 1993, p. 473). A mensuração de desempenho, objeto de avaliação e monitoramento, é uma das cinco áreas-foco da governança de TI, classificada como *driver* (direcionador) para se alcançar os resultados (ITGI, 2007, p. 19).

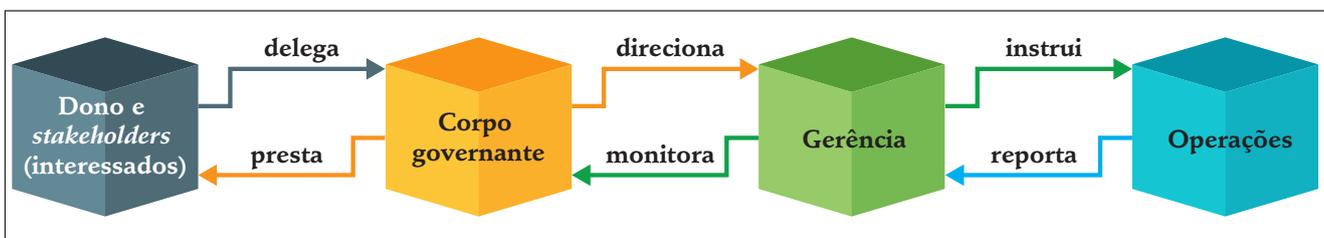
Para complementar a ilustração da aplicabilidade da nova métrica à governança de TI, o Quadro 6 ilustra alguns indicadores baseados em EF. Um mesmo indicador pode ser usado em perspectivas distintas do BSC-TI Desenvolvimento, dependendo da entidade alvo e do objetivo da medição, conforme exemplos que se seguem. A produtividade de um recurso (e.g. equipe, tecnologia) pode ser associada à perspectiva *Orientação futura*, enquanto se busca responder se a TI está preparada para as necessidades futuras. O mesmo indicador, se associado a um processo interno, codificação, por exemplo, traduz uma visão de sua capacidade produtiva, na perspectiva *Excelência operacional*. Na perspectiva *Orientação ao cliente*, a produção pode ser fracionada por área cliente beneficiada, demonstrando a proporção da produção da TI para cada uma das áreas de negócio. A avaliação da variação da produção da área de TI em contraste com a produção do negócio seria um exemplo do uso do indicador na perspectiva de *Contribuição para o negócio*.

A escolha dos indicadores procurou abarcar as cinco dimensões fundamentais citadas na seção 2.1.1: tamanho, esforço, prazo de entrega, qualidade e retrabalho. E agregou-se outra dimensão: o benefício esperado. Segundo Rubin (2003, p. 1), todo investimento em TI, da criação de um sistema a um simples treinamento, deve estar alinhado a uma prioridade de negócio cujo sucesso deve ser medido em termos de um valor específico²³. A dimensão de cada indicador figura na terceira coluna do Quadro 6.

Algumas medidas foram normalizadas ao ser divididas pelo número de elementos funcionais do produto ou processo, artifício usado para permitir a comparação entre projetos e sistemas de tamanhos diferentes. A habilidade de normalizar comparações, como em um BSC, é uma das características-chave de métricas de software (HUFSCHMIDT, 2002, p. 493). Trata-se de

Figura 3:

Papéis, atividades e relacionamentos da governança de TI



Fonte: Adaptado de ISACA (2012a, p. 24)p. 7)

Quadro 5:

Perspectivas de um BSC-TI

Perspectiva	Pergunta base	Perspectiva do BSC corporativo
Contribuição para o negócio da empresa	Como os executivos de negócio veem a área de TI?	Financeira
Orientação ao cliente	Como os usuários veem a área de TI?	Cliente
Excelência operacional	Quão efetivos e eficientes são os processos de TI?	Processos internos
Orientação futura	Como a TI está preparada para as futuras necessidades?	Aprendizado

Fonte: inspirado em ITGI (2007, p. 31)

Quadro 6:

Descrição de indicadores ilustrativos

Métrica	Unidade	Dimensão	Descrição do cálculo do indicador para um sistema
Tamanho funcional	EF	Tamanho	somatório do tamanho funcional das funcionalidades que compõem os sistemas ao final do período
Produção no período	EF	Esforço	somatório do tamanho funcional das solicitações de inclusão, alteração e exclusão implementadas no período
Produção em retrabalho *	EF	Retrabalho	somatório do tamanho funcional das solicitações de alteração e exclusão implementadas no período
Produtividade	EF / hora	Esforço	somatório do tamanho das solicitações atendidas no período / somatório dos esforços de todas as pessoas alocadas às atividades do sistema no período
Densidade de erros	Falhas / EF	Qualidade	quantitativo de falhas decorrentes do uso do sistema no período / tamanho do sistema ao final do período
Velocidade de entrega	EF / Hora	Prazo	somatório do tamanho das funcionalidades produzidas no período / tempo decorrido
Densidade do benefício esperado	R\$ / EF	Benefício	benefício esperado pelo sistema no período / tamanho do sistema

* A taxa de retrabalho pode ser calculada pela razão entre a produção em retrabalho e a produção no período.

uma prática similar a normalizar métricas de construção com base no metro quadrado, uma prática comum (DEKKERS, 2002, p. 161).

Como argumenta Dennis (2002, p. 302), não se deve tomar decisões baseadas em um único indicador, mas a partir de uma visão formada por vários indicadores complementares. À medida que a TI tem assumido maior proeminência como facilitadora da realização da estratégia de negócio, o uso de painéis de controle gráficos (*dashboards*) para monitorar o seu desempenho, sob apropriados critérios, tem se tornado popular entre os gerentes de empresa (ISACA 2010, p. 74). Fernandes e Abreu (2009, p. 167) propõem alguns tópicos que podem compor painéis estratégicos e táticos de controle da TI.

O Gráfico 3 ilustra com dados fictícios o comportamento das métricas indicadas no Quadro 6 para instâncias de sistemas, com apuração anual. O traço contínuo vertical indica como estava a métrica para o sistema no período anterior, permitindo uma visualização da proporção do acréscimo ou decréscimo

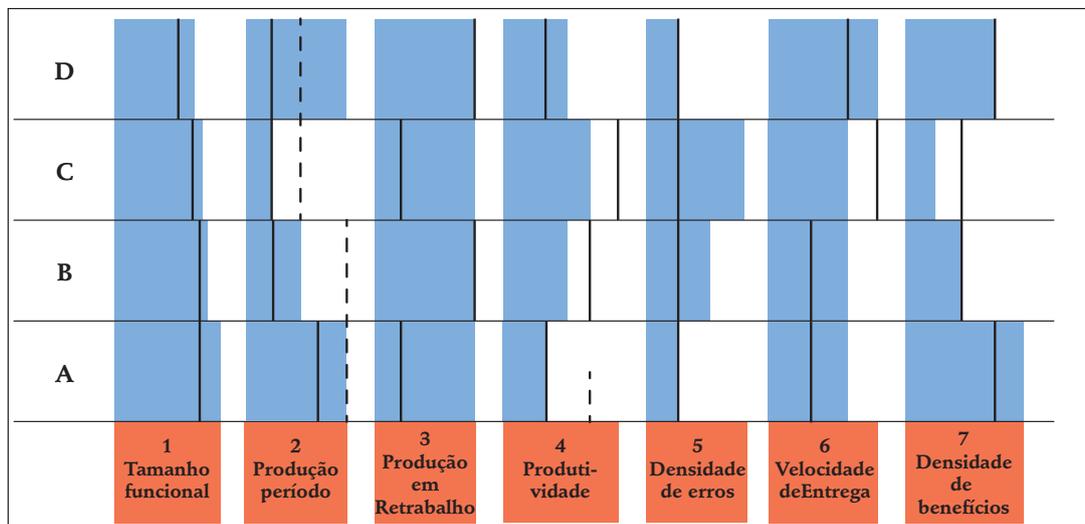
dos valores no período. Na coluna da produtividade (coluna 4) percebe-se uma linha curta em sua base que indica, por exemplo, um valor padrão obtido por *benchmarking*. A linha tracejada vertical associada à métrica *produção período* (2) indica a meta traçada para o período em cada sistema: alcançada para o sistema A, ultrapassada para o sistema D e não atingida para os sistemas B e C.

Em uma análise superficial ilustrativa dos dados para o sistema C, pode-se associar a causa do não alcance da meta da produção no período (2) ao decréscimo da velocidade de entrega (6) e ao aumento da produção em retrabalho (3), resultado provável do crescimento na densidade de erros (5). A diminuição da velocidade de entrega (6), que pode estar associada à diminuição da produtividade (4), levou a um baixo crescimento do sistema no período (1). Esses resultados negativos levaram a uma diminuição na densidade do benefício esperado (7).

O Gráfico 3 traduz uma opção de visualização dos indicadores de governança indicados no Quadro 6:

Gráfico 3:

Indicadores anuais dos Sistemas A, B, C e D



um gráfico multimétricas de multi-instâncias de uma mesma entidade ou atributo alvo. A largura da coluna vertical é variável conforme os valores das métricas (eixo horizontal) associados às instâncias diferentes de entidades, ou atributos, de interesse (eixo vertical)²⁴. Reserva-se o mesmo espaço vertical para cada instância da entidade. A largura da área colorida, delineada e valorada da esquerda para a direita, indica graficamente o valor do indicador para a instância.

Nas mãos do comitê de governança, os indicadores certos podem ajudar a alta administração, diretamente ou por intermédio de alguma estrutura de governança, a identificar como a gestão de TI está se comportando e a identificar os problemas e as providências devidas, quando necessárias.

3. CONSIDERAÇÕES FINAIS

Os cinco objetivos específicos propostos para este trabalho na seção 1.4 foram alcançados, ainda que com limitações e possibilidades de melhoria que se traduzem em propostas de trabalhos futuros.

O principal resultado foi a proposição da uma nova métrica EF e de sua submétrica EFt. As novas métricas, livres de algumas deficiências da APF, métrica tomada por base para sua derivação, alcançaram uma correlação superior com o esforço do que a APF, no contexto dos dados analisados.

Também foi ilustrada a conexão entre métricas e atividades de governança de TI, seja na avaliação e monitoramento, pelo seu uso em *dashboards*, seja no direcionamento dos objetivos, pelo seu uso em BSC-TI.

Há possibilidades de trabalhos futuros em relação a cada um dos objetivos específicos.

Quanto à conceituação e categorização de métricas de software, necessita-se de um trabalho mais detalhado e abrangente de pesquisa bibliográfica com a produção de uma categorização mais ampla e atualizada de métricas de software.

Quanto às críticas à APF, apenas foram apresentadas aquelas tratadas pela nova métrica proposta. Um trabalho de pesquisa mais amplo no tema serviria inclusive para fundamentar outras proposições de métricas de software.

Quanto ao processo de criação da nova métrica, poderia ser aprimorado em novos trabalhos ou poderia ser aplicado a outras métricas de qualquer área do conhecimento que, à semelhança da APF, se basearem em valores ordinais derivados de tabelas de complexidade (e.g. métrica proposta por Karner (1993): Pontos de Caso de Uso). Trabalhos futuros poderiam propor e avaliar alterações nas regras e no escopo da EF. O processo de criação poderia ser aprimorado, por exemplo, com tratamentos diferentes para as faixas ilimitadas (seção 2.3.2). Pesos poderiam ser atribuídos aos tamanhos das faixas limitadas, por exemplo, em função da proporção de funções integrantes dessas faixas em uma amostra com funcionalidades de vários sistemas.

Quanto à avaliação da nova métrica, a limitação do uso de dados de apenas uma organização poderia ser superada em novos trabalhos. Aplicações práticas da métrica poderiam ser ilustradas, por exemplo, em contratações com processo de entrega incremental. Novos trabalhos poderiam comparar os resultados da EF com a submétrica EFt, bem como comparar as duas com outras métricas de software. Modelos estatísticos diferentes poderiam ser usados para avaliação de sua correlação com o esforço inclusive em contextos espe-

cíficos (e.g. desenvolvimento, manutenção, plataformas de desenvolvimento). Espera-se uma maior correlação das novas métricas com o esforço de desenvolvimento com métodos ágeis em comparação com a APF, por sua capacidade de dimensionar alterações em partes de uma funcionalidade (6ª crítica da seção 2.2).

Quanto à conexão com a governança de TI, mostra-se promissor um trabalho mais detalhado do uso de métricas como apoio para a governança de TI em todas as suas atividades. O gráfico proposto²⁵ para visualização de múltiplos indicadores de múltiplas instâncias em colunas com larguras variáveis em sua extensão também poderia ser padronizado e melhorado em trabalhos futuros.

Destaca-se uma sugestão de trabalho futuro: a definição de um indicador que sinalize o nível de maturidade de uma empresa em relação ao uso de métricas para a governança de TI. Entre outros aspectos que podem ser considerados na composição desse indicador, destacam-se: a abrangência das entidades avaliadas (e.g. sistemas, projetos, processos, equipes), as dimensões consideradas (e.g. tamanho, retrabalho, qualidade, benefício) e o uso dos indicadores (e.g. monitoramento, direcionamento).

Espera-se, por fim, que a métrica EF e sua submétrica EFt permitam aumentar a contribuição da TI para o negócio das empresas de forma objetiva, confiável e visível.

NOTAS

- 1 Uma versão deste trabalho, apenas em inglês, não contemplando todo o conteúdo aqui presente, foi apresentada no XXVII SBES (Simpósio Brasileiro de Engenharia de Software) promovido pela SBC (Sociedade Brasileira de Computação) e foi publicada na base digital IEEE Xplore: M. V. B. D. Castro and C. A. M. Hernandez, "A Metric of Software Size as a Tool for IT Governance", *Software Engineering (SBES), 2013 27th Brazilian Symposium on*, Brasília, 2013, pp. 99-108. doi: 10.1109/SBES.2013.13.
- 2 Em sua versão, 4.3.1 (IFPUG, 2010), no apêndice C, há a possibilidade de se ajustar o tamanho funcional com um fator que reflete uma avaliação do sistema em relação a 14 características gerais não funcionais. Segundo Fenton e Pfleeger (1998, p. 262), a determinação é subjetiva e, segundo Kemerer (1987, p. 9), o ajuste não aumenta a correlação da métrica com o esforço. Essa parte foi apartada da regra padrão dos pontos de função, até porque a APF é um padrão ISO de métrica funcional somente sem a aplicação do ajuste.
- 3 Há vários acórdãos sobre o assunto: 1.782/2007, 1.910/2007, 2.024/2007, 1.125/2009, 1.784/2009, 2.348/2009, 1.274/2010, 1.647/2010, todos do Plenário do TCU.
- 4 Também conhecido como CIO (*Chief Information Officer*).
- 5 Kitchenham et al. (1995) apresentam um *framework* para métrica de software em que são colocados os conceitos associados ao modelo formal em que uma métrica se baseia (e.g. tipo de escala usado).
- 6 A visão geral apresentada resulta da experiência do autor Marcus com a APF. Em 1993 ele coordenou a implantação do uso da APF na área de desenvolvimento de sistemas do Tribunal Superior do Trabalho. No Tribunal de Contas da União, também atua na área de métricas.
- 7 Requisitos funcionais são apenas uma dimensão de várias que impactam o esforço. Todas devem ser levadas em consideração em estimativas. Requisitos não funcionais (e.g. experiência da equipe, tecnologia) e estimativas não são objetivo deste trabalho.
- 8 Funcionalidades de complexidade muito baixa também não são adequadamente dimensionadas pela APF, pois assumem o valor mínimo quando deveriam assumir um valor ainda menor.
- 9 Por serem conceitos amplamente conhecidos pelos medidores, espera-se que a nova métrica possa ter aceitação entre os profissionais da área.
- 10 Esses atributos correspondem neste trabalho ao conceito de elementos funcionais, nome da métrica proposta.
- 11 A alternativa de se atribuir à terceira faixa a soma dos elementos das duas primeiras faixas também foi avaliada. Contudo, esta abordagem se mostrou menos eficiente na correlação com o esforço, para os dados avaliados.
- 12 Optou-se por distinguir a métrica EFt para aplicação em casos onde o esforço de tratamento das estruturas de dados (Efd) não é objeto de avaliação ou de contratação. Embora não tenha sido avaliada, a submétrica Efd tem o seu papel enquanto traduz a complexidade estrutural dos dados de uma aplicação.
- 13 Kemerer (1987, p. 421) justificou o uso de regressão linear como um meio para se avaliar a correlação da métrica APF com o esforço.

- 14 O órgão fornecedor dos dados informou que cada sistema está implementado em somente uma linguagem: Java, DotNet ou Natural.
- 15 A ordem dos sistemas segue o critério do quantitativo de OS.
- 16 Também foi realizada uma regressão não linear logística, com constante, usando-se o software Gretl, ferramenta livre de código aberto (<http://gretl.sourceforge.net>). Contudo, o R2 da correlação não linear se mostrou bem inferior em relação à regressão linear para todas as métricas e por isso o modelo não linear de correlação foi descartado. O R2 da regressão não linear alcançou os seguintes valores: (sistema, R2_APF, R2_PM, R2_EF, R2_EFt) – (H, 0,316, 0,470, 0,434, 0,426); (B, 0,013, 0,313, 0,442, 0,443); (C, 0,327, 0,262, 0,16, 0,152) e (D, 0,02, 0,127, 0,087, 0,087).
- 17 Ou seja, a reta passa pela origem dos eixos.
- 18 Justifica-se essa restrição, por exemplo, com a informação passada pelo órgão fornecedor dos dados de que a linguagem de desenvolvimento é única por sistema e que a equipe técnica é, regra geral, também a mesma por sistema. A linguagem e a equipe são fatores que influenciam o esforço. Os fatores que influenciam o esforço e o grau dessa correlação foram discutidos em diversos artigos. Para mais detalhes sobre o tema, orienta-se o acesso aos artigos da base BestWeb (<http://www.simula.no/BESTweb>), criada como resultado da pesquisa de Jørgensen e Shepperd (2007).
- 19 Para se considerar uma correlação estatisticamente significativa em um nível de confiança de X%, o *p-value* deve ser inferior a $1 - X$ (ORLOV, 1996, p. 11). Para um nível de 95%, o *p-value* precisa ser inferior a 0,05.
- 20 Critério usado para se considerar uma correlação C1 superior à correlação C2: C1 ser significativa e C2 não ser ou, se ambas forem significantes, C1 possuir um maior R2 do que C2.
- 21 Percebe-se uma maior correlação da métrica EFt em relação à EF no sistema H, único sistema que permitiu uma diferenciação do resultado das duas métricas por apresentar comandos relativos à alteração de arquivos lógicos em suas OS. Percebe-se assim uma tendência favorável à submétrica EFt em relação à EF, reforçando-se a hipótese de que a submétrica EFd que compõe a métrica EF não impacta o esforço de codificação e teste, tarefas tratadas nas OS avaliadas.
- 22 Uma comparação entre a correlação da métrica EF e a correlação da métrica PM (Nesma) não era objetivo do

trabalho. Porém, como os dados traziam também as medidas em PM, métrica usada no contrato do fornecedor dos dados, fez-se também uma avaliação da métrica PM.

Sistema		H	B	C	D
PM	R2	67,0%	48,1%	63,7%	60,4%
	<i>p-value (teste-f)</i>	4,7E-12	6,7E-05	1,1E-05	4,8E-04
	comparação R2 PF	+13%	+329%	-6%	+17%

Observações:

1. todas as correlações foram consideradas significantes a um nível de confiança de 95%;
 2. como as novas métricas, PM alcançou correlações médias ($0,8 > R\text{-quadrado} > 0,5$) para os quatro sistemas e um resultado superior à APF nos mesmos três sistemas (H, B e D);
 3. no sistema H houve uma equivalência da correlação da PM com a EFt, ficando a diferença entre as duas correlações em 0,9%.
 4. as correlações da PM foram superiores em 2 sistemas em relação às novas métrica (C e D) e inferior em um sistema (B).
 5. no contexto dos dados avaliados, os números mostram uma tênue superioridade da PM em relação às novas métricas. Vislumbra-se, contudo, um potencial de melhores resultados para a EF, uma vez que a PM contém as cinco primeiras falhas citadas na seção 2.2.
 6. a PM apresenta outra falha conceitual: diferentemente da APF e da EF, ela dimensiona diferentemente software em desenvolvimento de software em manutenção, com fatores de ajustes diferenciados. O custo pode realmente ser diferente para manutenção em relação ao desenvolvimento. Como também pode ser diferente por outros fatores (e.g. linguagem usada, metodologia aplicada). Fatores que não alteram o tamanho do software, mas sim o custo. Trazer a preocupação com o custo para o tamanho do software, como faz a PM, parece não ser a melhor opção, pois são conceitos diferentes.
 7. outro ponto desfavorável à PM em relação às métricas propostas é que ela requer a contagem do tamanho da funcionalidade antes da manutenção, esforço extra não necessário para as novas métricas.
- 23 Não faz parte do escopo do trabalho investigar os conceitos e os processos associados à determinação do valor de uma

funcionalidade ou de um sistema ou da área de TI. Trata-se de um tema complexo e ainda imaturo.

- 24 O valor máximo de cada indicador no período foi associado com a largura máxima definida para a coluna. As larguras das áreas coloridas dos demais sistemas foram derivadas por regra de três simples.
- 25 Em <<http://learn.wordpress.com>> (acesso em 04 nov. 2012) encontra-se um gráfico que funcionalmente se assemelha ao proposto, *heatmap plotting*, porém distinto na formatação e nas possibilidades de evolução. Por não se ter encontrado gráfico semelhante, fica a presunção de ser um novo formato para visualização do comportamento de múltiplos indicadores em múltiplas instâncias em colunas com larguras variáveis em sua extensão (MIMICoVaWE – *Multiple Indicators about Multiple Instances through Columns with Varying Widths along their Extension*). Dois exemplos dentre as diversas evoluções possíveis para o gráfico: troca de posição entre as métricas e as instâncias, passando estas para o eixo horizontal; variação na tonalidade da cor da célula que relaciona uma métrica e uma instância conforme algum critério (e.g. em relação ao atingimento da meta especificada)

REFERÊNCIAS

- ABRAN, A.; ROBILLARD P. N. Function points: a study of their measurement processes and scale transformations. *Journal of Systems and Software* 25(2): 171-184. 1994.
- ALBRECHT, A. J. (1979). "Measuring application development productivity." *Proceedings of the Joint SHARE/GUIDE/IBM Application Development Symposium*, SHARE Inc. and GUIDE International Corp. Monterey, CA.
- ALBRECHT, A. J.; GAFFNEY JR. J. E. Software function, source lines of code, and development effort prediction: a software science validation. *Software Engineering, IEEE Transactions on*(6): 639-648. 1983.
- ANDRADE, M. M. Como preparar trabalhos para cursos de pós-graduação: noções práticas, Atlas. 2002.
- ANTONIOL, G. et al. Object-oriented function points: An empirical validation. *Empirical Software Engineering* 8(3): 225-254. 2003.
- BASIL, V. R.; WEISS D. M. A methodology for collecting valid software engineering data. *Software Engineering, IEEE Transactions on*(6): 728-738. 1984.
- BECKER, S. A. ; BOSTELMAN M. L. Aligning strategic and project measurement systems. *Software, IEEE* 16(3): 46-51. 1999.
- BEYERS, C. P. Estimating software development projects. In: IFPUG. *IT Measurement: Practical Advice from the Experts*. Addison-Wesley Indianapolis. p. 337-362. 2002.
- BRASIL. MCT - Ministério da Ciência e Tecnologia. Pesquisa de Qualidade no Setor de Software Brasileiro 2009. Brasília: Ministério da Ciência e Tecnologia. Secretaria da Política de Informática. 204p. Disponível em: <http://www.mct.gov.br/upd_blob/0214/214567.pdf>. Acesso em: 16 set. 2012. 2009.
- BRASIL. SISP - Sistema de Administração dos Recursos de Tecnologia da Informação. Roteiro de Métricas de Software do SISP. Versão 2.0 Brasília: Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. Disponível em: <http://www.sisp.gov.br/ct-gcie/download/file/Roteiro_de_Metricas_de_Software_do_SISP_-_v2.0.pdf>. Acesso em: 31 out. 2012. 2012.
- BUNDSCHUH, M.; DEKKERS C. The IT measurement compendium: estimating and benchmarking success with functional size measurement. Springer. 2008.
- CHIKOFFSKY, E. e RUBIN H. A. Using metrics to justify investment in IT. *IT professional* 1(2): 75-77. 1999.
- DEKKERS, C. A. How and when can functional size fit with a measurement program? In: IFPUG. *IT Measurement: Practical Advice from the Experts*. Addison-Wesley Indianapolis. p. 161-170. 2002.
- DENNIS, S. P. Avoiding obstacles and common pitfalls in the building of an effective metrics program. In: IFPUG. *IT Measurement: Practical Advice from the Experts*. Addison-Wesley Indianapolis. p. 295-304. 2002.
- FENTON, N. E.; PFLEEGER S. L. *Software metrics: a rigorous and practical approach*. PWS Publishing Co. 1998.
- FERNANDES, A. A.; ABREU V. F. *Implantando a governança de TI: da estratégia à gestão de processos e serviços*. Brasport. 2009.
- GENCEL, C.; DEMIRORS, O. Functional size measurement revisited. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 17(3): 15. 2008.

HUFSCHMIDT, B. Software balanced scorecards: the icing on the cake. In: IFPUG (2002). IT Measurement: Practical Advice from the Experts. Addison-Wesley Indianapolis. p. 491-502. 2002.

IFPUG - International Function Point Users Group. Manual de Práticas de Contagem de Pontos de Função, Versão 4.3.1 IFPUG. 2010.

ISACA - Information Systems Audit and Control Association. CGEIT Review Manual 2010. ISACA. 2010.

_____. COBIT 5: A Business Framework for the Governance and Management of IT. ISACA. 2012a.

_____. COBIT 5: Enabling Processes. Isaca. 2012b.

ISO - International Organization for Standardization; IEC - International Electrotechnical Commission. 14143-4: Information Technology - Software Measurement - Functional Size Measurement - Part 4: Reference Model. 2002a.

_____. 20968: MkII Function Point Analysis - Counting Practices Manual. 2002b.

_____. 14143-3: Information Technology - Software Measurement - Functional

Size Measurement - Part 3: Verification of Functional Size Measurement Methods. 2003.

_____. 14143-5: Information Technology - Software Measurement - Functional

Size Measurement - Part 5: Determination of Functional Domains for Use with Functional Size Measurement. 2004.

_____. 24570: NESMA functional size measurement method version 2.1 -- Definitions and counting guidelines for the application of Function Point Analysis. 2005.

_____. 14143-6: Guide for the Use of ISO/IEC 14143 Series and Related International Standards. 2006.

_____. 14143-1: Information Technology - Software Measurement - Functional Size Measurement - Part 1: Definition of Concepts. 2007.

_____. 38500: Corporate governance of information technology. 2008.

_____. 20926: Software measurement - IFPUG functional size measurement method. 2009.

_____. 29881: FISMA 1.1 functional size measurement method. 2010.

_____. 14143-2: Information Technology - Software Measurement - Functional Size Measurement - Part 2: Conformity Evaluation of Software Size Measurement Methods to ISO/IEC 14143-1. 2011a.

_____. 19761: COSMIC: a functional size measurement method. 2011b.

ITGI. IT Governance Institute. Board briefing on IT Governance. 2nd Ed. Isaca. 2007.

JORGENSEN, M.; SHEPPERD, M. A systematic review of software development cost estimation studies. Software Engineering, IEEE Transactions on 33(1): 33-53. 2007.

KAPLAN, R. S.; NORTON, D. P. The balanced scorecard - measures that drive performance. Harvard business review 70(1): 71-79. 1992.

KAPLAN, R. S.; NORTON D. P. Using the balanced scorecard as a strategic management system. Harvard business review 74(1): 75-85. 1996.

KARNER. G. Metrics for Objectory. Diploma thesis at the university of Linköping, Sweden. No. LiTH-IDA-Ex-9344; 21. December 1993.

KEMERER, C. F. An empirical validation of software cost estimation models. Communications of the ACM 30(5): 416-429. 1987.

KITCHENHAM, B. et al. Towards a framework for software measurement validation. Software Engineering, IEEE Transactions on 21(12): 929-944. 1995.

KITCHENHAM, B. The problem with function points. Software, IEEE 14(2): 29-31. 1997.

KITCHENHAM, B.; KANSALA, K. Inter-item correlations among function points, Software Engineering, 1993. Proceedings, 15th International Conference on, Baltimore, MD, pp. 477-480, 1993.

KRALJ, T. et al. Improved standard FPA method - resolving problems with upper boundaries in the rating complexity process. Journal of Systems and Software 77(2): 81-90. 2005.

MOSER, S. Measurement and estimation of software and software processes, PhD Thesis, University of Berne, Switzerland. 1996.

NESMA. Netherlands Software Metrics Association. Function Point Analysis For Software Enhancement. Disponível em: <[http://www.nesma.nl/download/boeken_NESMA/N13_FPA_for_Software_Enhancement_\(v2.2.1\).pdf](http://www.nesma.nl/download/boeken_NESMA/N13_FPA_for_Software_Enhancement_(v2.2.1).pdf)>. Acesso em: 16 out. 2012. 2009.

ORLOV, M. L. Multiple Linear Regression Analysis Using Microsoft Excel. Chemistry Department, Oregon State University. 1996.

PFLEEGER, S. L. et al. Status report on software measurement. Software, IEEE 14(2): 33-43. 1997.

ROHM, H.; MALINOSKI, M. Strategy-Based Balanced Scorecards for Technology. Balanced Scorecard Institute. 2010.

RUBIN, H. A. Software process maturity: measuring its impact on productivity and quality. In: Proceedings of the 15th international conference on Software Engineering, IEEE Computer Society Press: 468-476. 1993.

_____. How to Measure IT Value. CIO insight. 2003.

SARTORIS, A. Estatística e introdução à econometria. 2008.

TURETKEN, O. et al. The Effect of Entity Generalization on Software Functional Sizing: A Case Study. Product-Focused Software Process Improvement: 105-116. 2008.

VAN GREMBERGEN, W.; VAN BRUGGEN R. Measuring and improving corporate information technology through the balanced scorecard. The Electronic Journal of Information Systems Evaluation 1(1). 1997.

VAN GREMBERGEN, W. The balanced scorecard and IT governance.

Information Systems Control Journal, Vol 2, 40-43, 2000.

VAZQUEZ, C. E. et al. Análise de Pontos de Função: Medição, Estimativas e Gerenciamento de Projetos de Software. Editora Érica, São Paulo. 2005.

XIA, W. et al. Updating weight values for function point counting. International Journal of Hybrid Intelligent Systems 6(1): 1-14. 2009.

WEILL, P.; ROSS, J. W. Governança de TI, tecnologia da informação. São Paulo: M. Books do Brasil. 2006.

