

INSTITUTO FEDERAL DO PARANÁ

TALITA FATIMA PAULA MONTEIRO

**DESENVOLVIMENTO DE BUSINESS INTELLIGENCE PARA O PERFIL DA
PRODUÇÃO CIENTÍFICA DO IFPR CAMPUS PALMAS**

PALMAS – PR

2017

TALITA FATIMA PAULA MONTEIRO

**DESENVOLVIMENTO DE BUSINESS INTELLIGENCE PARA O PERFIL DA
PRODUÇÃO CIENTÍFICA DO IFPR CAMPUS PALMAS**

Trabalho de Conclusão de Curso apresentado
ao Curso Superior Bacharelado em Sistemas de
Informação do Instituto Federal do Paraná,
como requisito parcial de avaliação.

Orientadora: Lauriana Paludo

PALMAS – PR

2017

ATA DA BANCA DE AVALIAÇÃO DO TRABALHO DE CONCLUSÃO

Instituto Federal do Paraná (IFPR) - Campus Palmas – PR – Brasil
Bacharelado em Sistemas de Informação

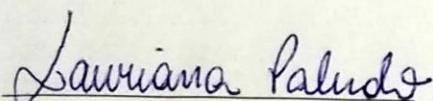
Ata da 1^a sessão de defesa de Trabalho de Conclusão de Curso (TCC) do ano de 2017.

Aos 29 dias do mês de novembro de 2017 em sessão pública, reuniu-se a Banca Examinadora constituída pelos professores: **Lauriana Paludo** (presidente/orientador), e os membros: **Vagner Scamati** e **Débora R M Lima Reis**, na sala F26 Bloco F, do IFPR Campus Palmas, para avaliar o Trabalho de Conclusão de Curso (Monografia) intitulado: “**Desenvolvimento de Business Intelligence para o Perfil da Produção Científica do IFPR Campus Palmas (PR)**”, sendo apresentado pelo (s/a) aluno (s/a) **Talita Fatima Paula Monteiro**.

Iniciados os trabalhos, o (a) orientador (a) do trabalho deu conhecimento aos membros da banca e ao (a) aluno (a) das normas que regem a apresentação. A seguir, o (a) aluno (a) passou à apresentação de seu trabalho. As notas apontadas pelos professores conforme as Fichas de Avaliação da Banca indicam que o aluno está Aprovada sendo recomendada nesta ocasião a concessão do grau de “**Bacharel em Sistemas de Informação**” ao candidato, uma vez atendidas as correções apontadas pela banca e o cumprimento integral do currículo do curso. Encerrada a apresentação, procedeu-se a arguição da banca. Nada mais havendo a tratar, lavrou-se a presente ata que foi lida na presença dos membros da banca e do aluno que foi convidado a retornar a sala, que vai assinada pelos membros da banca examinadora, pelo Coordenador do Trabalho de Conclusão e pelo aluno orientando.

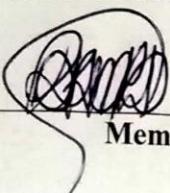
Observações:

- 1) A média aritmética final para aprovação é 7,0 (sete);
- 2) O(a) aluno(a) terá prazo de até 30 (trinta) dias para fazer as correções solicitadas pela banca e apresentá-las ao Presidente da Banca (professor orientador) com a finalidade de entrega definitiva do trabalho (capa dura).



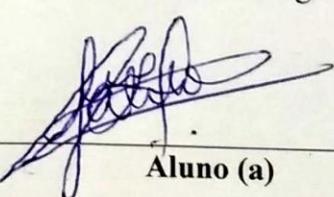
Lauriana Paludo

Professor orientador
Presidente da Banca



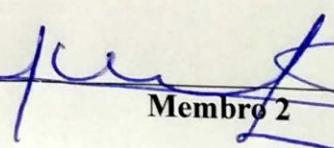
Ronaldo

Membro 1



Jessica

Aluno (a)



Paulo

Membro 2



Paulo

Coordenador(a) de Curso

Dedico este trabalho aos meus pais AMILTON JOSÉ MONTEIRO e DIOMAR DE CÁSSIA MOURA MONTEIRO por todo apoio e todo amor dedicado a mim, não apenas no período dessa graduação mas em toda a minha existência.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais Amilton e Diomar por me apoiarem no momento que decidi sair da minha zona de conforto e ir cursar o ensino superior em uma cidade à 500km de distância de casa, por não questionarem minhas decisões, aturarem meu mau humor e me amarem acima de tudo e sob qualquer circunstância. Aos meus irmãos Rodrigo, Milena, Vinicius e Lavínia pelo apoio e compreensão nesse período de ausências. A minha “boadrasta”, Profª Dra. Rosineide da Paz, pela disponibilidade, apoio e incentivo de sempre.

A minha orientadora, Profª. Me. Lauriana Paludo, por todo apoio e dedicação destinado, não somente neste trabalho, mas em todo período da minha formação. Pela parceria e pelo respeito aos alunos dentro e fora das salas de aula e, acima de tudo, por ter me apresentado o caminho da pesquisa e me feito enxergar que o aprendizado vai além das salas de aulas. Não existem formas de mensurar sua importância nessa graduação e em todos os momentos que pude contar com sua amizade durante esse período. Que a vida lhe retribua em dobro e lhe seja generosa.

Aos colegas e amigos que fiz através do Movimento Estudantil do IFPR, da União Nacional dos Estudantes – UNE, da União Paranaense dos Estudantes – UPE e da União dos Jovens Socialistas – UJS, dos quais pude participar ativamente durante esses quatro anos, lutando e resistindo. Aos amigos que fiz durante a graduação que, apesar de serem em pouco número, foram gigantes em sua importância, em especial ao casal mais parceiro do IFPR, Carla e Fernando e a querida companheira de trabalhos, seminários e afins, Patrícia. Vocês foram incríveis, serei sempre grata por terem me acolhido no grupo e me tratado sempre tão bem!

Aos poucos colegas que fiz em Palmas, que durante a graduação me presentearam com sua companhia, fazendo desses quatro anos um período menos tenso e solitário. E por fim, mas não menos importante, aos meus verdadeiros amigos de Maringá/PR que transcendem a distância e se fizeram presentes nos momentos em que tanto precisei.

RESUMO

Esse trabalho apresenta uma solução de *Business Intelligence (BI)* que tem como objetivo identificar o perfil da produção científica do IFPR campus Palmas, através de recursos como *Extraction, Transform and Load (ETL)*, *Data Warehouse (DW)* e *Online Analytical Processing (OLAP)*. Para tanto, utiliza tecnologias *open source* da suíte de aplicativos *Pentaho* (*Pentaho Data Integration, Mondrian Scheme Workbench e Pentaho BI Server*) em conjunto com o banco de dados *PostgreSQL*. O perfil é traçado a partir de indicadores bibliográficos extraídos dos currículos Lattes e da base de dados do Comitê de Pesquisa e Extensão dos pesquisadores do campus, tais como, tipologia da produção bibliográfica, artística/cultural e produção técnica. O BI desenvolvido como resultado deste estudo pode ser utilizado para subsidiar a comunidade acadêmica e da região com informações sobre as publicações e projetos desenvolvidos pelos docentes do campus, a compreender melhor a atuação dos mesmos, em quais áreas de conhecimento se concentram as produções científicas, bem como apoiar o processo de tomada de decisão da gestão do campus na aplicação de recursos e oportunidades de continuidade de estudos para a comunidade e egressos dos cursos ofertados, entre outras aplicações.

Palavras-Chave: Business Intelligence, Data Warehouse, Pentaho, Produção Científica IFPR campus Palmas.

ABSTRACT

This paper presents a Business Intelligence (BI) solution that aims to identify the profile of scientific production of the IFPR campus Palmas, through resources such as Extraction, Transformation and Load (ETL), Data Warehouse (DW) and Online Analytical Processing (OLAP). To do so, use Opensource technologies of the Pentaho application suite (Pentaho Data Integration, Mondrian Scheme Workbench and Pentaho BI Server) in conjunction with the PostgreSQL database. The profile is drawn from bibliographic indicators extracted from curriculum Lattes and the Research and Extension Committee's database of campus researchers, such as bibliographic, artistic/cultural and technical production typology. The BI was developed as a result of the study can be used to subsidize an academic community and the region with information on publications and projects developed by campus teachers, improved upgrades in which areas of knowledge are concentrated as scientific productions as well as support the campus management decision-making process in applying resources and continuing education opportunities to a community and graduates of business courses, among other applications.

Keywords: Business Intelligence, Data Warehouse, Pentaho, Scientific Production IFPR campus Palmas.

LISTA DE IMAGENS

FIGURA 1- Arquitetura do BI.....	21
FIGURA 2 - Comparativo entre sistemas OLTP e Data Warehouse.	23
FIGURA 3 - Arquitetura do Data Warehouse	24
FIGURA 4 – Modelagem Dimensional	25
FIGURA 5 - Arquitetura da ferramenta Pentaho Data Integration.	35
FIGURA 6 - Metodologia da Pesquisa	39
FIGURA 7 - Modelo Entidade-Relacionamento do DW da Produção Científica.	41
FIGURA 8 - Modelagem Dimensional do DW através do esquema estrela.	42
FIGURA 9 - Arquitetura do Business Intelligence da Produção Científica do IFPR campus Palmas.	43
FIGURA 10 - Áreas e Subáreas do Conhecimento de acordo com a CAPES.	43
FIGURA 11 - Processo de extração de dados e criação das tabelas de área e subáreas de conhecimento da CAPES.	44
FIGURA 12 - Parâmetros de configuração para mapear identificação das áreas e subáreas do conhecimento, através do step Value Mapper.	45
FIGURA 13 - Parâmetros de configuração do step table output que persiste os dados extraídos da planilha eletrônica em uma tabela na base de dados.	45
FIGURA 14 - Parâmetros de configuração do step table output responsável por criar a tabela de subáreas de conhecimento.	46
FIGURA 15 - Script SQL utilizado para criação e preenchimento da tabela de colegiado na base de dados.	47
FIGURA 16 - Planilha eletrônica contendo o resultado da Busca Pessoas utilizando a plataforma Stela Expertia.	47
FIGURA 17 - Extração, transformação e carga dos dados para a criação da tabela de pesquisadores na base de dados.	48
FIGURA 18 - Extração, transformação e carga dos dados para a criação e inserção de dados na tabela de cidades.	48
FIGURA 19 - Extração, transformação e carga dos eventos para a staging area.	50
FIGURA 20 - Processo de ETL para substituir o nome pela identificação de cada evento na planilha de publicações.	51
FIGURA 21 - Processo de ETL para substituição do nome pelo número de identificação de cada cidade.	52
FIGURA 22 - Processo de ETL para substituição do nome pelo número de identificação de cada pesquisador, criação e inserção de dados na tabela publicação da staging area.	53
FIGURA 23 - Configurando do step Table Output para persistência dados da planilha de publicações para a staging area.	54
FIGURA 24 - Processo de ETL para substituição do nome pelo número de identificação de cada pesquisador, criação e inserção de dados na tabela projeto da staging area.	56
FIGURA 25 - Modelo Entidade-Relacionamento da Staging Area.	57
FIGURA 26 - Criação das tabelas de dimensão e inserção de dados através de SQL.	58
FIGURA 27 - Transformação que realiza a carga de dados da planilha de publicação para a tabela publicação da staging area.	59
FIGURA 28 - Configuração do step Table Input com o select utilizado para listar informações das tabelas pesquisador, colegiado e subáreas.	60
FIGURA 29 - Processo de ETL para criação da tabela fato de projetos.	61
FIGURA 30 - Resultado do SQL de consultas aplicado na tabela fato projetos do DW.	61
FIGURA 31 - Tela inicial da ferramenta Schema Workbench.	62
FIGURA 32 - Configurando a conexão do Schema Workbench com o DW.	63

FIGURA 33 - Adicionando um novo cubo no esquema.....	64
FIGURA 34 - Definindo os parâmetros de configuração da tabela fato no cubo de Produção Científica.....	64
FIGURA 35 - Criação da medida quantidade publicação no cubo Produção Científica.....	65
FIGURA 36 - Criando uma medida calculada para consultar o percentual da Produção Científica.....	65
FIGURA 37 - Configurando a dimensão em conformidade Tempo.	66
FIGURA 38 - Criando e configurando uma hierarquia na dimensão tempo.	67
FIGURA 39 - Mapeando a tabela do DW que compõe a dimensão tempo do cubo.	67
FIGURA 40 - Configuração dos parâmetros para criar o nível ano na dimensão tempo.	68
FIGURA 41 - Composição da dimensão em conformidade Tempo	68
FIGURA 42 - Composição das dimensões Tempo, Pesquisador, Colegiado, Área de Conhecimento e Subáreas de Conhecimento criadas em conformidade com o esquema.	69
FIGURA 43 - Referenciando a dimensão em conformidade Tempo dentro do cubo da Produção Científica.....	69
FIGURA 44 - Cubo Produção Científica com todos os componentes mapeados (fato, medidas e dimensões).	70
FIGURA 45 - Cubo Projetos com todos os componentes mapeados (fato, medidas e dimensões).	71
FIGURA 46 - Confirmação da conexão entre o esquema e o DW.....	72
FIGURA 47 - Consultas MDX realizadas nos cubos Produção Científica e Projetos.	72
FIGURA 48 - Inicialização do Servidor Pentaho	73
FIGURA 49 - Painel de Acesso do usuário.	74
FIGURA 50 - Local de extração do plugin Saiku Alalitycs.	75
FIGURA 51 - Criando consultas no esquema OLAP através do Saiku Analitycs.	75
FIGURA 52 - Consulta da quantidade de publicações de colegiado por ano.	76
FIGURA 53 - Consulta da quantidade de publicações de colegiado por ano representada em um gráfico de barras.	76
FIGURA 54 - Gráfico Boxplot da Produção Científica dos docentes do IFPR campus Palmas	88
FIGURA 55 - Gráfico das Produções Científicas de acordo com o tipo de publicação	89
FIGURA 56 - Gráfico da Produção Científica por Áreas de Conhecimento	91
FIGURA 57 - Percentual dos projetos mantido pelo COPE por colegiado.....	93
FIGURA 58 - Gráfico dos Projetos mantidos pelo COPE nos últimos 3 anos, por tipo.	94
FIGURA 59 - Código de segurança solicitado para acesso aos Currículos cadastrados na Plataforma Lattes	96

LISTA DE TABELAS

TABELA 1 - Produção Científica dos docentes do IFPR campus Palmas	78
TABELA 2 - Estatística descritiva da Produção Científica dos docentes do IFPR campus Palmas por ano.....	87
TABELA 3 - Quantidade de Produção Científica de acordo com o tipo de trabalho publicado em eventos de cada colegiado.....	88
TABELA 4 - Análise estatística descritiva das produções científicas por tipo de produção e colegiado.....	90
TABELA 5 - Quantidade da Produção Científica por Área de Conhecimento	90
TABELA 6 - Análise estatística descritiva das produções científicas por área de conhecimento e ano.....	91
TABELA 7- Quantidade de projetos de ensino, pesquisa, ensino e inovação dos docentes dos colegiados do IFPR campus Palmas	92
TABELA 8 - Resultados da estatística descritiva dos projetos submetidos ao COPE.	94

LISTA DE APÊNDICES

APENDICE I - XML do esquema dos cubos OLAP, gerado através da ferramenta Schema Workbench.....	102
APENDICE II - Artigo completo submetido para a VII Contextos e Conceitos – Mostra de Produção Científica do IFPR campus Palmas.	106
APENDICE III - Pôster apresentado no VI SE ² PIN – Seminário de Extensão, Ensino, Pesquisa e Inovação do IFPR.	119
APENDICE IV - Oficina proposta ao VI SE ² PIN – Seminário de Extensão, Ensino, Pesquisa e Inovação do IFPR.	120

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 JUSTIFICATIVA	15
1.2 OBJETIVOS	17
1.2.1 <i>Objetivo Geral</i>	17
1.2.2 <i>Objetivos Específicos</i>	18
1.3. ESTRUTURA DO TRABALHO	18
2. FUNDAMENTAÇÃO TEÓRICA	19
2.1 SISTEMAS DE APOIO A DECISÃO	19
2.3 DATA WAREHOUSE.....	22
2.4 PROCESSAMENTO ANALÍTICO ONLINE - OLAP	26
2.5 ESTADO DA ARTE.....	29
3. MATERIAIS E MÉTODOS	32
3.1 A PLATAFORMA PENTAHO	32
3.1.1 <i>PENTAHO DATA INTEGRATION</i>	34
3.1.2 <i>MONDRIAN SCHEMA WORKBENCH</i>	35
3.1.3 <i>BUSINESS ANALYTICS PLATAFORM</i>	36
3.1.4 <i>SAIKU BUSINESS ANALYTICS</i>	36
3.2 A PLATAFORMA STELA EXPERTA	37
3.3 O ARMAZENAMENTO DE DADOS NO <i>POSTGRESQL</i>	37
3.4 BASES DE DADOS EXISTENTES DO COPE E CURRICULOS LATTES	38
3.5 METODOLOGIA DO TRABALHO	39
4. BI DA PRODUÇÃO CIENTÍFICA DO IFPR CAMPUS PALMAS	40
4.1 MODELAGEM ENTIDADE-RELACIONAMENTO (ER) E MULTIDIMENSIONAL DO <i>DATA WAREHOUSE</i>	40
4.2 O PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DE DADOS.....	42
4.3 CRIAÇÃO DO <i>DATA WAREHOUSE</i> DE PRODUÇÃO CIENTÍFICA.....	57
4.4 CRIAÇÃO E PUBLICAÇÃO DOS CUBOS OLAP	62
4.5 VISUALIZAÇÃO E CONSULTA DOS DADOS A PARTIR DA PUBLICAÇÃO DOS CUBOS OLAP	73
5. ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	77
5. 1. PRODUÇÃO CIEN TÍFICA DE ACORDO COM O CURRÍCULO LATTES	78
5. 2. PRODUÇÃO CIEN TÍFICA PUBLICADA EM ANAIS DE EVENTOS	88
5. 3. PRODUÇÃO CIEN TÍFICA POR ÁREA DE CONHECIMENTO	90
5. 4. PROJETOS DESENVOLVIDOS DE ACORDO COM A BASE DE DADOS DO COPE DO IFPR CAMPUS PALMAS	92
6. CONSIDERAÇÕES FINAIS	95
REFERÊNCIAS BIBLIOGRÁFICAS.....	98
APENDICES	102

1. INTRODUÇÃO

Os Institutos Federais são instituições de educação superior, básica e profissional, pluricurriculares e multicampi, especializados na oferta de educação profissional e tecnológica em diferentes modalidades de ensino. Entre os objetivos definidos em sua criação através da Lei nº 11.892/2008, destacam-se a realização de pesquisas aplicadas, a produção cultural, o empreendedorismo, o cooperativismo, o desenvolvimento científico e tecnológico e, também o desenvolvimento de técnicas e tecnologias.

De acordo com a Lei, uma das características e finalidades da instituição é realizar e estimular a pesquisa e o desenvolvimento científico tecnológico. A aplicabilidade da pesquisa dentro das Instituições de Ensino (IE), como os Institutos Federais, traz novas expectativas no desenvolvimento profissional a cargo da produção do conhecimento em prol da sociedade e benefícios a instituição.

A Lei de criação dos Institutos traz ainda que o desafio colocado no campo da pesquisa é ir além da descoberta científica. Em seu compromisso com a humanidade, a pesquisa, que deve estar presente em todo o trajeto da formação do trabalhador, deve representar a conjugação do saber e de mudar e se construir, na indissociabilidade da pesquisa, ensino e extensão. E mais, os novos conhecimentos produzidos pelas pesquisas deverão estar colocados a favor dos processos locais e regionais numa perspectiva de seu reconhecimento e valorização no plano nacional e global.

Com o objetivo de identificar e mapear numa perspectiva quantitativa o perfil da produção bibliográfica e técnica dos docentes de um campus do Instituto Federal do Paraná escolhido para estudo de caso, o campus Palmas, este trabalho apresenta seu foco na aplicabilidade de técnicas de *Business Intelligence* (BI) sobre os dados de publicações científicas do campus.

Business Intelligence é uma área interdisciplinar da tecnologia da informação que tem como objeto de estudo a elaboração de sistemas de informação computacionais, tais como *data warehouse* (responsáveis por organizar grande volume de dados), *data mining* (responsáveis por facilitar a descoberta de relações entre os dados) e oferecer interfaces que facilitem para o usuário o entendimento das relações entre os dados, com a intenção de ajudar, por exemplo, na tomada de decisão (TURBAN, 2009 *apud* MATHEUS & PARREIRAS, 2004).

Dentre seus componentes pode-se destacar como principais o *data warehouse* (*DW*), as ferramentas de processo de carga (ETL), as ferramentas de acesso aos dados (OLAP), metadados, relatórios e consultas e *dashboards*.

Diversas instituições de ensino já fazem uso dessas tecnologias e indicadores em seus processos decisórios, como o trabalho de Santana *et al.* (2011) do Programa de Pós-Graduação em Sociologia (PPGS) da Universidade Federal do Pernambuco (UFPE), que através de uma análise qualquantitativa foi capaz de identificar os principais aspectos e tendências da produção científica do PPGS.

Filho *et al.* (2015) aplicou o conceito e uso do *Business Intelligence* através do ambiente *open source* de Ensino a Distância (EaD) Amadeu para apoiar as decisões estratégicas da instituição, possibilitando a simulação da geração de relatórios ad-hoc e *dashboards* capazes de possibilitar consultas mais rápidas, flexíveis e seguras.

Munaretto *et al.* (2017) identificou o conceito dos cursos de graduação da UFSM de Frederico Westphalen através da aplicação de técnicas da estatística descritiva sobre dados extraídos da plataforma Lattes e analisados através do indicador 2.14 do IACG utilizado pelo MEC/INEP.

Mena-Chalco e Junior (2013) desenvolveram o *scriptLattes*, um *software* de licença livre capaz de extrair informações dos Currículos Lattes e apresentá-las em forma de relatórios capazes de identificar informações como (i) colaboração entre autores através de grafos; (ii) mapa de geolocalização das publicações e (iii) informações quantitativas correspondentes a produções bibliográficas, técnicas e artísticas.

Dias *et al.* (2016) identificaram a caracterização dos bolsistas de produtividade em pesquisas do CNPq através da utilização de um *framework* para extração de Currículos Lattes.

Neste trabalho, para quantificar, explorar e descrever características da produção científica dos docentes do IFPR campus Palmas, será formalizado um ambiente de apoio à decisão tendo como ponto central um *Data Warehouse* criado a partir da extração, transformação e carga dos dados, do período compreendido entre janeiro de 2014 e dezembro de 2016, oriundos das fontes de dados dos projetos submetidos ao Comitê de Pesquisa e Extensão (COPE) do campus e dos Currículos Lattes dos seus pesquisadores obtidos através da Plataforma Lattes do CNPq e um cubo OLAP responsável por manipular e analisar os dados armazenados no DW sob múltiplas perspectivas tais como a quantidade de produção científica por ano, colegiado, área de conhecimento, tipo de produção, pesquisador.

Como resultados, esse trabalho pode auxiliar à gestão do ensino, pesquisa, inovação e extensão da instituição na compreensão da dinâmica da produção científica do campus, no

atingimento de metas previamente delineadas por instituições brasileiras de fomento à pesquisa, como a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e fornecer informações para gerir e ampliar o desenvolvimento de projetos visando atender as necessidades da comunidade local e regional.

Este trabalho integra a solução de *Business Intelligence (BI)* de produção científica do IFPR campus Palmas em conjunto a outros dois trabalhos de conclusão de curso de Bacharelado em Sistemas de Informação do corrente ano, o de Geoprocessamento da Produção Científica do IFPR campus Palmas utilizando ferramentas *Opensource* e o de Visualização de Dados e *Design de Dashboards* para Apresentação de Perfil da Produção Científica do IFPR campus Palmas, proporcionando uma análise descritiva da produção científica do campus e melhor compreensão dos resultados obtidos a partir de indicadores como, produtividade dos pesquisadores, áreas de conhecimento que mais atuam, eventos e regiões geográficas que costumam publicar seus trabalhos, entre outros.

1.1 JUSTIFICATIVA

O uso de indicadores da atividade científica é incentivado pela comunidade acadêmica, gestores das instituições e órgãos de fomento que buscam compreender melhor a dinâmica da produção científica no intuito de subsidiar e avaliar o planejamento e resultados das políticas voltadas a esse objetivo.

Órgãos públicos como o CNPq, CAPES, Financiadora de Estudos e Projetos (FINEP), Ministério da Ciência e Tecnologia (MCT), Fundações de Amparo à Pesquisa (FAPs) e as próprias instituições através de suas pró-reitorias de pesquisa e extensão fazem o uso de indicadores e demais informações consistentes e confiáveis para auxiliar em seus processos decisórios.

O órgão de assessoramento da Direção de Ensino, Pesquisa e Extensão de cada campus e da Pró - Reitoria de Extensão, Pesquisa e Inovação (PROEPI) do IFPR responsável por formular, subsidiar e acompanhar a execução da política de pesquisa científica e extensão no âmbito institucional é o Comitê de Pesquisa e Extensão (COPE), regido pela Portaria 3, de 9 de junho de 2009.

De acordo com o Regimento Interno do COPE – *Campus Palmas* de 30 de agosto de 2016, são atribuições do comitê divulgar, no âmbito do IFPR campus Palmas, as diretrizes, normas e editais relativos aos projetos de pesquisa, extensão, inovação e ensino-aprendizagem; receber, analisar, emitir parecer e esclarecer possíveis dúvidas referentes aos projetos de pesquisa, extensão, inovação e ensino-aprendizagem submetidos pelos servidores (técnicos e docentes) do campus Palmas; manter atualizado a base de dados e fornecer anualmente um relatório consolidado dos projetos em andamento ou concluídos de pesquisa, extensão, inovação e ensino-aprendizagem do campus Palmas a fim de assessorar a direção de ensino, pesquisa e extensão do campus no monitoramento destes projetos (COPE, 2016).

Em maio de 2016, o comitê era composto por seis membros que se reúnem mensalmente para assessorar, analisar e emitir pareceres dos projetos de aproximadamente 150 (cento e cinquenta) docentes organizados em 15 (quinze) colegiados diferentes, sendo seis de cursos de licenciatura (Química, Letras Português/Inglês, Ciências Biológicas, Educação Física, Pedagogia e Artes Visuais); oito de cursos de bacharelado (Sistemas de Informação, Administração, Ciências Contábeis, Engenharia Agronômica, Enfermagem, Farmácia, Direito e Engenharia Civil) e um colegiado de formação multidisciplinar.

O regime de trabalho dos docentes efetivos do quadro permanente do IFPR ou dos docentes com contrato de trabalho na qualidade de substitutos, definido segundo critérios de contratação previstos na legislação Federal – Lei nº 7.596/87, de 10/04/87, Decreto Federal 94.664, de 23/07/87, Portaria nº 475, de 26/08/87 e Lei 11.784, de 22/09/08, de tempo integral de 40 (quarenta) horas e dedicação exclusiva, exceto os docentes afastados na forma da lei, apresenta como obrigatório o cumprimento de 16 (dezesseis) horas da carga horária de trabalho semanal em atividades de pesquisa e/ou extensão.

Para realizar a demanda das atividades requeridas pelo COPE cada membro titular do comitê dispõe de duas horas de sua carga horária de trabalho semanal, com exceção do presidente, que dispõe de quatro horas de trabalho semanal. O volume de trabalho do COPE do campus é grande e oneroso para esse pequeno número de membros e baixa carga horária dos membros destinada às atividades do comitê. Ainda, não existe um sistema de informação na instituição disponível para uso do COPE que auxilie a manter atualizado o banco de dados das atividades dos pesquisadores do campus, requerendo um esforço adicional para realizar esse controle através de planilhas e e-mails.

O cadastro, monitoramento e acompanhamento dos projetos é realizado em uma planilha eletrônica que atende parcialmente às necessidades da instituição e não apresenta segurança, consistência ou persistência eficiente dos dados. Uma dificuldade muito grande é

encontrada principalmente na geração de relatórios semestrais e anuais dos projetos. Além de não serem de fácil compreensão e entendimento, não são gerados gráficos e os dados armazenados contribuem parcialmente e de maneira trabalhosa para a geração de conhecimento sobre o perfil dos projetos desenvolvidos no campus.

A gestão da pesquisa e extensão do campus tem frequentemente recorrido a consultas por e-mail, a coordenadores dos cursos e aplicação de auditorias internas, a fim de levantar informações sobre a produção científica do campus e utilizar os dados em seus processos decisórios.

Desta forma, a construção de um sistema de informação capaz de gerar indicadores de produção científica torna-se uma importante contribuição para o COPE do *campus* Palmas, para a própria gestão de ensino, pesquisa e extensão, comunidade acadêmica e da região que podem ser subsidiadas com informações sobre a atuação e projetos desenvolvidos pelos pesquisadores do campus.

1.2 OBJETIVOS

A seguir é apresentado uma descrição do objetivo geral deste trabalho, seguido dos objetivos específicos que nortearam o desenvolvimento do mesmo.

1.2.1 Objetivo Geral

Desenvolver um ambiente de apoio à decisão através da aplicação de técnicas de *Business Intelligence* para identificar o perfil da produção científica e atuação dos pesquisadores do IFPR campus Palmas.

1.2.2 Objetivos Específicos

Para atingir o objetivo geral apresentado, este trabalho é orientado por cinco objetivos específicos:

- a) Revisão bibliográfica sobre *Business Intelligence*, *Data Warehouse* e Processamento Analítico;
- b) Identificação de trabalhos relacionados ao desenvolvimento de *Data Warehouse* sobre dados educacionais e de produção científica a fim de identificar o estado da arte do tema;
- c) Identificar e estudar as fontes de dados que serão utilizadas no trabalho;
- d) Extraír os dados das diferentes origens, integrá-los e armazená-los em um repositório único central *Data Warehouse*;
- e) Criar o cubo OLAP para manipular e analisar os dados armazenados no *Data Warehouse* sob múltiplas perspectivas a partir de indicadores de produção científica que auxiliem a identificar o perfil próprio de atuação da instituição para a produção de ciência.

1.3. ESTRUTURA DO TRABALHO

Este trabalho está organizado em mais 5 (cinco) capítulos. O capítulo 2 (dois) apresenta os principais conceitos das áreas de *Business Intelligence* que fundamentam esse estudo. O capítulo 3 (três) descreve os materiais e métodos utilizados no desenvolvimento do mesmo. O processo de extração, transformação e carga dos dados, a arquitetura e modelagem do DW e do cubo OLAP, são apresentados no capítulo 4 (quatro). O capítulo 5 (cinco) apresenta uma discussão dos resultados alcançados. Por fim, no capítulo 6 (seis) são apresentadas as considerações finais e propostas para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Para o desenvolvimento deste trabalho serão abordados no referencial teórico os temas sobre Sistemas de Apoio à Decisão, *Business Intelligence*, *Data Warehouse*, processamento analítico e modelagem multidimensional. Na sequência, é apresentado o estado da arte em soluções de *business intelligence* e *data warehouse* sobre dados de produções científicas em instituições de ensino.

2.1 SISTEMAS DE APOIO A DECISÃO

Os Sistemas de Apoio a Decisão (SADs) visam obter informações de alto nível a partir das informações armazenadas nos sistemas de processamento e transação das organizações, com o objetivo de auxiliar gerentes e diretores na tomada de decisões (SILBERSCHATZ, KORTH, e SUDARSHAN, 2006).

De acordo com Inmon (1997), os SADs, conhecidos anteriormente como MIS (*Management Information Systems* ou Sistemas de Informação Gerenciais) surgiram na década de 80 quando o usuário final passou a controlar diretamente os sistemas e os dados, percebendo então a possibilidade da utilização desses mesmos dados para outros objetivos, além de atender o processamento de transações online de alta *performance*.

Os SAD's visam fornecer apoio em decisões através de informações obtidas de diversas fontes de dados da empresa (TAVARES, 2008). Silberschatz, Korth, e Sudarshan (2006) relatam que o armazenamento e a recuperação de dados para apoio à decisão levantam algumas questões, como: a proposta da área de processamento analítico on-line (OLAP – *On-Line Analytical Processing*); o armazenamento e recuperação de grandes volumes de dados para realização de análises estatísticas eficientes; a adoção de armazéns de dados (do inglês *Data Warehouse*) por parte de grandes empresas que possuem diversas fontes de dados sob diferentes esquemas; e a mineração de dados.

Ainda sobre as características de um Sistema de Apoio a Decisão, Ceci (2012, *apud* TURBAN, 1990) apresenta outras características como: incorporação de modelos e dados, foco no auxílio ao gestor na tomada de decisões para resolução de problemas estruturados,

semiestruturados e não estruturados, suporte a tomada de decisão em conjunto com a avaliação do gestor e a melhoria na qualidade das decisões.

Com relação a arquitetura básica de um sistema de apoio a decisão, Heinze (2010) explica que esta é composta por três subsistemas: subsistema de dados – composto pelo gerenciador de dados, subsistema de modelos – composto pelo banco de modelos e seu gerenciamento e o subsistema de interfaces – responsável pela interação entre o sistema e o usuário.

Atualmente, a implementação da arquitetura mais utilizada é a de *Business Intelligence* (BI), que é considerada a evolução dos sistemas de apoio a decisão (CECI, 2012).

2.2 BUSINESS INTELLIGENCE

Sob uma perspectiva histórica, o mundo dos sistemas de informações pode ser considerado “imaturo” quando comparado com outras áreas, como por exemplo a engenharia civil que, com suas raízes na antiguidade, possui atividades realizadas há mais de 2000 anos. Já a área de processamento de informações surgiu apenas no início da década de 60 (INMON, 1997). O termo *Business Intelligence* (BI) é ainda mais recente, de acordo com Loh (2014) seu surgimento se deu após metade da década de 90 com a evolução dos sistemas de geração de relatórios, também conhecidos como Sistemas de Informações Gerenciais (SIG) e Sistemas de Informações Executivas (EIS).

Loh (2014) salienta ainda que, apesar da existência de técnicas, tecnologias e softwares de *Business Intelligence*, o BI é um processo que envolve uma gama de métodos, métricas, ferramentas, pessoas, entre outros. As metas fundamentais de um sistema de *business intelligence* são, de forma genérica, recolher dados, transformá-los em informações (através da descoberta de padrões e tendências) e, sequencialmente, informação em conhecimento útil e oportunamente para a tomada de decisão (SANZÕES, OLIEIRA e BAPTISTA, 2006).

Sezões, Oliveira e Baptista (2006) definem BI como:

Business Intelligence – conceito que engloba um vasto conjunto de aplicações de apoio à tomada de decisão que possibilitam um acesso rápido, partilhado e interativo das informações, bem como a sua análise e manipulação; através destas ferramentas, os utilizadores podem descobrir relações e tendências e transformar grandes quantidades de informações em conhecimento útil.

Nesse sentido, Turban (2009) e Machado (2013) concordam ao definirem BI como um termo “guarda-chuva” que inclui arquiteturas, ferramentas, bancos de dados, aplicações e

metodologias cujos principais objetivos são: permitir o acesso interativo aos dados e proporcionar a manipulação desses dados fornecendo aos gerentes e analistas de negócios a capacidade de realizar a análise adequada conseguindo capturar valiosos *insights* que podem servir como base para decisões melhores e mais informadas.

A arquitetura e o fluxo de dados de um BI, representado na Figura 1 é, geralmente, formada por três elementos: o repositório de dados provenientes dos sistemas de informação da organização, ferramentas de mineração de dados com a função de explicitar as informações e conhecimentos implícitos nas bases de dados da organização, e as ferramentas OLAP responsáveis pela apresentação e cruzamento das informações a fim de apoiar no processo decisório (CECI, 2014).

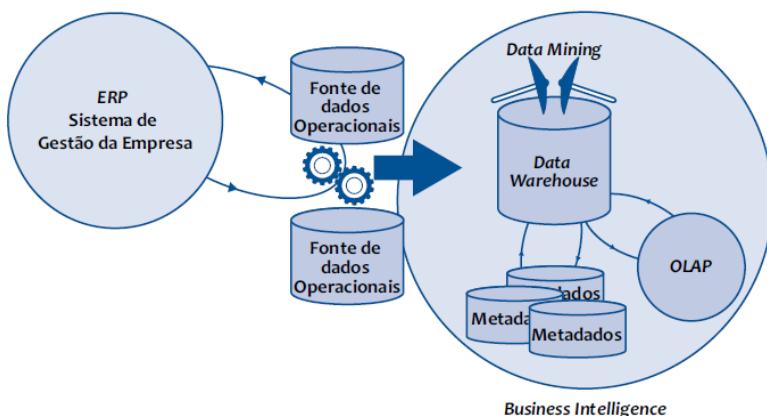


FIGURA 1- Arquitetura do BI
FONTE: Ceci, 2014.

Ceci (2014, *apud* REGINATO e NASCIMENTO, 2007) traz que as ferramentas provenientes da arquitetura de BI podem fornecer uma visão sistêmica dos negócios da organização, auxiliando na distribuição uniforme dos dados entre os usuários e facilitando na análise de vários segmentos das organizações como: tendências de transformação do mercado, preferências de clientes, recursos da empresa, condições de mercado, entre outros.

Contudo, o autor apresenta ainda, que não existe uma solução de BI genérica, capaz de atender diversas organizações. Cada organização possui um cenário particular e necessidades de respostas para diferentes perguntas. Embora a arquitetura do BI seja comum para diversas instituições, cada projeto possui suas particularidades. De forma geral, um projeto de BI inicia com o Planejamento Estratégico Corporativo (PEC) – procedimento que traça as principais metas e estratégias que devem ser alcançadas – e o Planejamento Estratégico da Informação (PEI) – responsável por identificar todos os sistemas e bases de dados da organização. Com o levantamento sistêmico da organização, deve-se identificar quais tipos de perguntas deseja-se

responder com o sistema de BI. Após levantadas essas informações, parte-se para a modelagem do repositório único de dados e informações, o *Data Warehouse* (DW). Com o DW modelado e carregado, é necessário identificar quais as técnicas de análises poderão ser aplicadas para a extração de informações e conhecimentos.

Para o desenvolvimento desse trabalho, as técnicas aplicadas de BI foram as de *data warehouse*, processamento analítico *on-line*, extração, transformação e carga de dados.

2.3 DATA WAREHOUSE

O *Data Warehouse* é um conjunto de dados baseado em assuntos, integrado, não-volátil, e variável em relação ao tempo, de apoio às decisões gerenciais, considerado o alicerce dos Sistemas de Apoio à Decisão sendo capaz de tornar a tarefa do analista de SAD muito mais fácil do que no ambiente clássico de informação em virtude de haver uma única fonte de dados integrados com condições de acesso (INMON, 1997).

Para Silberschatz *et al.* (2006) um *Data Warehouse* é um repositório de informações colhidas de várias origens, armazenadas sob um esquema unificado, em um único local.

Já Machado (2013) explica que o DW funciona como um grande conjunto de todos os dados de uma organização, formado por diversos *Data Marts*, que atuam como repositórios menores, orientados a áreas específicas.

Dentre as principais características do *Data Warehouse* já citadas por Inmon (1997), Machado (2013) destaca as seguintes:

a) Orientado por assunto: um DW armazena as informações agrupadas por assuntos de interesses da empresa que são mais importantes para a tomada de decisões, relativos a análise e desempenho de processos ou atividades críticas. Esses assuntos, são em realidade os principais processos de uma organização.

b) Variação de Tempo: no DW a data é elemento essencial e componente-chave, pois trata-se de uma estrutura sobre janelas de tempo. Diferente dos ambientes transacionais onde o horizonte de tempo gira em torno de 60 a 90 dias, no DW os espaços de tempo são significativamente maiores, situando-se entre cinco a dez anos. Isso implica que os dados de um DW não podem ser atualizados.

c) Não volátil: um DW tem duas operações básicas: a carga de dados (inicial e incremental) e a consulta a esses dados. Diferente dos ambientes transacionais onde os dados são carregados diretamente no banco de dados, no DW os dados provêm de um ambiente operacional após serem filtrados, limpos e transformados. Depois de carregado, um DW somente possui operações de consultas e novas inserções.

d) Integração: essa é uma das características de suma importância em um DW, sendo fundamental devido a necessidade de unicidade de informações. A existência de sistemas mais antigos com padrões de codificação diferentes, como por exemplo a convenção de nomes e valores de variáveis, tais como sexo masculino e feminino leva a existência de diferentes padrões entre os sistemas operacionais, que quando carregados no DW são resolvidos pelos processos de filtragem e agregação.

Além dessas características, existem outras premissas que diferem um *Data Warehouse* de uma base de dados transacional (OLTP – *On-Line Transaction Processing*), os autores Teorey, Lighstone e Nadeau (2007) demonstraram em uma tabela, conforme mostrado na Figura 2, as principais diferenças dos dois tipos de sistemas.

OLTP	Data Warehouse
Orientado a transação (operação)	Orientado ao processo do negócio (a assuntos)
Milhares de usuários	Poucos usuários (normalmente a camada gerencial)
Geralmente utiliza pouco espaço (MB até vários GB)	Utiliza muito espaço (de milhares de GB a vários TB)
Dados atuais	Dados históricos (fotografias)
Dados normalizados (muitas tabelas, poucas colunas por tabela)	Dados não normalizados (poucas tabelas com muitas colunas)
Atualização contínua	Atualizações em lote (processo de carga de tempos em tempos)
Consultas de simples a complexas	Normalmente, consultas muito complexas.

FIGURA 2 - Comparativo entre sistemas OLTP e Data Warehouse.

FONTE: Teory, Lightstone e Nardeu, 2007.

Sendo o *Data Warehouse*, repositórios integradores de informações relevantes às organizações, sua utilização pode ser realizada através dos *Data Marts*, que, segundo Ceci (2012, apud XAVIER e PEREIRA, 2009), podem ser descritos como um pequeno *Data Warehouse* que possibilita apoio à decisão de um pequeno grupo de pessoas e assuntos, concentrando-se na exigência de um departamento específico.

Colaço Junior (2004) faz o alerta que a presença de muitos *data marts* pode resultar no risco de redundância, por isso deve-se sempre ter a preocupação de compartilhar os dados, tabelas e relatórios em comum entre os departamentos.

A Figura 3 apresenta, de maneira reduzida, como funciona uma arquitetura padrão de um *data warehouse* e seus *data marts*:

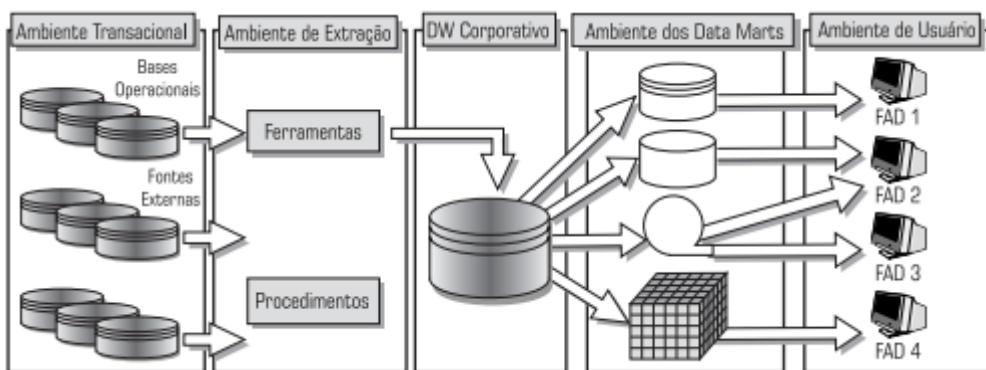


FIGURA 3 - Arquitetura do Data Warehouse
FONTE: Colaço Junior, 2004.

De acordo com a Figura 3, a arquitetura de um *Data Warehouse* começa com os dados oriundos de diversas fontes e dispostos em diversos formatos (ambiente transacional), para realizar a integração desses dados é possível a utilização de ferramentas de ETL – *Extraction, Transformation and Load* responsáveis por filtrar e padronizar os dados que serão carregados no *data warehouse*. De acordo com Colaço Junior (2004), os processos de ETL consomem mais de 70% do tempo de desenvolvimento e é necessário devido à diversidade existente em termos de estruturas de dados nos sistemas transacionais das empresas. Antes de carregar os dados para o DW, eles devem ser passados por uma área intermediária de armazenamento chamada *Operational Data Storage ou Staging Area* (ODS) a fim de facilitar a integração dos dados do ambiente operativo (MACHADO, 2013). Para economizar espaço de armazenamento em disco, muitos DW's são implementados sem ODS o que não gera implicações graves, contudo, se houver erros na carga de dados será necessário um tempo maior para realizar as devidas correções (COLAÇO JUNIOR, 2004). Somente após a integração e limpeza dos dados é que se realiza a exportação para o DW e transmissão para os *data marts* permitindo que os dados sejam consultados por usuários através das ferramentas de apoio a decisão.

Ceci (2012) destaca que umas das fases iniciais do desenvolvimento de DW é a identificação e entendimento das necessidades do negócio, onde serão levantadas as perguntas que se deseja responder com a análise dos dados e informações contidas no DW e também como

ele deverá ser organizado. Uma das diferenças entre os DWs e as bases de dados OLTP está na modelagem dos dados.

A técnica de projeto lógico de banco de dados mais utilizada no desenvolvimento *de Data Warehouse* é a modelagem dimensional que está representada na Figura 4.

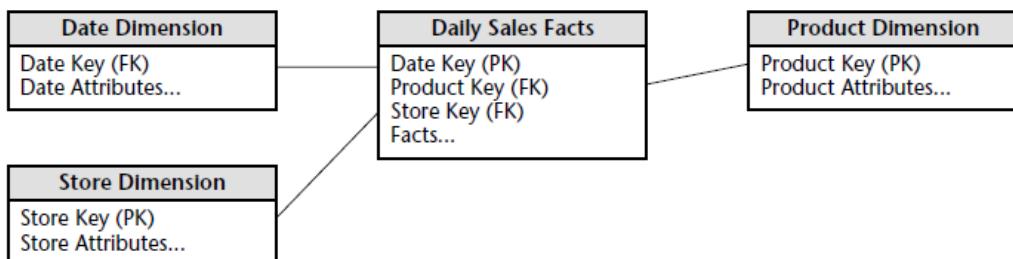


FIGURA 4 – Modelagem Dimensional

Fonte: Kimball, 1997.

O modelo dimensional é composto por uma tabela com várias chaves, chamada de tabela fato, e um conjunto de tabelas menores nomeadas tabelas de dimensão. Cada tabela de dimensão possui uma chave primária que corresponde exatamente a uma das chaves estrangeiras que compõe a tabela fato. Esta ligação entre as tabelas de dimensão à tabela fato possui uma forma parecida com uma estrela, por esse motivo, esse modelo é também conhecido como Modelo Estrela (do inglês *Star Schema*) (KIMBALL, 1998). Uma das principais características desse modelo é que ele é altamente reconhecível tanto para os desenvolvedores quanto para os usuários finais por descrever aspectos comuns de negócios.

Neste modelo, cada fato (representado através da tabela fato) representa um item ou evento do negócio e é utilizado para analisar o processo de negócio da empresa, a principal característica de um fato é que ele é representado por valores numéricos e aditivos (MACHADO, 2013). A aditividade é crucial devido à grande quantidade de registros recuperados por um *Data Warehouse*. Em contrapartida, as informações textuais e descritivas são encontradas nas tabelas de dimensões.

As tabelas de dimensão buscam responder quatro questões do fato: Quando? Quem? Onde? E O que? (CECI, 2012). Os atributos de dimensões são constantemente utilizados nas consultas de *data warehouse*, por isso sua estrutura deve ser muito bem planejada (KIMBALL, 1997).

De acordo com Ceci (2012), é nas dimensões que estão armazenadas as informações complementares aos dados cadastrados na tabela fato, que são utilizadas como variáveis de corte para auxiliar nos filtros de ferramentas OLAP.

2.4 PROCESSAMENTO ANALÍTICO ONLINE - OLAP

O processamento analítico online (OLAP) corresponde a atividade de analisar grandes quantidades de dados em tempo real, implicando que, mesmo existindo enormes quantidades de dados, o sistema deve responder a consultas com rapidez suficiente para permitir uma exploração interativa dos mesmos (MONDRIAN, 2017).

O termo OLAP (*Analytic Processing On-Line*, ou processamento analítico *on-line*) surgiu em 1993 em um ensaio apresentado pelo Dr. Codd, caracterizando o conjunto de ferramentas capazes de apresentar as informações contidas em um *Data Warehouse* em relatórios, respeitando suas peculiaridades e características. Antes desta data, já existiam alguns produtos com características OLAP no mercado que então passaram a ser referenciadas como ferramentas OLAP (COLAÇO JUNIOR, 2004).

Sezões (2006) define o processamento analítico *on-line* como um conceito que se refere a ferramentas capazes de efetuar, de forma rápida e partilhada, a análise de informações multidimensional, originária de diversas fontes de dados.

As ferramentas OLAP são caracterizadas por oferecer um grande potencial de recuperação e análise de informações rápida e fácil. O processamento OLAP é capaz de proporcionar acesso seguro e controlado aos dados de um *Data Warehouse* de forma que os usuários possam reestruturar os dados de uma base e dados relacional numa perspectiva multidimensional, podendo assim realizar comparações, analisar tendências, destacar problemas e manipular informações em qualquer ordem (COLAÇO JUNIOR, 2004 e SEZÕES, 2006).

Ao realizar o comparativo entre o processamento OLAP e OLTP, pode-se destacar que as principais diferenças entre ambos estão relacionadas ao modelo e tipo de dados, forma de acesso, finalidade, atualizações, complexidade e condições dos dados. Nesse sentido, enquanto no modelo relacional (OLTP) os dados são normalizados e bastante detalhados com atualizações e operações diárias que podem ser acessados apenas via SQL (*Structured Query Language*, ou linguagem de consulta estruturada), no processamento analítico (OLAP) a complexidade é maior, com dados consolidados, organizados sob uma perspectiva multidimensional, cuja finalidade é a recuperação e análise desses dados que podem ser acessados, não apenas via SQL, mas também através do uso de ferramentas analíticas.

A principal característica de um sistema OLAP é permitir uma visão conceitual e multidimensional dos dados, assim os dados são modelados em uma estrutura com o formato

de um cubo onde cada dimensão representa os temas mais importantes da empresa (COLAÇO JUNIOR, 2004). Segundo Ceci (2012, *apud* DATTA e THOMAS, 1999) um cubo de dados (ou cubo OLAP) é uma peça fundamental na criação e construção de um banco de dados dimensional, servindo como uma unidade básica de entrada e saída para todos os operadores definidos em um banco de dados, formado por células com valores ou medidas que tomam como base um conjunto de dimensões.

Silberschatz (*et al.*, 2006) explica que cubos OLAP permitem que os usuários vejam dados em qualquer nível de granularidade. Inmon (1997) define granularidade como o mais importante aspecto do projeto de um DW, pois afeta profundamente o volume de dados que residem no DW e, ao mesmo tempo, afeta o tipo de consulta que pode ser atendida. A granularidade diz respeito ao nível de detalhes ou resumo contido nas unidades de dados existentes no DW. Quanto mais detalhes, mais baixo o nível de granularidade. Quanto menos detalhes, mais alto o nível de granularidade.

Para Machado (2013) são as ferramentas OLAP que permitem aos usuários finais acesso para extração dos dados de suas bases e construção de relatórios capazes de responder às questões gerenciais. Para isso, são utilizadas as seguintes operações:

a) *Drill Down* e *Roll Up* – utilizam a navegação dos dados, modificando o nível de granularidade da consulta. O *Drill Down* ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo o nível de granularidade. Já o *Roll Up* (ou *Drill Up*) é o contrário, ocorrendo quando o usuário aumenta o nível de granularidade, diminuindo o nível de detalhamento da informação. Por exemplo: quando se deixa de analisar uma consulta trimestral e se passa a visualizá-la com dados anuais, utiliza-se a opção *roll up*, pois a consulta passa de um nível mais baixo de detalhe (um trimestre específico de um ano) e passa para um nível mais alto (todos os trimestres de um ano).

b) *Drill Across* – quando o usuário pula um nível intermediário dentro de uma mesma dimensão, como por exemplo: se a dimensão tempo for formada por ano, semestre, trimestre, mês e dia e o usuário passar a consulta do ano direto para trimestre, mês ou dia, essa execução é chamada de *drill across*.

c) *Drill Through* – quando o usuário muda de uma informação contida em uma dimensão para outra. Por exemplo: quando o usuário deixa de analisar as informações por tempo e passa a analisar as informações por região.

d) *Slice* e *Dice* – considerada uma característica de extrema importância dos cubos OLAP, essas operações realizam a navegação por meio dos dados na visualização de um cubo, resultando na redução do escopo dos dados em análise e mudança na ordem das dimensões,

permitindo a análise das informações sob diferentes prismas. Por exemplo: um cubo com informações de venda de celulares e *notebooks* de diversas lojas de varejo espalhadas pelo Brasil, todas da mesma franquia. *Slice* é a operação que corta o cubo, mantendo a mesma perspectiva de informações dos dados, permitindo que seja visualizada somente a venda de um tipo de produto, por exemplo, os celulares. *Dice* é a mudança de perspectiva da visão, a extração de um “subcubo” ou a intersecção de vários *slices*. Fazendo a alusão de um cubo, é como se o cubo fosse girado para ser visto sob outra perspectiva.

e) *Pivot* – corresponde à modificação da posição das dimensões em um gráfico ou troca de linhas por colunas em uma tabela.

Sezões, Oliveira e Baptista (2006), relatam que os cubos OLAP podem ser distinguidos essencialmente em três tipos: (i) o relacional (ROLAP), caracterizado por manter os dados nas tabelas relacionais originais, ao mesmo tempo que gera outras para armazenar os valores agregados tornando esse modelo globalmente mais lento – se comparado com as outras opções, uma vez que é o modelo relacional quem sustenta o trajeto de todas as uniões entre as tabelas; (ii) o multidimensional (MOLAP) é uma estrutura gerida por um motor de base multidimensional onde os cálculos das agregações acontecem de maneira automática, o que torna-o extremamente rápido em termos de respostas às questões dos utilizadores, mas, com desvantagens a nível do espaço ocupado e do tempo despendido em sua criação; (iii) o modelo HOLAP surge como um meio-termo entre os dois modelos anteriores, assumindo aspectos e atributos de ambos, onde os dados permanecem nas tabelas relacionais e as agregações ficam armazenadas em uma estrutura multidimensional, tornando-o uma opção viável tanto em termos de rapidez de resposta quanto em termos de dimensão.

Neste trabalho a manipulação e análise dos dados sob múltiplas perspectivas é baseada em bancos relacionais, ou seja, no modelo ROLAP. Ao montar o cubo de dados, com informações oriundas do *Data Warehouse*, o usuário poderá realizar as operações de *Drill Down*, *Roll Up*, *Drill Across*, *Drill Through*, *Slice* e *Dice*, permitindo a navegação e visualizando dos dados através de relatórios sob diferentes perspectivas, como por exemplo, a consulta de publicações ou projetos por ano, colegiado, pesquisador.

2.5 ESTADO DA ARTE

Para o desenvolvimento deste trabalho, foi realizada uma pesquisa sobre o estado da arte em sistemas de *Business Intelligence* desenvolvidos para instituições de ensino com o objetivo de constatar nos trabalhos similares características e funcionalidades que auxiliem na elaboração do sistema proposto nesse projeto e o atual estado da arte na área.

Deste modo, a coleta de dados foi realizada com base em materiais publicados em revistas especializadas, conferências, simpósios e seminários da área de ciência da computação, como o Simpósio Brasileiro da Computação (SBC) e a Escola Regional de Banco de Dados (ERBD). As ferramentas de pesquisa utilizada foram o Google Acadêmico (scholar.google.com.br), o Portal de Periódicos da CAPES (<http://www.periodicos.capes.gov.br>), Biblioteca Digital Brasileira de Computação – BDBComp (<http://www.lbd.dcc.ufmg.br/bdbcomp>) e a Biblioteca Eletrônica Científica On-Line – SciELO, do inglês *Scientific Electronic Library Online* (<http://search.scielo.org>).

Para a obtenção dos resultados da busca foram utilizados argumentos de pesquisa escritos em português como: (perfil) AND (produção científica), (extração de dados) AND (currículo lattes), partes e outras combinações desses argumentos, pelo nome de trabalhos encontrados, referências bibliográficas de documentos analisados como artigos publicados. Em um dos casos, o autor do trabalho foi contato por e-mail para esclarecimentos sobre a metodologia utilizada na extração das informações dos Currículos Lattes analisados. Em sua resposta, o autor informou que utilizou um *framework* desenvolvido em sua tese de doutorado e que, até a redação desta seção, ainda não havia sido publicado.

A seleção dos trabalhos para análise foi realizada no período de março a abril de 2017 e os resultados da mesma encontram-se a seguir.

Santana *et al.* (2011) apresenta uma análise da produção científica do Programa de Pós-Graduação em Sociologia (PPGS) da Universidade Federal de Pernambuco (UFPE) através do uso de indicadores. As informações foram extraídas do Currículo Lattes de cada professor membro do PPGS resultando em uma análise qualitativa capaz de identificar os principais aspectos e tendências da produção científica do PPGS. Através desse estudo, percebeu-se um aumento significativo no número de publicações de livros, capítulos de livros e artigos de periódicos considerados “produção qualificada” pela CAPES. Também foi possível identificar as redes de colaboração entre autores, a média de publicação de cada professor do PPGS e quais as áreas onde a concentração de artigos é maior.

Mena-Chalco e Cesar Junior (2013) desenvolveram um software, de licença livre conhecido como *scriptLattes*, capaz de extrair as informações do Currículo Lattes, a fim de auxiliar na elaboração de relatórios de produção científica. Após realizar extração dos dados, o sistema os apresenta através de relatórios capazes de identificar informações como (i) colaboração entre autores através de grafos; (ii) mapa de geolocalização das publicações e (iii) informações quantitativas correspondentes a produções bibliográficas, técnicas e artísticas.

Moraes Filho *et al.* (2015) aplicou o conceito e uso do *Business Intelligence* através do ambiente *open source* de Ensino a Distância (EaD) Amadeu para apoiar nas decisões estratégicas da instituição. Por meio de metodologias de pesquisas exploratória e qualitativa, foi disponibilizado uma ferramenta analítica através da suíte de aplicativos *Pentaho*. As ferramentas utilizadas foram: PDI, *Mondrian Scheme Workbench* e *Bi-Server*. Como resultados, o autor simulou a geração de relatórios ad-hoc e *dashboards* com as seguintes consultas: quantidade curso por ano, quantidade de módulo por ano e quantidade de curso por módulo e ano, fornecendo ao ambiente de EaD Amadeu uma ferramenta analítica capaz de possibilitar consultas mais rápidas, flexíveis e seguras.

Dias *et al.* (2016) realizou o levantamento do perfil dos bolsistas de produtividade do CNPq por meio de dados extraídos do currículo lattes de cada professor das categorias 1, 2 e *sênior* contemplados pelo CNPq. O trabalho é apresentado em três etapas, em um primeiro momento é realizado o levantamento de informações sobre bolsas e bolsistas de produtividade de pesquisa do CNPq, para isso, foi utilizado um *framework* responsável por realizar a extração de toda a base de Currículos Lattes do grupo dos bolsistas de produtividade. Na segunda etapa houve a organização dos dados levantados, onde constitui-se o perfil com informações referente a formação acadêmica, produção científica e orientações de cada um dos 14.154 pesquisadores bolsistas identificados anteriormente. Na terceira etapa, realizou-se o tratamento dos dados referente ao perfil dos pesquisadores, como problemas com ambiguidade de nomes, erros ortográficos, entre outros. Com isso, os autores puderam identificar o volume da produção científica em cada nível de bolsa, a faixa etária dos pesquisadores por nível, a distribuição da produção científica entre as modalidades, entre outros.

Munaretto *et al.* (2017), conseguiu identificar e descrever o conceito dos cursos de graduação da UFSM de Frederico Westphalen em relação ao indicador 2.14 do Instrumento de Avaliação dos Cursos de Graduação do MEC/INEP, que descreve os conceitos e os critérios de análise da produção científica, cultural, artística ou tecnológica dos docentes. Para tal, o autor coletou dados da produção científica dos 79 docentes da instituição, com dados de 2013 à 2015, por meio de pesquisa realizada na Plataforma Lattes do CNPq. Através da aplicação de

técnicas da estatística descritiva, foi possível determinar a frequência e média de produção por curso/departamento. Também identificou a maior média de produção dos docentes e o impacto causado no departamento alocado. Através do estudo, foi possível concluir que, através do item 2.14 do IACG todos os cursos estariam aptos a obter o conceito 5, caso avaliados pelo MEC/INEP.

Com base nos trabalhos analisados e como proposto por Moraes Filho *et al.* (2010), Santana *et al.* (2011) e Munaretto *et al.* (2017), este trabalho busca descobrir o perfil da produção científica do IFPR campus Palmas através do uso de indicadores, da extração de informações dos currículos Lattes e do COPE, utilizando as ferramentas da suíte de aplicativos *Pentaho* para realizar a limpeza dos dados, criação da *staging area* e do DW, criação de cubos OLAP e visualização de dados através de relatórios e gráficos.

Apesar de existirem diversas pesquisas para identificação do perfil da produção científica dentro de instituições de ensino, cada instituição possui suas peculiaridades e buscam responder perguntas específicas, não sendo possível desenvolver um BI genérico por exemplo ou utilizar um já desenvolvido para outra instituição.

3. MATERIAIS E MÉTODOS

Nesta seção são apresentadas as principais características e funcionalidades da suíte *Pentaho* utilizada para a extração, transformação e carga dos dados e criação do cubo OLAP; da ferramenta *Stela Experta* utilizada para extração de dados dos currículos lattes; do *PostgreSQL* utilizado para a implementação do DW de produção científica; das bases de dados existentes utilizadas como fonte de dados para a integração e carga no DW, bem como as etapas seguidas para o desenvolvimento do trabalho.

3.1 A PLATAFORMA PENTAHO

Pentaho é uma empresa do *Grupo Hitachi*, líder em integração de dados e análise de negócios através de uma plataforma de código aberto com suporte a diversas implementações de dados, que tem por objetivo atender usuários que trabalham com grande volume de dados (*Big Data*) e Internet das Coisas (do inglês *Internet of Things* - IoT). Fundada em 2004 e incorporada ao grupo Hitachi em 2015, a empresa se destaca pela facilidade de ser incorporada em qualquer aplicativo de fluxo de trabalho ou modelo de implantação e proporcionar a integração de dados de qualquer fonte através do uso de ferramentas do tipo “arrasta e puxa” (PENTAHO, 2017)

A versão 7.1 da suíte de ferramentas oferecida pela *Pentaho* oferece diversos recursos para *Business Intelligence*, entre eles: elaboração de relatórios, tabelas dinâmicas de OLAP, dashboards, etc. Todos os softwares disponíveis pela empresa são de código aberto, o que garante total liberdade ao usuário para usar, distribuir, estudar e, até mesmo, modificar o código se necessário (BOUMAN e DONGEN, 2009).

Distribuídos em duas versões: comercial (*Enterprise*) e comunitária (*Community*), a versão *Enterprise* não possui seu valor de comercialização divulgado no site da empresa. Desta forma, é necessário que o usuário preencha um formulário *online* e aguarde o contato da empresa. Ao ser contatado, o suporte da Pentaho realiza uma configuração para demonstração personalizada com base no caso de uso informado pelo cliente. Após instalação, a empresa oferece um período de 30 dias de uso gratuito para teste do produto. Dependendo do plano contratado, o usuário além de ter acesso à documentação exclusiva e ter em mãos um ambiente

de BI configurado de acordo com o seu modelo de negócio, poderá ter também acesso a outras vantagens, como: gerenciamento de incidentes *on-line*, assinatura de notificação, suporte 24 horas/dia via site ou telefone, orientações individuais e discussões sobre práticas recomendadas, convites para eventos exclusivos, sessões técnicas programadas com especialistas, entre outras (PENTAHO, 2017).

A outra versão, a *Community*, é distribuída gratuitamente sobre os termos da licença GNU *General Public Licence* Versão 2 (GPLv2) e pode ser baixada diretamente através do site da empresa. Apesar de não possuir todo o suporte ofertado pela versão comercial, o conjunto de ferramentas para BI da *Pentaho* conta com praticamente as mesmas ferramentas distribuídas na versão *Enterprise* e material de apoio produzido e divulgado pelos membros dos grupos de usuários espalhados por todo o mundo (PENTAHO, 2017). No Brasil, o principal grupo de usuários da comunidade *Pentaho* conta, atualmente, com 2003 membros que compartilham documentação, recursos e troca de experiência além de promoverem todos os anos o *Pentaho Day*, um evento dedicado à apresentação e troca de experiências sobre desenvolvimento *Open Source* na plataforma *Pentaho* (GRUPO DE USUÁRIOS PENTAHO BRASIL, 2017 e PENTAHO BRASIL, 2017).

A suíte de aplicativos *Pentaho* é composta por 8 (oito) ferramentas capazes de atender todo o processo de BI, desde a extração de dados das mais diferentes bases até a apresentação de relatórios e gráficos. Essas ferramentas podem ser baixadas em conjunto todas de uma vez ou, caso o usuário prefira, é possível realizar a transferência individual de cada aplicação. De acordo com informações do site da empresa, as 4 (quatro) ferramentas mais baixadas são: (i) *Business Analytics Platform* – uma ferramenta de análise com recursos avançados que vão desde a criação de relatórios básicos até modelagem preditiva dos dados; (ii) *Pentaho Data Integration* – também conhecida como *Kettle* ou *Spoon*, essa ferramenta oferece recursos de extração, transformação e carga de dados; (iii) *Report Designer* – uma ferramenta gráfica para geração de relatórios de dados em vários formatos, entre eles PDF, XML, CSV, Planilhas de Excel, entre outros; e (iv) *Market Place* – oferece acesso aos *plugins* desenvolvidos pela comunidade *Pentaho*. Além dessas, ainda é possível encontrar diversas outras ferramentas como o *Aggregation Designer* – que fornece uma interface simples para criação e implementação de tabelas agregadas para melhorar o desempenho de cubos OLAP, o *Schema Workbench* – que permite criar e testar esquemas de cubos OLAP através de uma interface gráfica, e *Metadata Editor* – que permite a criação de relatórios e domínios de metadados e modelos de dados relacionais. A plataforma *Pentaho* conta ainda com uma ferramenta de *Big Data* conhecida como *Hadoop Schims* que facilita o suporte de um amplo conjunto de

distribuições *Hadoop* e atualizações de versões através de uma camada de abstração (PENTAHO COMMUNITY, 2017).

Para realização deste trabalho serão utilizados apenas três das oito ferramentas citadas, que serão apresentadas mais detalhadamente a seguir, o Pentaho Data Integration, o *Mondrian Schema Workbenck* e *Business Analytics Plataform*.

3.1.1 Pentaho Data Integration

O *Pentaho Data Integration* (PDI) é formado por três ferramentas: a IDE de integração gráfica, conhecida como *Spoon*, a ferramenta de linha de comando para execução de *jobs*, chamada *Kettle* e a ferramenta de linha de comando para execução de transformações, chamada *Pan*. Sua principal funcionalidade é a realização da Extração, Transformação e Carga de dados, ou ETL (do inglês *Extracition, Transformation and Load*).

Através das transformações, compostas por uma sequência de passos (*steps*) e das rotinas automatizadas denominadas *jobs*, o usuário pode extrair e tratar dados de diversas fontes e armazená-los em diversos outros formatos, entre eles, base de dados, planilhas eletrônicas, arquivos de texto.

O motor de integração que compõe a ferramenta é responsável por interpretar e executar os *jobs* e as transformações, sendo fundamental para seu funcionamento. Apesar do ambiente ser formado por várias ferramentas, todas elas encontram-se disponíveis para download em um único arquivo zipado na página do projeto *Pentaho* em <https://sourceforge.net/projects/pentaho> (BOUMAN e DONGEN, 2009).

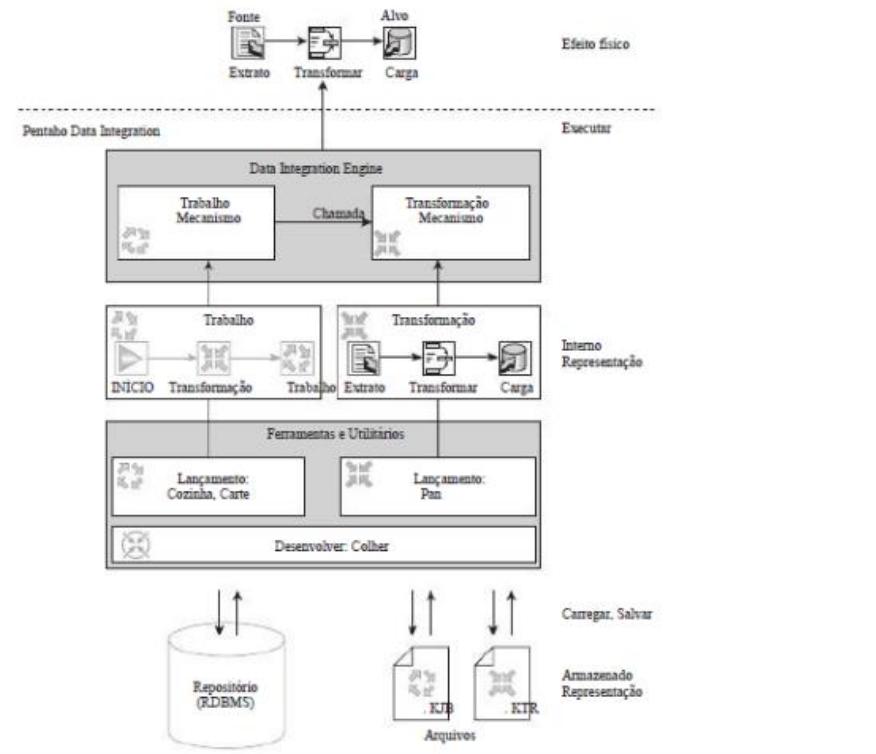


FIGURA 5 - Arquitetura da ferramenta Pentaho Data Integration.

FONTE: Bouman e Dongen, 2009.

Neste trabalho o PDI foi utilizado para realizar a limpeza e integração dos dados extraídos da Plataforma Stela Experta e da Planilha Eletrônica do COPE campus Palmas.

3.1.2 Mondrian Schema Workbenck

O *Mondrian Schema Workbench* é a ferramenta que através de uma interface gráfica gera os códigos XML que compõem os cubos OLAP. O mapeamento das dimensões e tabelas fatos no formato de cubo é feito essencialmente através do *Mondrian*.

Sua principal funcionalidade é executar consultas MDX (do inglês, *Multidimensional Expressions*), lendo dados de uma base relacional e os apresentando em formato multidimensional (MONDRIAN, 2017). MDX é um padrão criado e implementado pela Microsoft para realizar consultas OLAP (BOUMAN e DONGEN, 2009).

Neste trabalho ele foi utilizado para criação e teste dos esquemas de cubos OLAP de produção científica e projetos do campus Palmas. Os cubos foram criados para responder questões como: “Qual quantidade de projetos publicados por pesquisadores da área de ciências

exatas e da terra, no ano de 2014, 2015 e 2016?”, “Qual a média anual de publicação de pesquisadores por colegiado?”, entre outras.

3.1.3 *Business Analytics Platform*

A Plataforma Analítica de Negócios da suíte Pentaho (do inglês, *Business Analytics Platform*), é formada por um conjunto de componentes que auxiliam os usuários na criação e visualização de relatórios e painéis informativos. Assim, é possível que as consultas dos cubos OLAP criados a partir do *Mondrian Scheme Workbench* sejam representados através de relatórios, gráficos ou painéis criados pelo próprio usuário, através de uma interface web (PENTAHO, 2017).

Neste trabalho foi utilizado para aumentar o desempenho e a facilidade de uso, o *plugin Saiku Business Analytics* em conjunto com a plataforma do *Business Analytics*.

3.1.4 *Saiku Business Analytics*

O *Saiku Business Analytics* é uma suíte analítica disponível livremente para *download* através do site do fabricante sob a licença Apache podendo ser utilizada livremente em soluções internas.

Sua principal característica é a capacidade em se conectar com diversos servidores OLAP permitindo que seus usuários escolham as medidas e dimensões a serem analisadas e utilizem as operações OLAP, como *slice and dice*, *drill across*, entre outras, para descobrir relacionamentos, oportunidades e problemas através dos dados visualizados, tudo isso através de uma interface intuitiva *look and feel* desenvolvida para internet.

3.2 A PLATAFORMA *STELA EXPERTA*

A plataforma *Stela Expert*a foi criada pela mesma equipe que desenvolveu a Plataforma Lattes para o CNPq e apresenta-se como uma solução na avaliação de desempenho de docentes e discentes em processos internos de distribuição de recursos.

Através do uso da ferramenta, o usuário pode localizar especialistas, gerar relatórios e indicadores da produção científica da instituição da qual faz parte, identificar *gaps* de conhecimento, entre outras consultas. Tudo isso a partir da análise das informações do currículo Lattes e dos grupos de pesquisa do CNPq.

O uso da plataforma é atrelado à um investimento financeiro que contempla uma taxa fixa para implantação da solução e capacitação dos usuários para uso dos sistemas mais uma mensalidade estabelecida de acordo com a quantidade de currículos de docentes e discentes que a instituição pretende gerenciar, essa mensalidade refere-se à licença de uso, suporte, manutenção e acesso à novas versões do sistema que compõe a versão contratada (STELA EXPERTA, 2017).

Com o uso da plataforma *Stela Expert*a foi possível obter as informações dos Currículos Lattes de todos os professores lotados no IFPR campus Palmas.

3.3 O ARMAZENAMENTO DE DADOS NO *POSTGRESQL*

O *PostgreSQL*, criado pelo professor da Universidade da Califórnia em Berkeley, Michael Stonebracker em 1986, é um sistema de banco de dados objeto-relacional de código aberto, desenvolvido para todos os principais sistemas operacionais com recursos sofisticados como Controle de Concorrência de Multi-Versão (MVCC) e recuperação de ponto-a-tempo. Algumas de suas características são a oferta de suporte pela comunidade; programas de treinamentos mais econômicos, gerenciáveis e práticos quando comparados com outros bancos de dados proprietários; confiabilidade; disponibilização do código-fonte de forma gratuita, tornando-o extensível; ferramentas de design e administração de banco de dados GUI; entre outros (POSTGRESQL, 2017).

A utilização do *PostgreSQL* nesse trabalho se dá na criação da *Staging Area*, um armazenamento intermediário dos dados com o objetivo de facilitar a integração dos dados do

ambiente operativo antes da persistência no *Data Warehouse* e no desenvolvimento do próprio *Data Warehouse*.

3.4 BASES DE DADOS EXISTENTES DO COPE E CURRÍCULOS LATTES

Os dados que compõem o *Data Warehouse* e visualizados nos cubos OLAP são oriundos de dois sistemas: (i) a base de dados do COPE do Instituto Federal do Paraná – campus Palmas, e (ii) a base de Currículos Lattes do CNPq.

A base de dados do COPE consiste em uma planilha eletrônica, disponibilizada e armazenada em nuvem pelo Google Docs. Essa planilha pode ser acessada através de qualquer dispositivo com acesso à internet desde que o usuário tenha permissão e faça parte da rede de compartilhamento do COPE, que, no período de desenvolvimento deste trabalho, é composta pela direção de ensino, pesquisa e extensão do campus, coordenação de ensino, coordenação de pesquisa e extensão e presidência do COPE. Os cadastros na planilha são realizados mensalmente e se dividem em quatro tipos: projetos de pesquisa, projetos de extensão, projetos de ensino e projetos de inovação, que estão separados entre os projetos que estão em andamento e os que já foram encerrados. As informações cadastradas nesta planilha dizem respeito ao título do projeto, número de cadastro no sistema interno da instituição (SIPAC), número de cadastro no COPE, nome do pesquisador que coordena o projeto e quantidade de horas, nome dos pesquisadores que colaboram com o projeto e quantidade de horas, número de discentes que atuam no projeto, se o projeto precisou ser submetido a alguma Comissão de Ética em Pesquisa (CEP) ou Comissão de Ética na Utilização Animal (CEUA), data de submissão do projeto ao COPE, data de aprovação do projeto no COPE, situação do projeto perante ao COPE, data parcial da entrega do relatório parcial, situação do relatório parcial, data de previsão para entrega do relatório final e situação do relatório final.

A segunda fonte de dados para o desenvolvimento deste trabalho trata-se dos Currículos Lattes que representam um histórico das atividades científicas acadêmicas e profissionais de pesquisadores cadastrados na Plataforma Lattes. Os Currículos Lattes foram projetados para mostrar informação pública e individual de cada usuário cadastrado na plataforma (MENA-CHALCO e JUNIOR, 2013, *apud* AMORIN, 2003). Nesse sentido, através do cadastro desses currículos a plataforma Lattes é capaz de fornecer dados estratégicos que são utilizados não somente pelo CNPq, mas também por outros órgãos públicos e privados em

suas atividades de planejamento e gestão, como por exemplo o Ministério de Ciência e Tecnologia que faz uso da plataforma para a formulação de suas políticas (CNQP, 2017).

3.5 METODOLOGIA DO TRABALHO

O presente trabalho trata de um estudo de caso de natureza descritiva, uma vez que busca apresentar dados através de sua descrição e quantificação. Para alcançar os objetivos propostos foram definidas várias etapas que juntas integram o trabalho desenvolvido.

A primeira etapa consistiu no levantamento de referencial teórico (leis, normas, regulamentos) sobre os Institutos Federais de Educação e o processo de atividades de pesquisa e extensão proposto por eles. Uma vez conhecido o processo e como é o fluxo no campus estudado, foi necessário um embasamento teórico (livros, artigos e internet) sobre técnicas de *Business Intelligence*, a tecnologia de *Data Warehouse* e Processamento Analítico.

Conhecendo as técnicas mais utilizadas, o passo seguinte foi a aplicação da metodologia estudada através do levantamento das bases de dados existentes e da concepção de uma proposta de modelo de DW de Produção Científica para o IFPR campus Palmas.

O trabalho utilizou as bases de dados do COPE do IFPR campus Palmas e dos currículos Lattes do CNPq dos docentes do campus, considerando os dados do período compreendido entre janeiro de 2014 e dezembro de 2016. Os dados foram extraídos para uma área intermediária, a ODS, onde foram tratados e organizados. Da ODS, os dados foram armazenados no DW. Ambos, ODS e DW persistem os dados no banco de dados *PostgreSQL*.

Tendo implementado o modelo de DW planejado, realizou-se a implementação do cubo OLAP para a manipulação e análise dos dados armazenados no DW sob múltiplas perspectivas.

Ao final, discute-se os resultados obtidos e apresentam-se as conclusões e sugestões de trabalhos futuros.

A metodologia utilizada na pesquisa é representada a seguir (FIGURA 6).



FIGURA 6 - Metodologia da Pesquisa

4. BI DA PRODUÇÃO CIENTÍFICA DO IFPR CAMPUS PALMAS

Neste capítulo são apresentados os métodos e as ferramentas utilizadas durante o processo de criação do BI para identificar o perfil da Produção Científica do IFPR campus Palmas. Desta forma, na primeira sessão é apresentada a modelagem aplicada para concepção do DW; a segunda sessão descreve o processo de ETL aplicado para criação de uma área intermediária para armazenamento dos dados; a terceira sessão apresenta os passos aplicados para criação do DW; na quarta sessão são apresentados os passos para criação e publicação dos cubos OLAP e, por fim, na quinta sessão é apresentado a visualização das consultas OLAP através do Painel de Usuários da *Pentaho*.

4.1 MODELAGEM ENTIDADE-RELACIONAMENTO (ER) E MULTIDIMENSIONAL DO DATA WAREHOUSE

O primeiro passo para a criação da solução em BI consistiu em entender o funcionamento do sistema e seu fluxo de dados no campus Palmas. Em seguida, identificou-se quais tipos de perguntas pretendia-se responder com o BI. Após levantadas as informações, foi realizado a modelagem do repositório único de dados e informações, o *Data Warehouse*, citado por diversos autores como o coração do BI (CECI, 2014, *apud* REGINATO e NASCIMENTO, 2007).

Para a implementação do DW, seguiu-se a orientação de Machado (2013) ao realizar duas modelagens: (i) a ER (Entidade-Relacionamento), tradicionalmente conhecida entre os desenvolvedores, o diagrama ER é uma ferramenta que ajuda a análise de requisito de negócio e a essência técnica da arquitetura de DW; (ii) a Modelagem Multidimensional, é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócio.

O modelo Entidade-Relacionamento (ER), representado pela Figura 7, é composto por três representações de dados: entidade, relacionamentos e atributos. Neste modelo ER são apresentadas sete classes: pesquisador, colegiado, área e subárea de conhecimento, evento, publicação e projeto, que estão interligadas através dos relacionamentos representados por

losangos. Desta forma é possível identificar a relação que existe entre cada objeto que compõe o sistema.

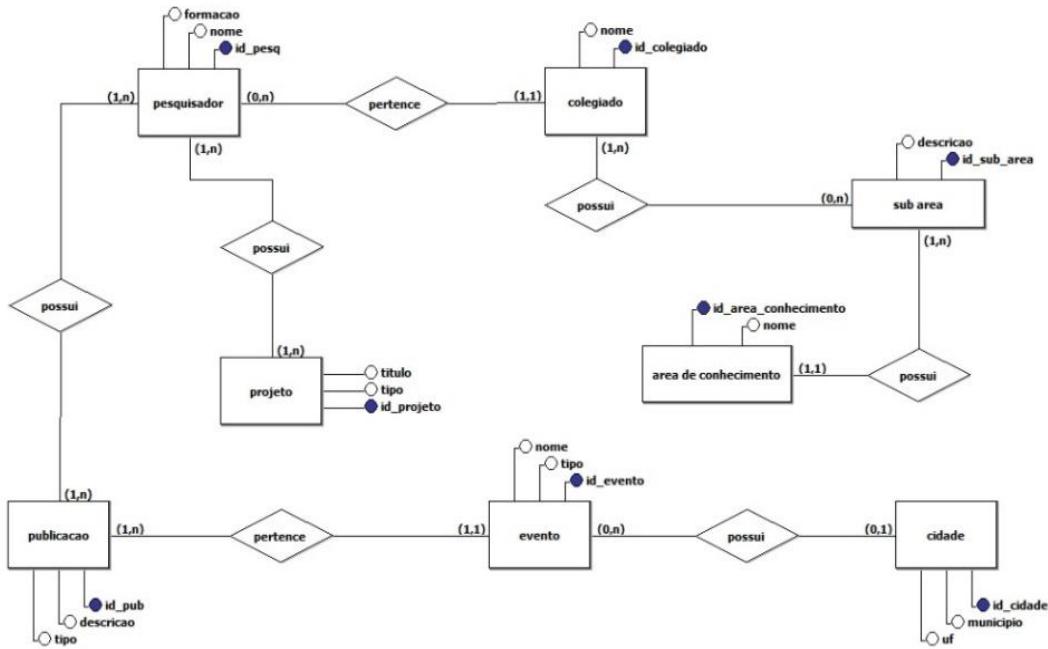


FIGURA 7 - Modelo Entidade-Relacionamento do DW da Produção Científica.

O modelo dimensional é composto por uma tabela com várias chaves, chamada de tabela fato, e um conjunto de tabelas menores nomeadas tabelas de dimensão. Cada tabela de dimensão possui uma chave primária correspondente a uma das chaves estrangeiras que compõe a tabela fato. Esta ligação entre as tabelas de dimensão à tabela faz alusão a forma parecida com uma estrela, por esse motivo, esse modelo é também conhecido como Modelo Estrela (do inglês, *Star Schema*) (KIMBALL, 1998).

A modelagem do DW da Produção Científica desenvolvida neste trabalho, conforme ilustra a Figura 8, é composta por nove dimensões: a dimensão tempo, publicação, projeto, evento, cidade, pesquisador, colegiado, área de conhecimento e subárea de conhecimento; e duas tabelas fato: publicação e projetos, composta pelas chaves estrangeiras referenciando cada dimensão, uma chave primária e a medida quantidade.

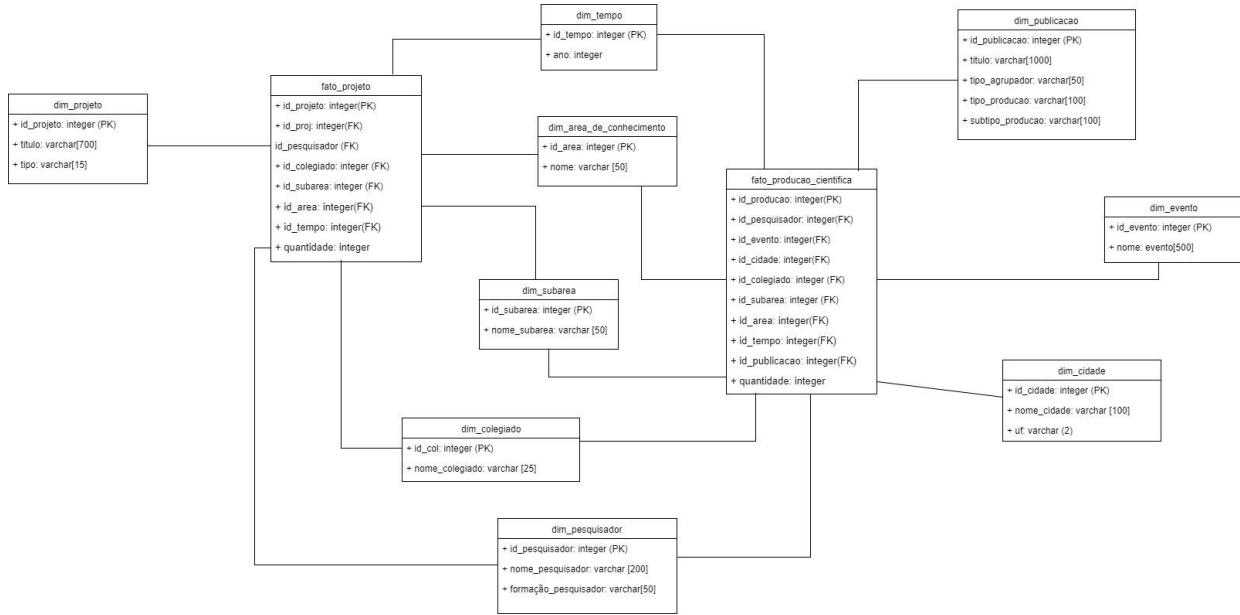


FIGURA 8 - Modelagem Dimensional do DW através do esquema estrela.

Através desse modelo é possível responder questões como: “Qual quantidade de projetos publicados por pesquisadores da área de ciências exatas no ano de 2016?”, “Qual a média anual de publicação dos pesquisadores por colegiado?”, “Quantas publicações teve determinado professor?”, entre outras.

4.2 O PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DE DADOS

Com o DW modelado iniciou-se o processo de extração, transformação e carga de dados. O primeiro passo foi a identificação das fontes de dados. Foram extraídos e integrados dados de quatro planilhas eletrônicas diferentes, sendo elas: (i) Tabela de Áreas de Conhecimento/Avaliação da CAPES (CAPES, 2017); (ii) Relação das cidades e unidades federativas (IBGE, 2017); (iii) Relação dos Projetos de Pesquisa, Extensão, Ensino e Inovação mantido pelo COPE; e (iv) Currículos Lattes dos Pesquisadores do IFPR campus Palmas, conforme ilustra a Figura 9.

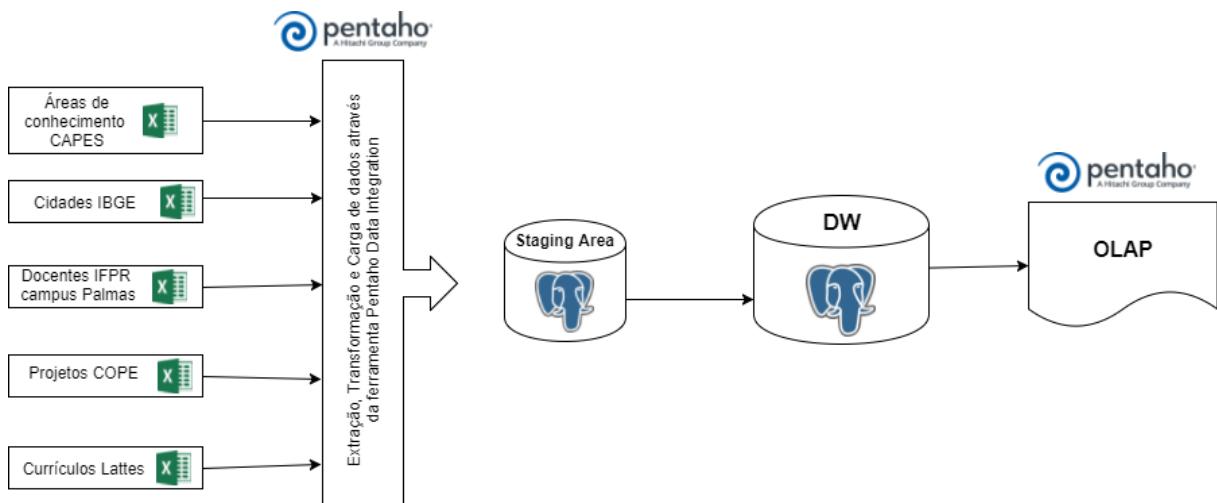


FIGURA 9 - Arquitetura do Business Intelligence da Produção Científica do IFPR campus Palmas.

A Tabela de Áreas de Conhecimento/Avaliação da CAPES é disponibilizada em dois formatos: PDF e DOC. Após o *download* do arquivo em formato PDF, as informações foram reorganizadas em dez abas de uma planilha eletrônica (FIGURA 10) onde a primeira aba contém o código e descrição das nove grandes áreas de conhecimento e as demais são compostas pelas subáreas de cada uma das grandes áreas.

	area
1	cod capes
2	10000003 Ciências Exatas e da Terra
3	20000006 Ciências Biológicas
4	30000009 Engenharias
5	40000001 Ciências da Saúde
6	50000004 Ciências Agrárias
7	60000007 Ciências Sociais Aplicadas
8	70000000 Ciências Humanas
9	80000002 Linguística, Letras e Artes
10	90000005 MULTIDISCIPLINAR
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	

	sub area
1	cod capes
2	10100008 Matemática
3	10200002 Probabilidade e Estatística
4	10300007 Ciência da Computação
5	10400001 Astronomia
6	10500006 Física
7	10600000 Química
8	10700005 GeoCiências
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	

FIGURA 10 - Áreas e Subáreas do Conhecimento de acordo com a CAPES.

A integração dos dados contidos na planilha à base de dados *PostgreSQL* foi realizada através da ferramenta de integração de dados *Pentaho Data Integration*.

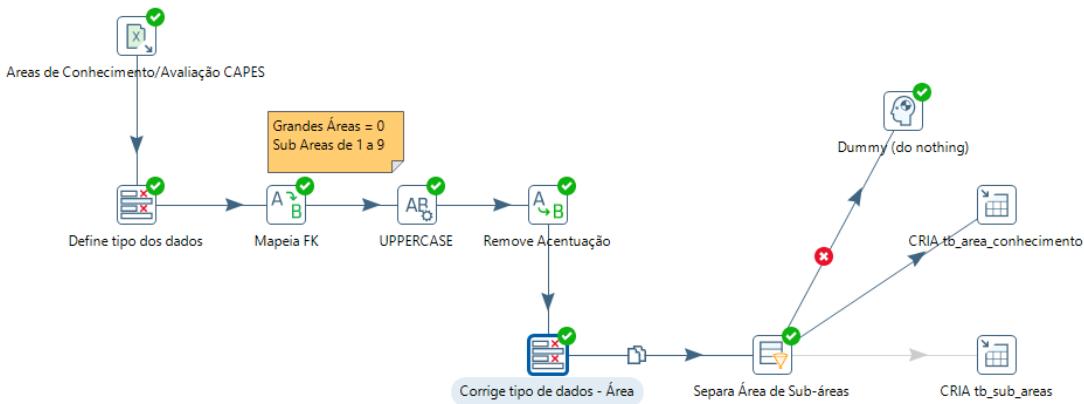


FIGURA 11 - Processo de extração de dados e criação das tabelas de área e subáreas de conhecimento da CAPES.

O processo de extração de dados para a criação da tabela de Áreas de Conhecimento e Subáreas de Conhecimento da CAPES (FIGURA 11) inicia com o *step Microsoft Excel Input*, que permite realizar a leitura de dados de uma ou mais planilha eletrônica. Após ler informações da planilha e antes de carregá-las na *staging area* foi realizado o tratamento dos dados, que consistiu em definir corretamente o tipo dos dados de cada coluna (*integer* para o código da CAPES e *string* para a descrição das áreas e subáreas de conhecimento), mapear as chaves estrangeiras, deixar todas as letras em caixa alta e remover a acentuação das descrições. Antes de persistir os dados na *staging area*, a coluna com o código capes foi descartada, uma vez que as consultas acontecerão através das descrições ou chaves primárias.

Como as áreas e subáreas de conhecimento foram extraídas da mesma planilha, os dados foram tratados na mesma transformação (FIGURA 11). Devido a isto, a transformação precisou ser executada duas vezes, onde no primeiro momento as informações foram persistidas na tabela de área de conhecimento e posteriormente na tabela de subáreas. Para possibilitar essa execução, as áreas de conhecimento foram mapeadas com o valor 0 e as subáreas na ordem de 1 a 9 através do *step Value Mapper* (FIGURA 12).

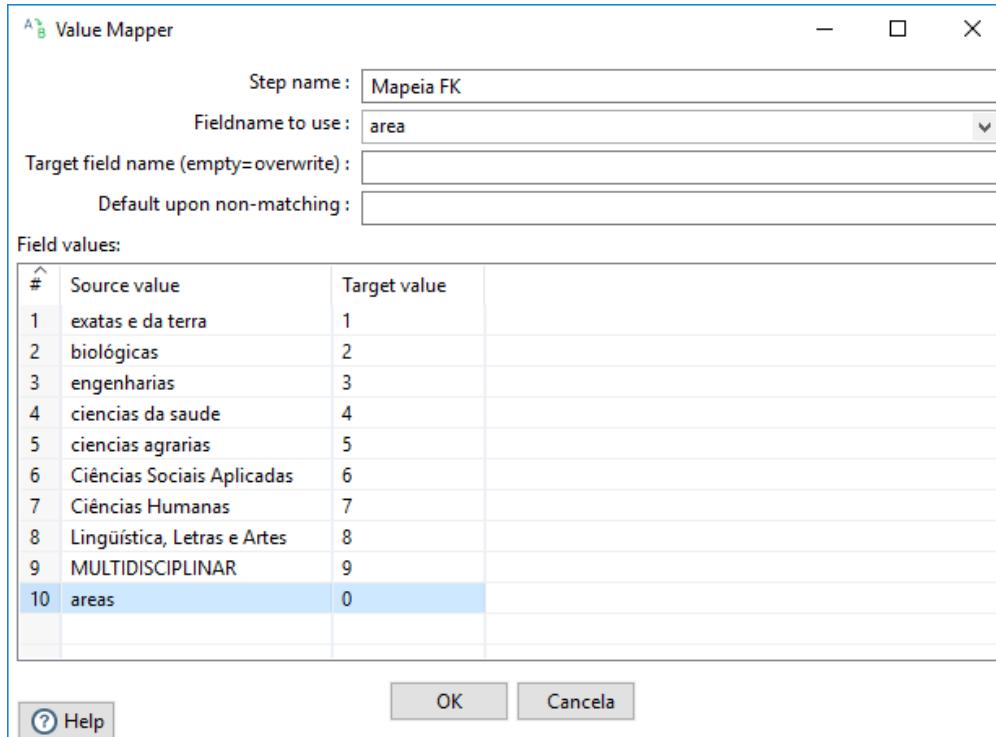


FIGURA 12 - Parâmetros de configuração para mapear identificação das áreas e subáreas do conhecimento, através do step *Value Mapper*.

Com as áreas e subáreas de conhecimento devidamente mapeadas, foi utilizado o *step Filter Row* para identificar e persistir os dados na base. Na primeira execução da transformação, o filtro “Separa Área de Sub-áreas” identificou os campos de área igual a 0 e os persistiu na tabela de áreas de conhecimento através do *step table output* (FIGURA 13).

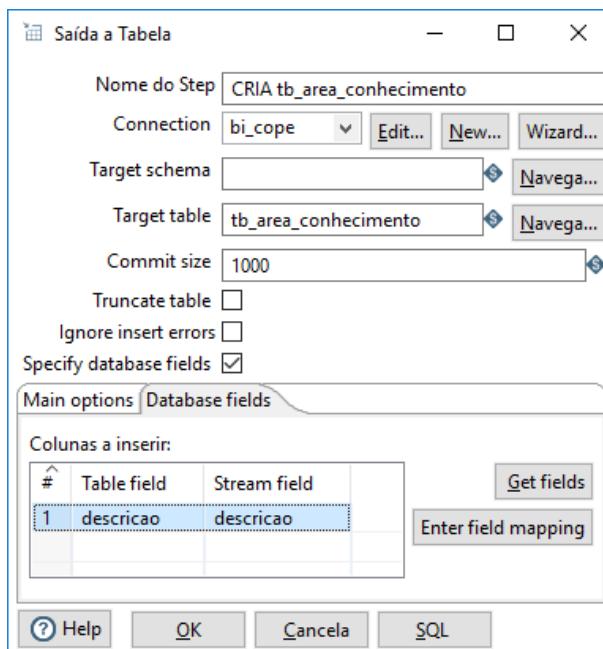


FIGURA 13 - Parâmetros de configuração do *step table output* que persiste os dados extraídos da planilha eletrônica em uma tabela na base de dados.

Com a tabela área de conhecimento criada, a transformação foi executada novamente, porém, desta vez, o filtro “Separa Áreas de Sub-áreas” foi configurado para identificar os registros onde a área é diferente de 0. Com os dados oriundos deste processo, foi possível utilizar o *step table output* para criar a tabela de subáreas de conhecimento composta por dois campos do tipo *integer* (id e área) e um campo do tipo *string* (FIGURA 14). O campo id trata-se de uma chave-primária criada sequencialmente, de forma automática, pelo próprio Sistema Gerenciador de Banco de Dados (SGBD). O campo “area” trata-se de uma chave-estrangeira que referencia um dos nove campos da tabela de áreas de conhecimento criada a partir da primeira execução da transformação. O campo *descrição* apresenta o nome de cada subárea armazenada.

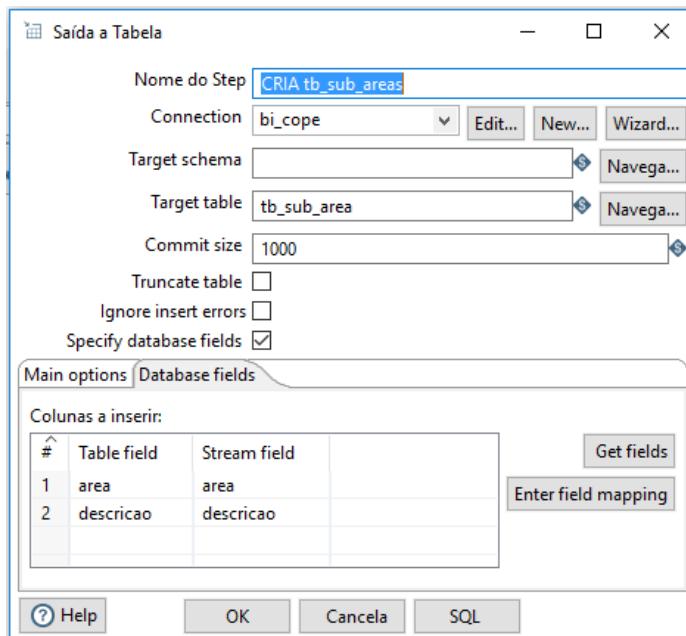


FIGURA 14 - Parâmetros de configuração do *step table output* responsável por criar a tabela de subáreas de conhecimento.

Com a criação e inserção de dados nas tabelas de áreas e subáreas de conhecimento, foi possível criar a tabela com a relação dos quinze colegiados existentes no IFPR campus Palmas. Por não possuir um número elevado de registros, a criação desta tabela aconteceu de forma manual na base de dados, utilizando SQL (FIGURA 15).

```

bi_cope on postgres@PostgreSQL 9.6
1 CREATE TABLE public.tb_colegiado
2 (
3     id_col bigint NOT NULL,
4     descricao character varying(40)NOT NULL,
5     sub_area bigint NOT NULL,
6     CONSTRAINT pk_tb_colegiado PRIMARY KEY (id_col),
7     CONSTRAINT tb_colegiado_sub_area_fkey FOREIGN KEY (sub_area)
8 )
9
10 insert into tb_colegiado (descricao, sub_area) values ('ADMINISTRAÇÃO',51), ('ARTES VISUAIS', 74),
11 ('CIÉNCIAS BIOLÓGICAS', 8), ('CIÉNCIAS CONTÁBEIS', 51), ('DIREITO', 50), ('EDUCAÇÃO FÍSICA', 42),
12 ('ENFERMAGEM', 37), ('ENGENHARIA AGRONÔMICA', 43), ('ENGENHARIA CIVIL', 21), ('FARMÁCIA', 36),
13 ('LETRAS PORTUGUÊS/INGLÊS', 73), ('PEDAGOGIA', 69), ('QUÍMICA',6), ('SISTEMAS DE INFORMAÇÃO', 3),
14 ('MULTIDISCIPLINAR', 75)

```

FIGURA 15 - Script SQL utilizado para criação e preenchimento da tabela de colegiado na base de dados.

Em seguida foi realizada a criação e carga de dados na tabela com informações dos docentes do campus. Para obter a relação dos docentes que atuam no IFPR campus Palmas foi feito uma busca por Pessoas na plataforma *Stela Experta* através dos campos Pessoas, Lotação Institucional, Enquadramento Funcional (Professor do Ensino Básico, Técnico e Tecnológico), Regime de Trabalho (Docente – Dedicação Exclusiva; Docente – 40 horas semanais; Docente – 20 horas semanais) e Vinculo Institucional (Ativo Permanente). A busca retornou em uma planilha eletrônica (FIGURA 16) contendo 132 registros com as seguintes informações: Nome, Endereço do Currículo Lattes, Área de Titulação, Titulação, Instituição da Titulação, Ano da Titulação, Lotação Institucional, Bolsista PQ/DT (CNPq), Vinculo Institucional, Ano de Ingresso na Instituição, Perfil na Instituição, Regime de Trabalho, Titulação Máxima Informada pela Instituição, Data de Atualização do Currículo e 3 (três) Palavras-Chave.

The screenshot shows an Excel spreadsheet with the following columns:

	A	B	C	D	E	F	G	H
	Nome	Currículo Lattes	Ano da titulação máxima informada no CV-Lattes	Titulação máxima informada no CV-Lattes	Instituição da titulação máxima	Ano da titulação máxima	Vinculo Institucional	
1	Marília Cezarina Mendes Sandri	http://lattes.cnpq.br/953046962414951		Doutorado	UEM	2016	CampusPalmar	
2	Ricardo Aparecido Ferreira	http://lattes.cnpq.br/2122320784506333		Feminino	Mestrado acadêmico	UNICENTRO	CampusPalmar	
3	Jacobi dos Santos Etival	http://lattes.cnpq.br/32621666415322	Lerato	Doutorado	UNESP	2015	CampusPalmar	
4	Janaina Lopes de Souza	http://lattes.cnpq.br/950704703503530		Educação	Especialização	UNIVERSELA	CampusPalmar	
5	Leutânia Paludo	http://lattes.cnpq.br/679464662782704		Cáncica da Computação	Mestrado acadêmico	UNIVALI	CampusPalmar	
6	Fábio Manchoso Pützenreuter	http://lattes.cnpq.br/750422953276748	Lerato	Doutorado	UFSC	2014	CampusPalmar	
7	Rafael Henrique Sartori	http://lattes.cnpq.br/6533267542400017		Educação	Mestrado acadêmico	UEM	CampusPalmar	
8	Christiano Santos Rocha-Pita	http://lattes.cnpq.br/75653325782521		Agronomia	Doutorado	UFSC	2012	CampusPalmar
9	Vagner Scarami	http://lattes.cnpq.br/8464055001366515		Cáncica da Computação	Mestrado acadêmico	FACAMP	CampusPalmar	
10	Lilian Moura de Lima Spagnoli	http://lattes.cnpq.br/036130052020204		Enfermagem	Doutorado	UFSC	2016	CampusPalmar
11	Diego Benelli Cochere	http://lattes.cnpq.br/953456404733278		Educação	Especialização	FEISL	2017	CampusPalmar
12	Rodrigo Ezequiel de Almeida	http://lattes.cnpq.br/97144848415166		Feminino	Mestrado acadêmico	UFSC	2003	CampusPalmar
13	Aurea e Isobel Stuckmann	http://lattes.cnpq.br/45365081200533		Engenharia de Produção	Mestrado acadêmico	UFSC	2001	CampusPalmar
14	Carolina Ribeiro Noronha de Souza	http://lattes.cnpq.br/3411794300675760		Biofísica Geral	Doutorado	USP	2014	CampusPalmar
15	Ana Carolina Vieira de Carvalho	http://lattes.cnpq.br/6347674032962739		Administrador	Mestrado acadêmico	UEL	2012	CampusPalmar
16	Laíla Maria Gralido	http://lattes.cnpq.br/017717830000005051		Odonto	Doutorado	UNB	2017	CampusPalmar
17	Douglas Silva do Prado	http://lattes.cnpq.br/543095795179421		Educação	Especialização	FAMPER	2010	CampusPalmar
18	Alcine Rocha Borges	http://lattes.cnpq.br/12945500125642		Química	Doutorado	UFSC	2013	CampusPalmar
19								

FIGURA 16 - Planilha eletrônica contendo o resultado da Busca Pessoas utilizando a plataforma *Stela Experta*.

O primeiro passo após acessar a planilha eletrônica disponibilizada pela Plataforma *Stela Expert* foi realizar a análise e limpeza dos dados. No campo Currículo Lattes optou-se por manter apenas a ID (uma sequência numérica de 16 dígitos apresentada no final de cada endereço eletrônico) de cada currículo e não o endereço eletrônico completo da Plataforma Lattes. Outra alteração foi com relação as palavras-chave, na planilha original elas eram apresentadas em três colunas diferentes que foram concatenadas em um único campo separadas por “;”. Após tratadas, as informações: Nome, ID Lattes, Colegiado, Área de Titulação, Titulação, Regime de Trabalho e Palavra-Chave foram salvas na tabela pesquisador da base de dados através de uma nova transformação feita no PDI (FIGURA 17).



FIGURA 17 - Extração, transformação e carga dos dados para a criação da tabela de pesquisadores na base de dados.

Conforme apresentado na modelagem multidimensional, um dos pontos que este trabalho busca auxiliar é na identificação de publicações realizadas em outros municípios/estados brasileiros e, a partir daí, apontar quais são eles e com qual frequência ocorrem. Para isso, criou-se uma tabela na base de dados com a relação das cidades brasileiras e seus respectivos estados. Para tanto, utilizou-se a tabela de estados e municípios disponibilizada na internet através do site do Instituto Brasileiro de Geografia e Estatística - IBGE (IBGE, 2017).

A partir da planilha eletrônica das cidades do IBGE foi possível realizar a persistência dos dados na *staging area* através da tabela de cidade utilizando o PDI (FIGURA 18). Após realizar a leitura dos dados com o *step Microsoft Excel Input* os dados foram colocados em ordem crescente pelo número de identificação de cada município através do *step Sort Row*. Para que todos os dados fossem padronizados e deixados em caixa alta, utilizou-se o *step String Operations*. O *step Select Values* foi utilizado para definir o tipo de dados de cada coluna da planilha - *string* para a descrição do município e *UF* e *integer* para o número de identificação. Por fim, a criação da tabela cidades e inserção de dados aconteceu através do *step Table Output*.

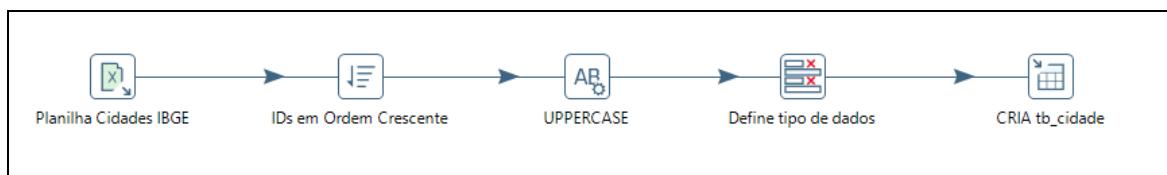


FIGURA 18 - Extração, transformação e carga dos dados para a criação e inserção de dados na tabela de cidades.

Os dois últimos processos de ETL são referentes as tabelas de publicações e projetos, onde: as publicações são oriundas das informações extraídas do Currículo Lattes de cada pesquisador e estão organizadas de acordo com os indicadores de pesquisa proposto pelo CNPq: produção bibliográfica, participação em eventos, produção técnica, produção artística, entre outros. Os dados foram obtidos através da Plataforma *Stela Expert*a após realizar uma busca por Produções, com os seguintes filtros: Tipo da Produção (Produção Bibliográfica, Evento, Produção Técnica, Outro Tipo de Produção e Produção Artística/Cultural), Ano da Produção, Lotação Institucional do Autor no Período de vinculo institucional do autor. Foram encontrados 1876 agrupamentos de produções com essas características. O segundo ETL é realizado sobre os projetos de pesquisa, ensino-aprendizagem, extensão e inovação oriundos da planilha eletrônica mantida pelo COPE de onde foram extraídas informações dos 235 cadastrados no período de 2014 a 2016.

Antes de realizar o processo de ETL nas planilhas, foi preciso analisar e corrigir algumas informações manualmente. Na planilha de publicações, os nomes dos autores tiveram de ser alterados respeitando a ortografia dos dados persistidos na tabela de pesquisadores anteriormente para que o PDI fosse capaz de identificar seus pares. Apesar da tabela não possuir colunas específicas com informações da cidade, unidade federativa e nome do evento onde ocorreu a publicação, foi possível obter estes dados e registrá-los em colunas específicas, na maior parte dos registros, através da coluna ABNT. Nesta análise, 621 registros duplicados foram identificados e descartados no decorrer da conferência. Por fim, a planilha de publicações passou a ter efetivamente 1254 publicações datadas entre 2014 e 2016.

Ao fim deste processo, pode-se realizar o ETL para carregar as informações da planilha de publicações para a *staging área*. A partir da coluna com o nome dos eventos de cada publicação, criou-se a tabela de eventos na *staging área*. A leitura e seleção dos nomes dos eventos aconteceu por meio dos *steps Microsoft Excel Input* e *Select Values*. O *step Sort Row* disponibilizou os eventos em ordem alfabética, o *step Value Mapper* corrigiu a ortografia da descrição dos nomes, o *step Unique Rows* excluiu os registros duplicados e o *step Table Output* criou a tabela de eventos na *staging area* e persistiu os dados (FIGURA 19).

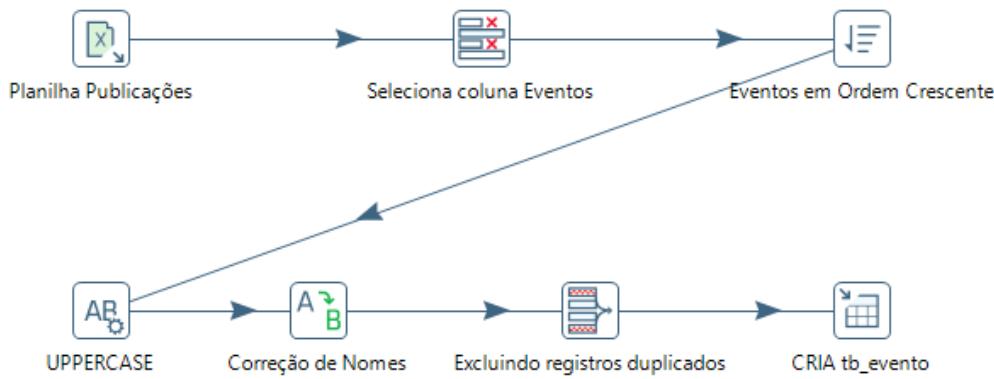


FIGURA 19 - Extração, transformação e carga dos eventos para a *staging area*.

Foram realizadas três processos de ETL antes de persistir os registros da planilha de publicações no banco de dados: (i) substituição do nome do evento pelo número de identificação da tabela de eventos existente na *staging área* (FIGURA 20); (ii) substituição do nome do município pelo número de identificação da tabela de cidades da *staging area*; e (iii) substituição do nome dos pesquisadores envolvidos na publicação pelo número de identificação da tabela de pesquisadores da *staging area*.

Para substituir o nome pela identificação do evento, o primeiro passo consistiu na padronização dos registros da planilha de publicações. Todos os registros da planilha foram deixados em caixa alta e sem acentuação, através dos *steps Strings Operations* e *Replace in String*. O *step Concat Fields* foi responsável por concatenar as colunas “Palavra-Chave 1”, “Palavra-Chave 2” e “Palavra-Chave 3” uma única coluna. A disponibilização dos registros em ordem alfabética de acordo com o nome do evento foi realizada através do *step Sort Rows*. Com os dados tratados, o *step Filter Rows* foi utilizado identificar os registros com informações de evento e compará-los com os dados da tabela de evento criada na *staging área*. Só então, a descrição do nome do evento pode ser substituída pelo número de identificação que referencia o registro da tabela de evento criada na *staging area*. Após substituição, os registros foram novamente integrados com os dados sem registro de evento, voltando aos 1254 registros originais da tabela de publicação. O resultado deste ETL foi a criação de uma nova planilha eletrônica contendo todos os registros de publicações agora com a coluna de eventos atualizada com o número de identificação de cada evento e não mais a descrição do nome.

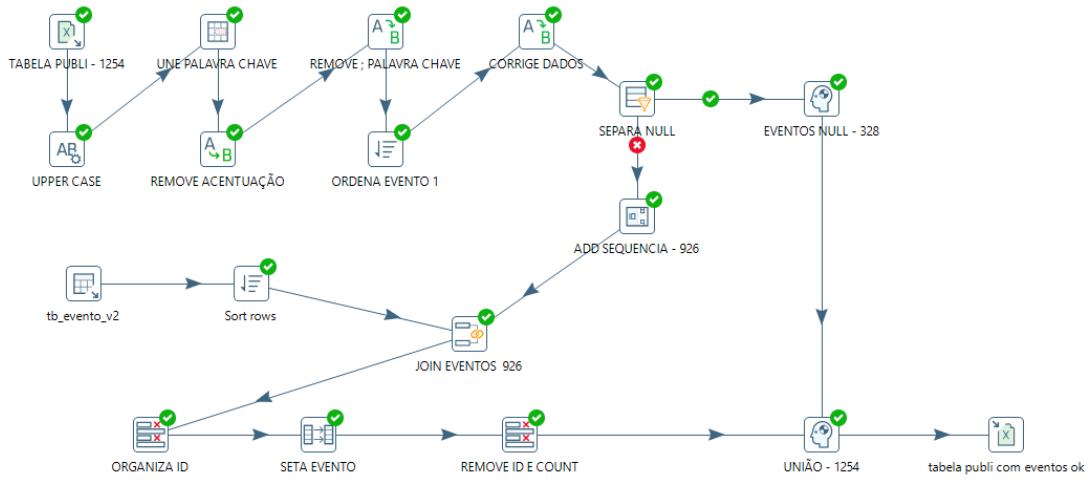


FIGURA 20 - Processo de ETL para substituir o nome pela identificação de cada evento na planilha de publicações.

O processo de ETL que substitui o nome da cidade pelo número de identificação de acordo com os registros da tabela de cidades criada na *staging area* (FIGURA 21) segue a mesma lógica utilizada na substituição dos dados dos eventos, porém neste processo a fonte de dados é a planilha eletrônica resultante do processo de ETL dos eventos. Como o ETL dos eventos padronizou todos os dados da tabela, após o *step Microsoft Excel Input* realizar a leitura da planilha de publicações, a coluna “UF” foi renomeada para “estado” pelo *step Select Value* e os registros colocados em ordem alfabética de acordo com o nome de cada município pelo *step Sort Row*. As publicações com informação de cidade foram separadas das demais através do *step Filter Row* e as colunas cidade e estado foram comparadas e unidas com os registros da tabela de cidades da *staging area* através do *step Join Rows*. Somente assim os dados da coluna cidade puderam ser substituídos pelo número de identificação da tabela de cidades, com o auxílio do *step x*, para então serem novamente integrados com os demais registros da planilha de publicações. Durante todo esse processo de ETL (FIGURA 21), outros 3 casos de registros duplicados foram identificados e descartados da planilha que passou a ser constituída por 1251 registros. Assim como no ETL de eventos, o resultado do ETL das cidades consistiu na criação de uma terceira versão da planilha de publicação, agora com as colunas eventos e cidades atualizadas.

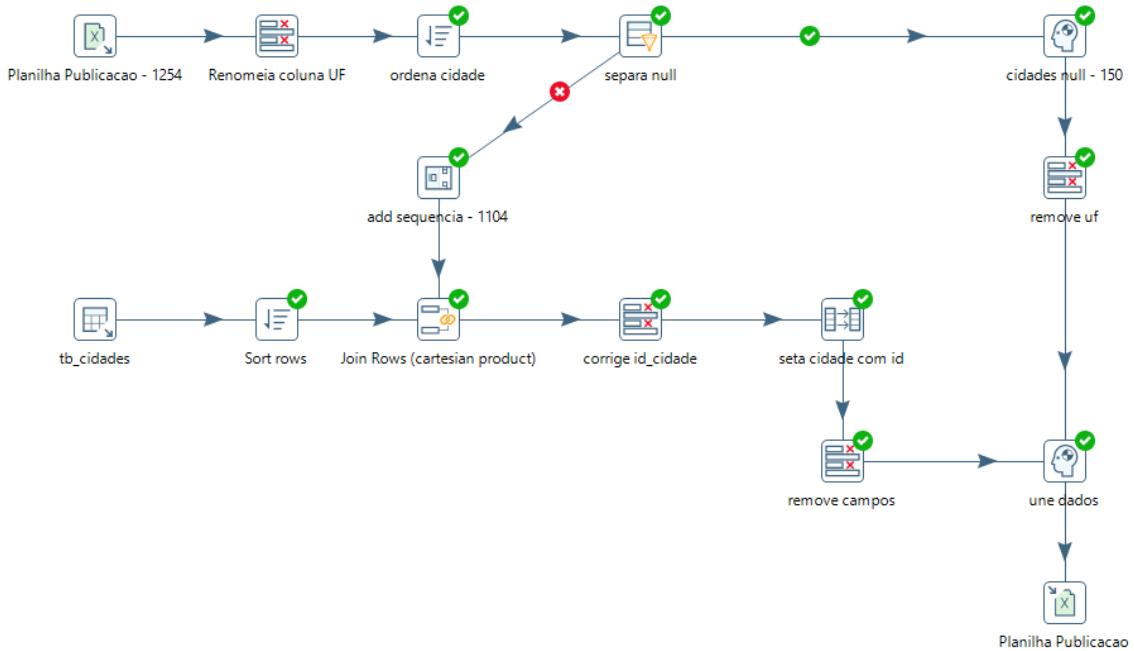


FIGURA 21 - Processo de ETL para substituição do nome pelo número de identificação de cada cidade.

O último ETL realizado antes de persistir os dados da planilha de publicações na *staging area*, é para realizar a alteração do nome de cada pesquisador pelo número de identificação salvo na tabela pesquisador criada na *staging area*. Na planilha publicação, existem registros com até cinco pesquisadores envolvidos, divididos entre as colunas “Autor 1”, “Autor 2”, “Autor 3”, “Autor 4”, e “Autor 5”. Esta situação torna o ETL (FIGURA 22) um pouco mais complexo, exigindo que o processo de substituição se repita cinco vezes antes de criar e persistir os dados na tabela publicação na *staging área*.

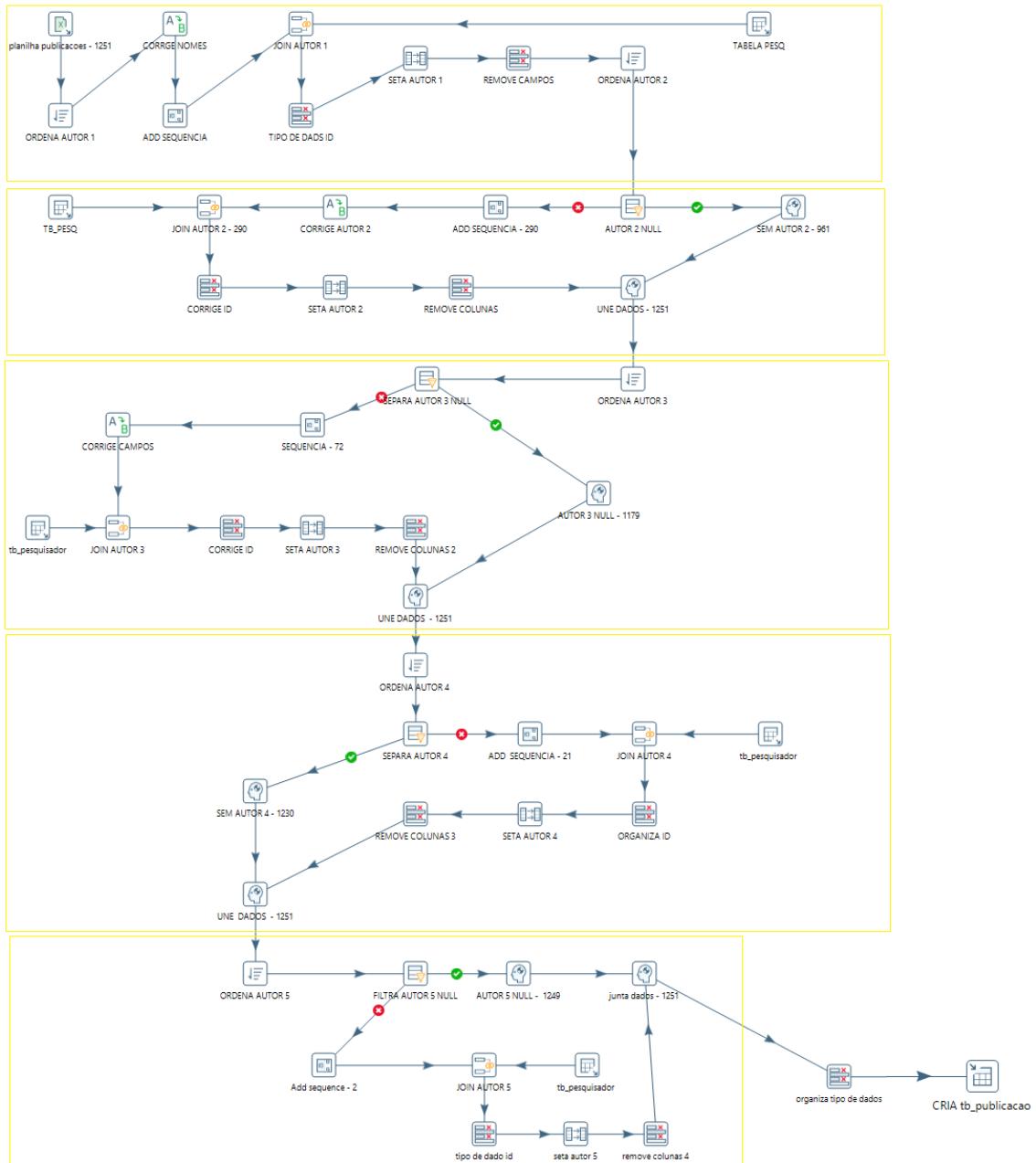


FIGURA 22 - Processo de ETL para substituição do nome pelo número de identificação de cada pesquisador, criação e inserção de dados na tabela publicação da *staging area*.

O processo de ETL da substituição dos nomes dos autores inicia com a leitura dos dados da planilha de publicações criada a partir do ETL de cidades. Após leitura dos dados, os registros são organizados por ordem alfabética de acordo com o nome do “Autor 1”. Como todas as publicações obrigatoriamente possuem pelo menos um autor, não foi preciso buscar por campos vazios nesta coluna. Para verificar a existência de erros gramaticais na descrição de cada campo da coluna “Autores 1”, os registros da planilha foram colocados em ordem alfabética de acordo com o nome de cada autor através da utilização do *step Sort Row*. Neste momento foi utilizado o *step Add Sequence* para criar uma nova coluna com valores

sequenciais cujo objetivo consistem em auxiliar na identificação dos registros caso haja alguma perda. Em seguida, a coluna “Autor 1” foi comparada, através do *step Join Rows*, com os registros da tabela de pesquisadores da *staging área*, extraídas a partir do *step Table Input*. Através do *step Set Field Value* foi feita a substituição do nome pela identificação de cada registro do campo “Autor 1”.

Todo este processo foi repetido outras quatro vezes no mesmo ETL para os campos “Autor 2”, “Autor 3”, “Autor 4” e “Autor 5”. Porém, como o preenchimento desses campos não são obrigatórios, antes de realizar a comparação entre os dados da tabela de pesquisadores da *staging area*, foi utilizado o *step Filter Row* para identificar as publicações com mais de um autor, separa-las das demais, executar a substituição para então integrá-las novamente. Após realizar a substituição nas cinco colunas de autores é que foi possível utilizar o *step Table Output* para persistir a tabela publicação na *staging area* com dados da planilha de publicações (FIGURA 23).

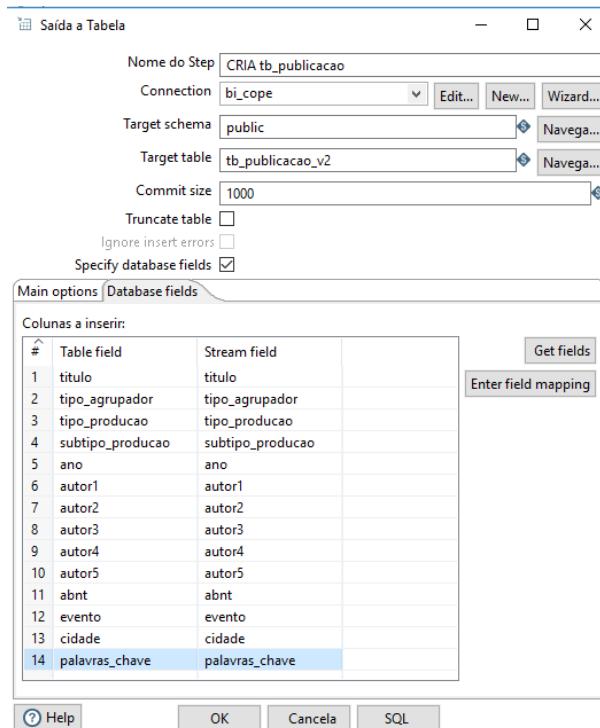


FIGURA 23 - Configurando do *step Table Output* para persistência dados da planilha de publicações para a *staging area*.

A última tabela criada na *staging area* foi a tabela de projetos. Não foi preciso realizar inclusão de dados ou colunas na planilha de projetos mantida pelo COPE. Porém, diversos registros foram inseridos através do uso da quebra de linha. Esta prática atrapalha significativamente o processo de ETL, pois impede que o PDI reconheça de forma correta os

dados registrados. Portanto, antes de iniciar o ETL, foi preciso remover essas quebras de linhas manualmente nos 250 projetos com data de 2014 a 2016.

Para persistir os dados da planilha de projetos na *staging área* apenas um ETL foi criado. Isto porque, não foi preciso realizar substituições dos dados referente a cidade e evento, visto que os projetos são realizados – em sua grande maioria – na instituição e a planilha do COPE não possui informações que permitam identificar locais e publicações atreladas ao projeto. Apesar de extenso, o processo de ETL para substituição dos nomes dos autores em cada projeto (FIGURA 24) é praticamente o mesmo realizado para a planilha de publicação e segue o mesmo roteiro: leitura dos dados da planilha eletrônica, disponibilização dos dados em ordem alfabética de acordo com o nome do autor, união e comparação com os registros extraídos da tabela pesquisador da *staging área*. Assim como no ETL das publicações, o ETL dos projetos também precisou repetir a ordenação e substituição dos nomes dos pesquisadores cinco vezes, uma vez para substituir o nome pela identificação do coordenador do projeto e outras quatro vezes para identificar e substituir as informações dos colaboradores. Só então, os dados foram persistidos na *staging area*.

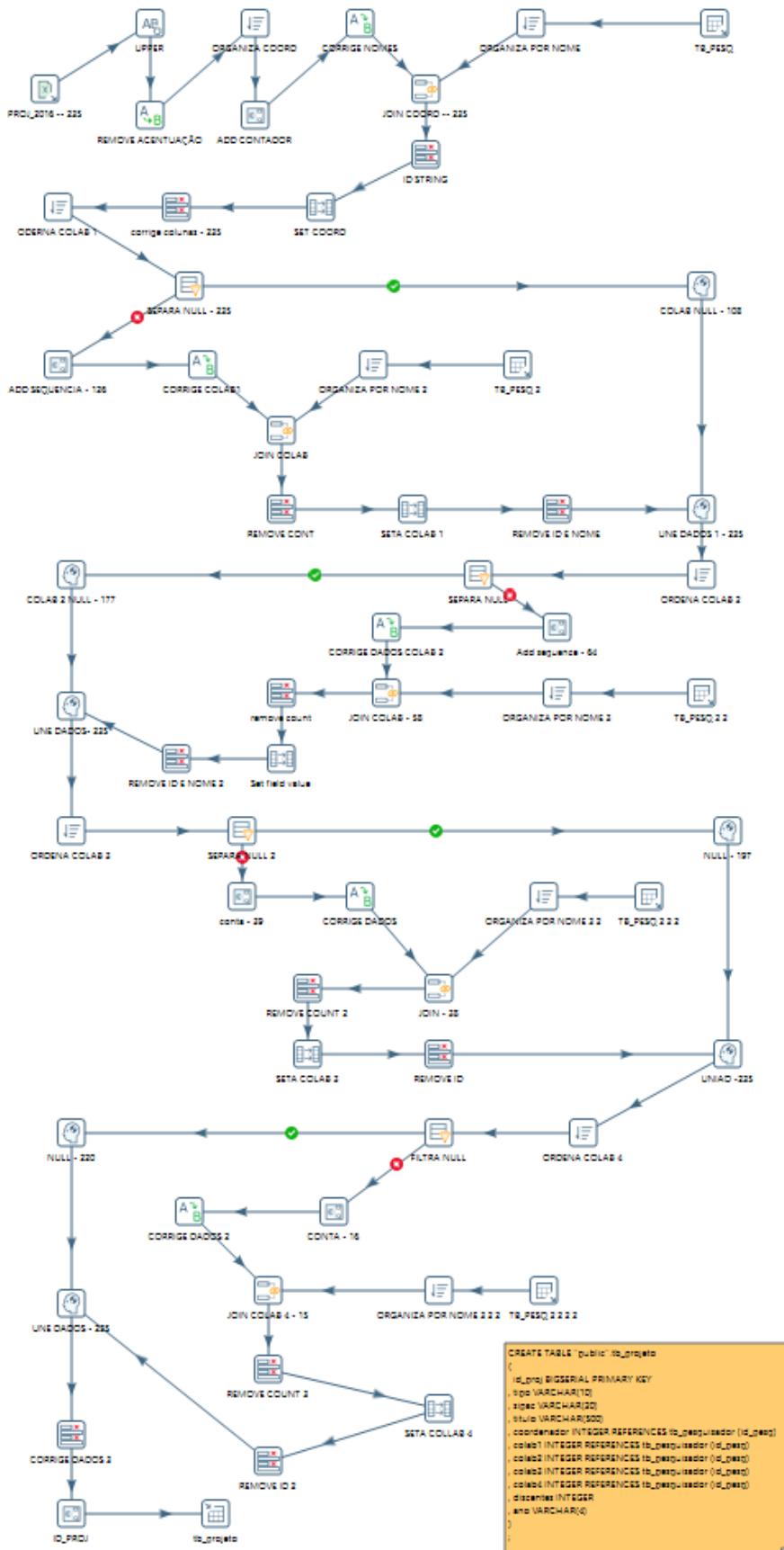


FIGURA 24 - Processo de ETL para substituição do nome pelo número de identificação de cada pesquisador, criação e inserção de dados na tabela projeto da *staging area*.

Como resultado das transformações descritas obteve-se a criação de uma base de dados relacional denominada *staging area*, conforme demonstrado na arquitetura do BI (FIGURA 9) deste trabalho.

4.3 CRIAÇÃO DO DATA WAREHOUSE DE PRODUÇÃO CIENTÍFICA

A *staging area* que compõe este trabalho, é formada por oito tabelas que representam as seguintes informações: área de conhecimento, subárea, evento, cidade, colegiado, pesquisador, projeto e publicação (FIGURA 26).

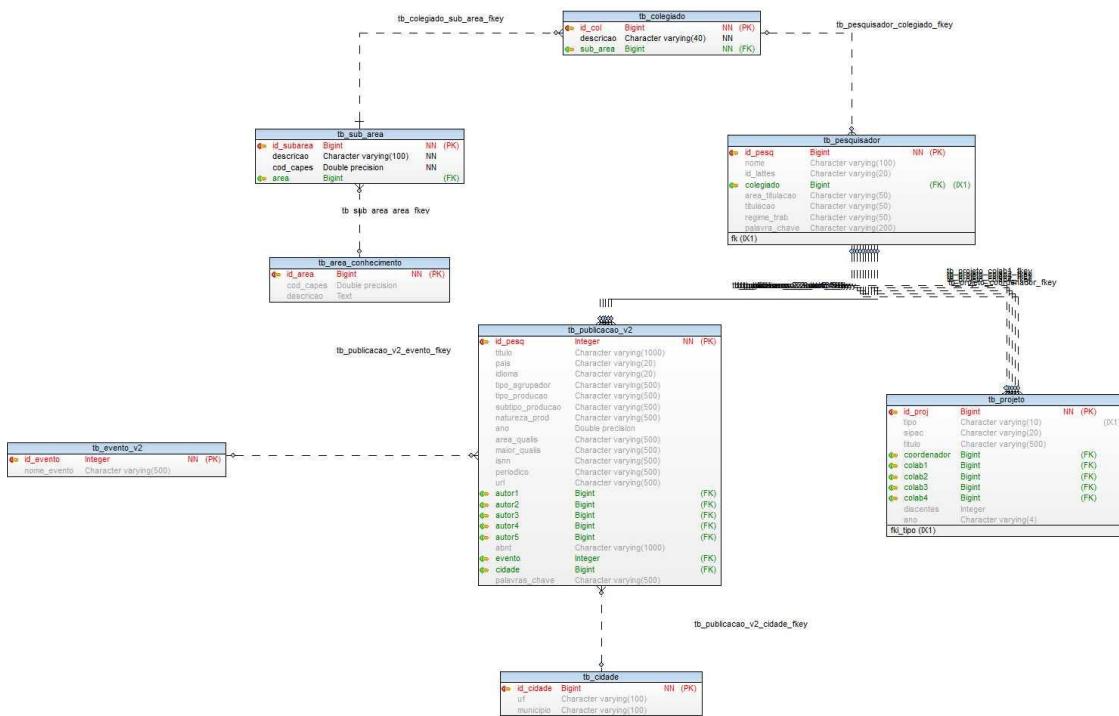


FIGURA 25 - Modelo Entidade-Relacionamento da Staging Area.

De acordo com o modelo ER (FIGURA 25) é possível identificar os relacionamentos que ocorrem entre as tabelas. Os projetos, por exemplo, estão ligados aos pesquisadores, que pertencem a um colegiado que estão relacionados a uma das subáreas de uma das áreas de conhecimento da CAPES. Já as publicações estão relacionadas a um evento, relacionado a uma cidade e aos pesquisadores que pertencem a um colegiado que estão relacionados a uma das subáreas de uma das áreas de conhecimento da CAPES.

O *Data Warehouse* da Produção Científica é criado a partir das oito tabelas que compõem a *staging área*. A criação e inserção de dados nas tabelas dimensões que compõem o DW é feita através do script SQL (FIGURA 27) que acessa informações das tabelas da *staging area* e as persiste no DW.

```

1  create table dim_cidade(
2    id_cidade INTEGER PRIMARY KEY,
3    nome_cidade VARCHAR(100),
4    uf VARCHAR(2)
5  )
6
7  INSERT INTO dim_cidade(id_cidade, nome_cidade, uf)
8    SELECT id_cidade, municipio, uf FROM
9      dblink ('user=postgres password=root dbname=bi_cope' :: text,
10      'SELECT id_cidade, municipio, uf FROM tb_cidade':: text, false)
11      tb_cidade(id_cidade integer, municipio character varying, uf character varying)
12
13 -----
14
15  create table dim_evento(
16    id_evento INTEGER PRIMARY KEY,
17    nome_evento VARCHAR(200),
18  )
19
20 INSERT INTO dim_evento(id_evento, nome_evento)
21   SELECT id_evento, nome FROM
22     dblink ('user=postgres password=root dbname=bi_cope' :: text,
23      'SELECT id_evento, nome FROM tb_evento_v2':: text, false)
24      tb_evento_v2(id_evento integer, nome character varying)
25
26 -----
27
28  create table dim_pesquisador(
29    id_pesquisador INTEGER PRIMARY KEY,
30    nome_pesquisador VARCHAR(200),
31    formacao_pesquisador VARCHAR(50)
32  )
33
34
35 INSERT INTO dim_pesquisador(id_pesquisador, nome_pesquisador, formacao_pesquisador)
36   SELECT id_pesq, nome, titulacao FROM
37     dblink ('user=postgres password=root dbname=bi_cope' :: text,
38      'SELECT id_pesq, nome, titulacao FROM tb_pesquisador':: text, false)
39      tb_pesquisador(id_pesq integer, nome character varying, titulacao character varying)
40
41 -----
42
43  create table dim_colegiado(
44    id_colegiado INTEGER PRIMARY KEY,
45    nome_colegiado VARCHAR(25)
46  )
47
48 INSERT INTO dim_colegiado(id_colegiado, nome_colegiado)
49   SELECT id_col, descricao FROM
50     dblink ('user=postgres password=root dbname=bi_cope' :: text,
51      'SELECT id_col, descricao FROM tb_colegiado':: text, false)
52      tb_colegiado(id_col integer, descricao character varying)
53
54 -----
55
56  create table dim_area_de_conhecimento(
57    id_area INTEGER PRIMARY KEY,
58    nome_area VARCHAR(50)
59  )
60
61 INSERT INTO dim_area_de_conhecimento(id_area, nome_area)
62   SELECT id_area, descricao FROM
63     dblink ('user=postgres password=root dbname=bi_cope' :: text,
64      'SELECT id_area, descricao FROM tb_area_conhecimento':: text, false)
65      tb_area_conhecimento(id_area integer, descricao character varying)
66
67 -----
68
69  create table dim_subarea(
70    id_subarea INTEGER PRIMARY KEY,
71    nome_subarea VARCHAR(50)
72  )
73
74 INSERT INTO dim_subarea(id_subarea, nome_subarea)
75   SELECT id_subarea, descricao FROM
76     dblink ('user=postgres password=root dbname=bi_cope' :: text,
77      'SELECT id_subarea, descricao FROM tb_sub_area':: text, false)
78      tb_subarea(id_subarea integer, descricao character varying)
79
80 -----
81
82  create table dim_tempo(
83    id_tempo bigserial primary key,
84    ano integer
85  )
86
87 insert into dim_tempo(ano) values (2014), (2015), (2016)
88
89 -----
90
91  create table dim_publicacao(
92    id_publicacao INTEGER PRIMARY KEY,
93    titulo VARCHAR(1000),
94    tipo_agrupador VARCHAR(50),
95    tipo_producao VARCHAR(100),
96    subtipo_producao VARCHAR(100)
97  )
98
99 INSERT INTO dim_publicacao(id_publicacao, titulo, tipo_agrupador, tipo_producao, subtipo_producao)
100   SELECT id_pesq, titulo, tipo_agrupador, tipo_producao, subtipo_producao FROM
101     dblink ('user=postgres password=root dbname=bi_cope' :: text,
102      'SELECT id_pesq, titulo, tipo_agrupador, tipo_producao, subtipo_producao FROM tb_publicacao_v2':: text, false)
103      tb_publicacao(id_pesq integer, titulo character varying, tipo_agrupador character varying,
104      tipo_producao character varying, subtipo_producao character varying)
105
106 -----
107
108  create table dim_projeto(
109    id_projeto BIGSERIAL PRIMARY KEY,
110    tipo varchar(15),
111    titulo varchar(700)
112  )
113
114 INSERT INTO dim_projeto(id_projeto, tipo, titulo)
115   SELECT id_proj, tipo, titulo FROM
116     dblink ('user=postgres password=root dbname=bi_cope' :: text,
117      'SELECT id_proj, tipo, titulo FROM tb_projeto':: text, false)
118      tb_proj(id_proj integer, tipo character varying, titulo character varying)

```

FIGURA 26 - Criação das tabelas de dimensão e inserção de dados através de SQL.

Além das dimensões criadas a partir do SQL (FIGURA 26), o DW conta ainda com duas tabelas fato: produção e projeto. A fato produção contém os registros da produção científica extraída do currículo lattes de cada pesquisador no período de 2014 a 2016. Estes dados foram persistidos na tabela fato publicação da *staging area* através de uma nova transformação utilizando o PDI (FIGURA 27).

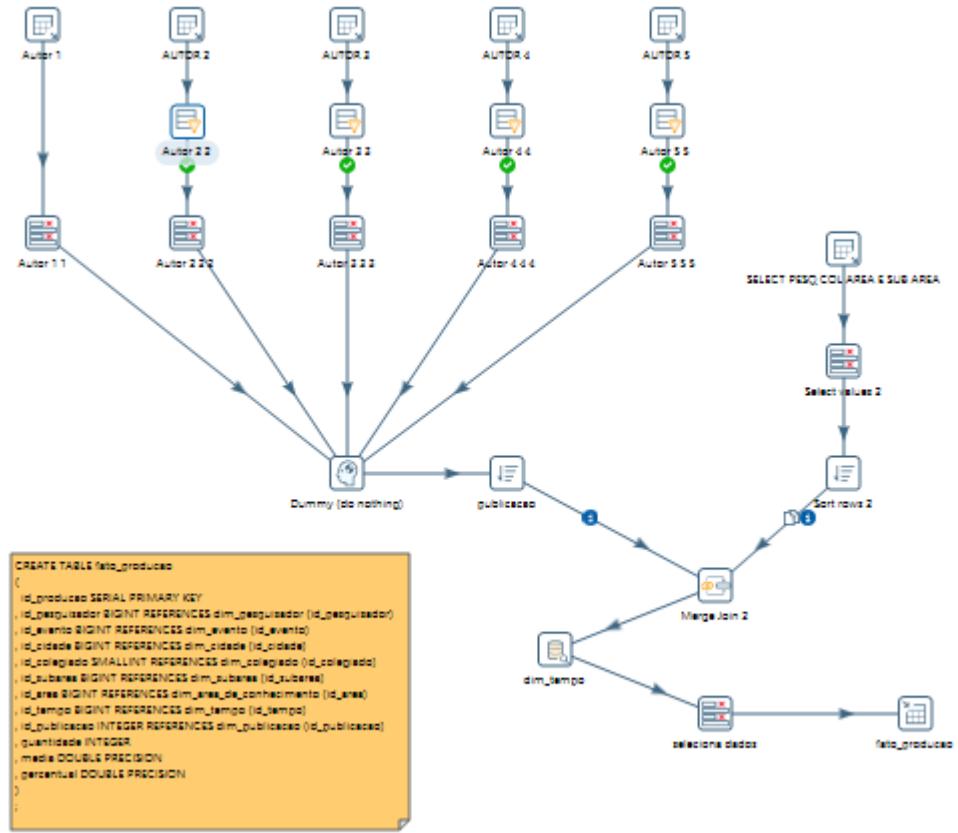


FIGURA 27 - Transformação que realiza a carga de dados da planilha de publicação para a tabela publicação da *staging area*.

A principal característica de uma tabela fato é armazenar dados históricos. Quando ocorre alguma alteração ou inclusão de dados de algum registro, tratando-se de uma tabela fato ao invés de realizar a alteração um novo registro é adicionado mantendo assim um histórico dos dados. No caso das publicações, cada registro será lançado de acordo com o pesquisador. Ou seja, as publicações com mais de um autor, deverão ser escritas na tabela fato uma vez para cada um dos autores, portanto, publicações com três autores, por exemplo, aparecerão na tabela fato três vezes.

Desta forma, os primeiros *steps* do processo de ETL da tabela fato consistem em duplicar os registros de cada publicação de acordo com o número de autores. Em seguida, os

dados foram disponibilizados em ordem crescente de acordo com o número de identificação de cada pesquisador, através do *step Sort Rows*, e integrados com o resultado da consulta SQL realizada pelo *step Table Input* (FIGURA 28). O *step Lockup Dimension* acrescentou a informação dos anos de cada publicação através da dimensão tempo, para só então criar a tabela fato produção no DW.

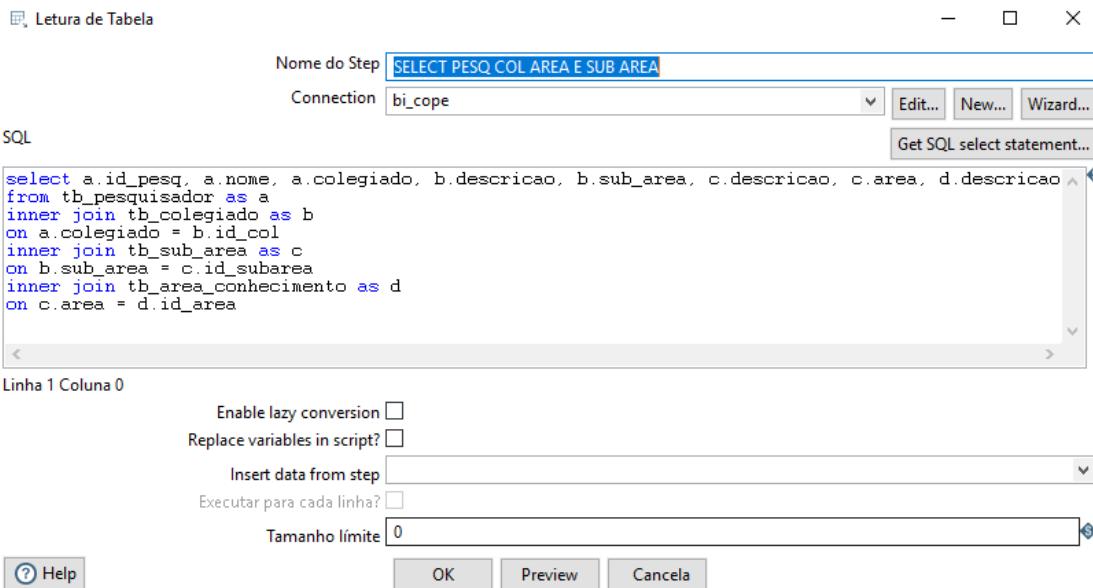


FIGURA 28 - Configuração do *step Table Input* com o *select* utilizado para listar informações das tabelas pesquisador, colegiado e subáreas.

O processo de ETL para criar a tabela fato de projetos (FIGURA 29) segue a mesma lógica apresentada no ETL de criação da tabela fato publicação (FIGURA 27), a principal diferença entre uma tabela e outra está na fonte de dados, uma vez que os registros que compõem a fato de projetos são oriundos da planilha de projetos do COPE, e na ausência de informações sobre eventos e cidades.

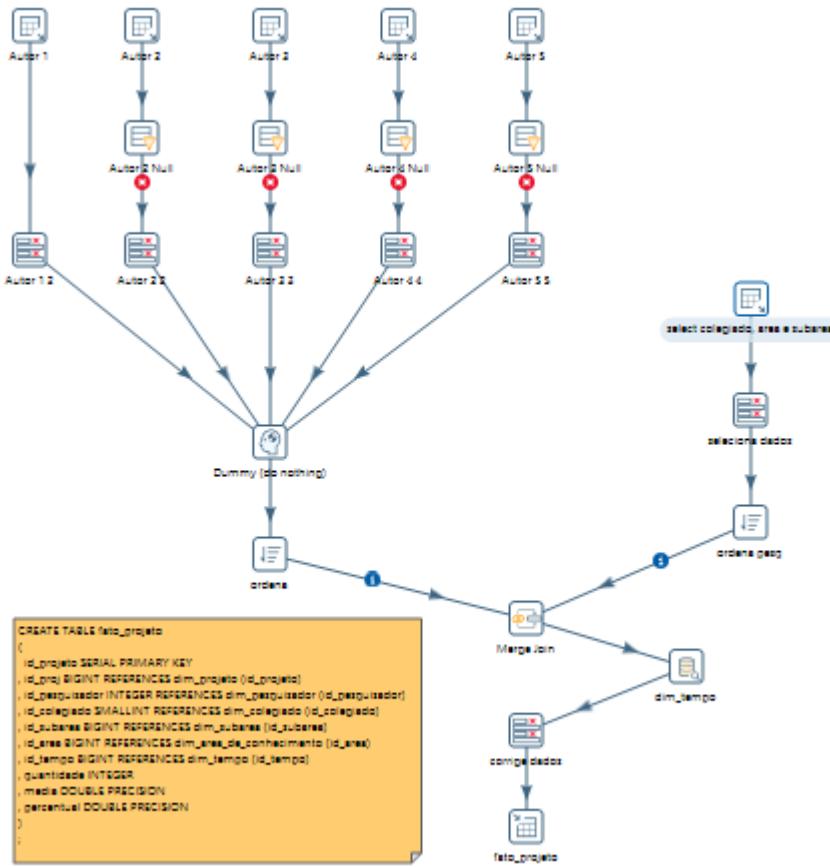


FIGURA 29 - Processo de ETL para criação da tabela fato de projetos.

O resultado das transformações realizadas pelos ETL de projetos e produção é a criação de duas tabelas fato no DW, a fato produção com 1636 registros e a fato projetos com 472 registros. A composição de ambas as tabelas pode ser visualizada diretamente no PostgreSQL (FIGURA 30) através de um SQL de consulta do DW.

	id_projeto [PK] integer	id_proj bigint	id_pesquis integer	id_colegiado smallint	id_subarea bigint	id_area bigint	id_tempo bigint	quantidade integer	media double pre... cision	percentual double pre... cision
1	1	3	1	13	6	1	3	[null]	[null]	[null]
2	2	186	2	10	36	4	3	[null]	[null]	[null]
3	3	13	2	10	36	4	1	[null]	[null]	[null]
4	4	11	2	10	36	4	2	[null]	[null]	[null]
5	5	12	2	10	36	4	3	[null]	[null]	[null]
6	6	7	2	10	36	4	3	[null]	[null]	[null]
7	7	2	2	10	36	4	2	[null]	[null]	[null]
8	8	14	3	11	73	8	3	[null]	[null]	[null]
9	9	74	3	11	73	8	2	[null]	[null]	[null]
10	10	171	4	14	3	1	3	[null]	[null]	[null]
11	11	172	4	14	3	1	1	[null]	[null]	[null]
12	12	173	4	14	3	1	1	[null]	[null]	[null]

FIGURA 30 - Resultado do SQL de consultas aplicado na tabela fato projetos do DW.

Como resultado dos passos apresentados nesta sessão, obteve-se a criação do DW da produção científica do IFPR campus Palmas, composto por nove tabelas de dimensão e duas tabelas fato.

4.4 CRIAÇÃO E PUBLICAÇÃO DOS CUBOS OLAP

Terminada a criação do DW foi possível utilizar a ferramenta *Schema Workbench* para criar os cubos OLAP. Uma das principais vantagens do uso desta ferramenta é que ela não exige nenhum processo de instalação, além de ser distribuída livremente sem custo financeiro pelo site do projeto. Para poder utilizá-la o usuário precisa apenas realizar o *download* da pasta *Schema Workbench* no site do projeto (SOURCEFORCE, 2017) e descompactar a pasta em um diretório de sua preferência. Com a pasta descompactada, basta executar o arquivo .bat denominado *workbench* (FIGURA 31).

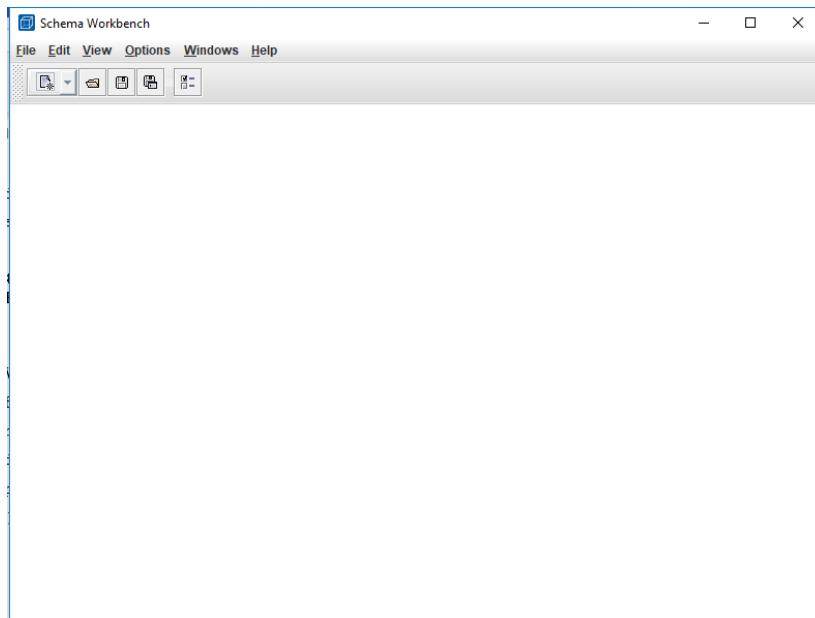


FIGURA 31 - Tela inicial da ferramenta *Schema Workbench*.

O primeiro passo para mapear um cubo OLAP é configurar a conexão do *Schema Workbench* com a base de dados *PostgreSQL* onde o DW é armazenado. Para isso, basta acessar o campo “*Conection...*” na aba “*Option*” no menu superior da ferramenta *Schema Workbench* e preencher os campos com os dados de conexão do DW de acordo com as informações da base de dados (FIGURA 32).

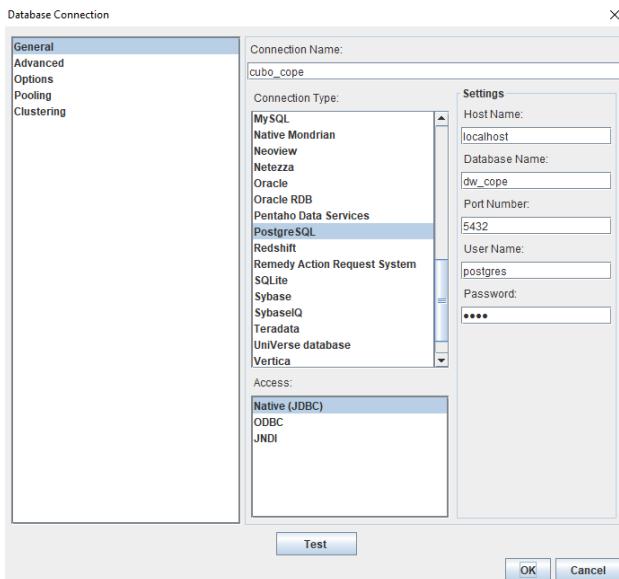


FIGURA 32 - Configurando a conexão do *Schema Workbench* com o DW.

As primeiras informações a serem preenchidas no momento de configurar a conexão são definir um nome para a conexão e informar o tipo do banco de dados onde o DW está armazenado, através dos campos “*Connection Name*” e “*Connection Type*.“ Neste trabalho, o DW e os cubos OLAP estão no mesmo equipamento, por isso o campo “*Host Name*” foi configurado como “*localhost*”. No campo “*Database Name*” é preciso informar o nome utilizado na criação do DW. O campo “*Port Name*” e o campo “*Acess*” são preenchidos de forma automática de acordo com o tipo da base de dados selecionado. Já os campos “*User Name*” e “*Password*” referem-se ao nome de usuário e senha cadastrados por cada usuário no momento de instalação do banco de dados e variam de usuário para usuário.

Com a conexão realizada é preciso criar um novo esquema dentro do *Schema Workbench*, para isso basta selecionar o campo “*Scheme*” dentro do campo “*New*” na opção “*File*” do menu superior. É no esquema que os cubos serão mapeados através de suas dimensões, fatos e medidas já pré-definidos no decorrer do projeto.

O esquema criado para realizar as consultas da produção científica dos pesquisadores do IFPR campus Palmas é formado por dois cubos OLAP, o cubo Produção Científica criado a partir da tabela fato produção e o cubo Projeto criado a partir da tabela fato de projetos. As particularidades de cada cubo estão nas tabelas fato e nas dimensões que compõem cada cubo. Para evitar erros de leitura dos dados e na criação do XML que compõem os cubos, não foi utilizado nenhum tipo de acentuação ortográfica em nenhum componente mapeado.

Para criar um novo cubo no esquema, basta clicar com o botão direito sobre o esquema e selecionar a opção “*Add Cube*” (FIGURA 33).

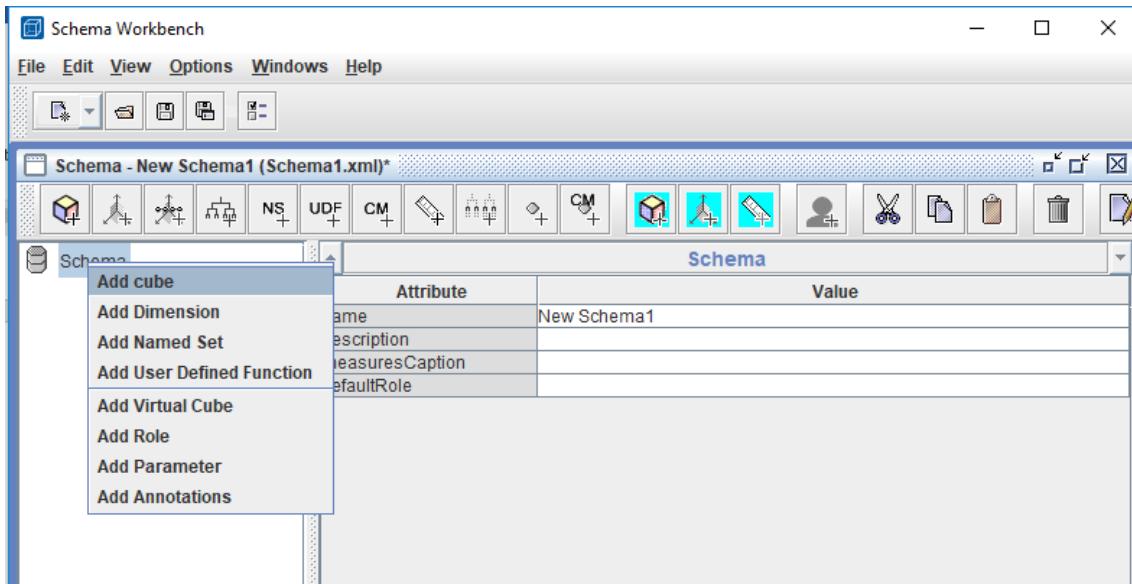


FIGURA 33 - Adicionando um novo cubo no esquema.

Com o esquema e o cubo criados, foi iniciado o processo de mapeamento da tabela fato, das dimensões e medidas que vão compor o cubo. Foi mapeado primeiro a tabela fato, depois as medidas e por último as dimensões de cada cubo. Para mapear a tabela fato no cubo, bastou clicar com o botão direito sobre ele, escolher a opção “Add Table” e definir as configurações apresentadas no lado direito do ambiente de trabalho (FIGURA 34). Os campos “schema”, “name” e “alias” devem ser preenchidos com as informações referentes ao nome da tabela fato criado no esquema e no nome que a tabela possui no DW, já o campo “alias” é opcional. Abaixo é apresentado como foi mapeada a tabela fato do cubo Produção Científica (FIGURA 34).

Table for 'Producao Cientifica' Cube		
	Attribute	Value
schema	public	
name	fato_producao	
alias		

FIGURA 34 - Definindo os parâmetros de configuração da tabela fato no cubo de Produção Científica

Com a tabela fato mapeada, iniciou-se o processo para mapear as medidas do cubo. Foram mapeadas duas medidas, quantidade e percentual de publicação, e serão essas medidas que permitirão mensurar a quantidade, média e percentual referente às publicações por área de conhecimento, colegiado, cidade, entre outros. Para mapear uma medida no cubo basta clicar com o botão direito do mouse sobre o cubo, selecionar o campo “Add Measure” e preencher os parâmetros de configuração da tabela que será apresentada do lado direito do *Schema Workbench* (FIGURA 35).

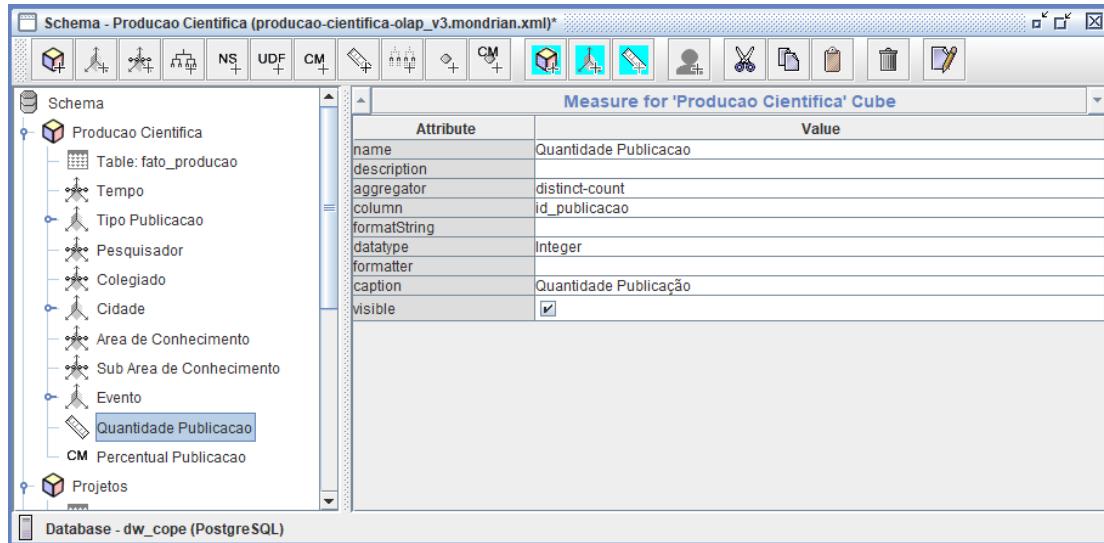


FIGURA 35 - Criação da medida quantidade publicação no cubo Produção Científica.

A medida quantidade publicação soma o número de publicações através do campo “id_publicacao” da tabela fato, desta forma, ela faz a distinção das publicações com o mesmo número de identificação. Além da medida quantidade publicação, o cubo Produção Científica conta também com a medida calculada denominada de Percentual Publicação que verifica a quantidade percentual de cada consulta, como por exemplo o percentual das publicações por colegiado ou tipo de publicação.

Para criar uma medida calculada, basta clicar com o botão direito sobre o cubo e escolher a opção “Add Calculated Member” e preencher os parâmetros de configuração (FIGURA 36). A diferença de uma medida calculada para uma medida normal é a possibilidade da definição de fórmulas personalizadas, escritas em MDX, para realizar a consulta desejada.

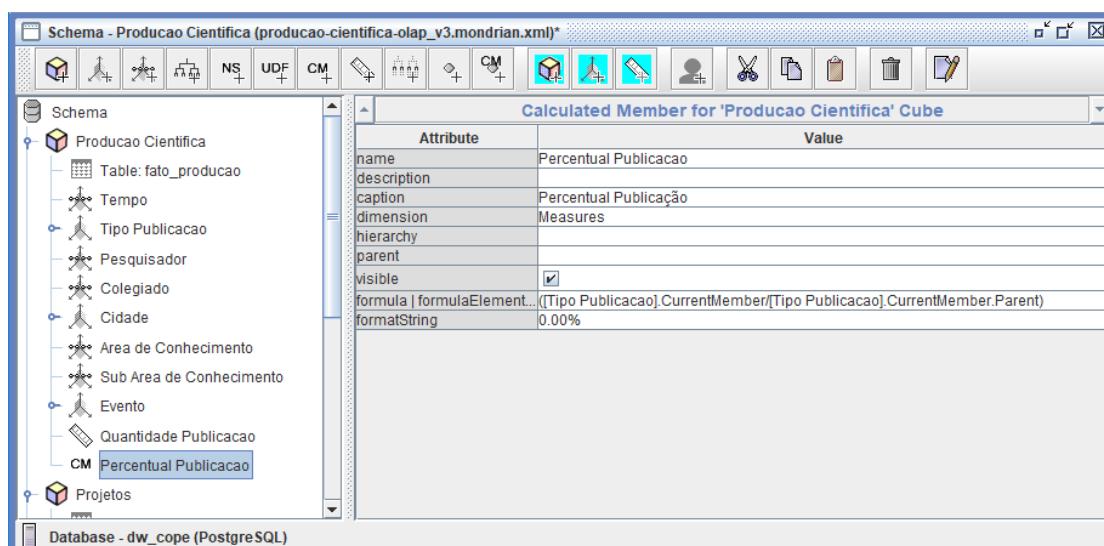


FIGURA 36 - Criando uma medida calculada para consultar o percentual da Produção Científica.

Com a tabela fato e as medidas mapeadas, pôde-se começar a mapear as dimensões do cubo.

O *Schema Workbench* permite a criação de dois tipos de dimensões: a *Dimension* a *Dimension Usage*. A diferença entre elas está no nível onde se encontram e a utilização nos cubos. Quando uma dimensão possuí informações genéricas que podem ser utilizadas por mais de um cubo, como por exemplo a dimensão tempo, elas podem ser chamadas de dimensão em conformidade e são criadas a partir da opção *Dimension*, assim a dimensão fica dentro do esquema no mesmo nível de criação dos cubos e pode ser referenciada em um ou mais cubos. Mas, se a dimensão trata de informações específicas a um cubo, então utiliza-se a *Dimension Usage* e a dimensão é criada dentro do cubo em questão, no mesmo nível da tabela fato e das medidas.

Neste trabalho, as dimensões Tempo, Pesquisador, Colegiado, Área de Conhecimento e Subárea serão utilizadas nos dois cubos do esquema, por isso serão criadas em conformidade no esquema. Para criar uma dimensão em conformidade, basta clicar com o botão direito sobre o esquema do projeto (*Schema*), selecionar o campo “Add Dimension” e preencher os parâmetros da tabela de configuração. Para configurar uma dimensão em conformidade, o primeiro passo é definir o nome e tipo da dimensão através dos campos “name” e “type” (FIGURA 38).

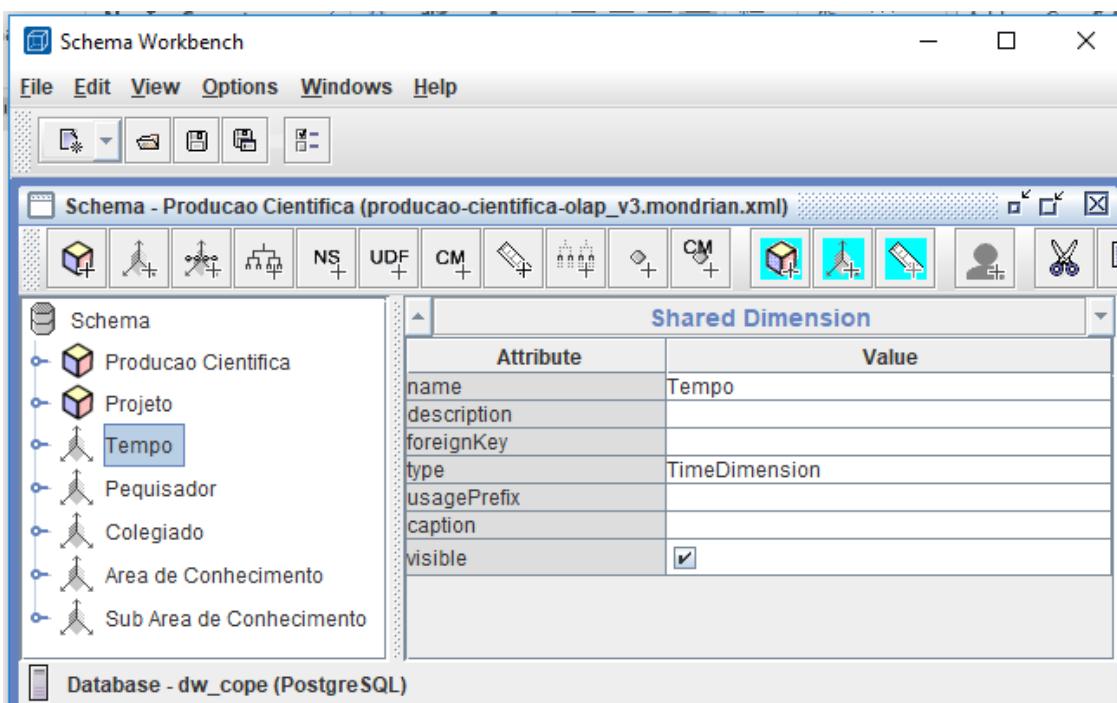


FIGURA 37 - Configurando a dimensão em conformidade Tempo.

Para mapear uma dimensão é preciso definir sua hierarquia. Toda dimensão possui pelo menos uma hierarquia, que nada mais é que um conjunto de membros organizados em uma estrutura para análise. Basicamente é nela onde os atributos de consulta de cada dimensão são mapeados. Para criar uma hierarquia, basta clicar com o botão direito sobre a dimensão, selecionar o campo “*Add Hierarchy*” e definir os parâmetros da tabela de configuração (FIGURA 38).

Hierarchy for 'Tempo' Dimension	
Attribute	Value
name	Ano
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	Todos os anos
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	id_tempo
caption	
visible	<input checked="" type="checkbox"/>

FIGURA 38 - Criando e configurando uma hierarquia na dimensão tempo.

A dimensão tempo possui apenas uma hierarquia que corresponde ao Ano das publicações e dos projetos. Com a hierarquia criada, é preciso informar qual a dimensão do DW que irá fornecer os dados para a dimensão do cubo. Para isso, clicar com o botão direito do mouse sobre a hierarquia, selecionar o campo “*Add Table*” e informar o esquema e o nome da dimensão através dos campos “*schema*” e “*name*” (FIGURA 39).

Table for 'Ano' Hierarchy	
Attribute	Value
schema	public
name	dim_tempo
alias	

FIGURA 39 - Mapeando a tabela do DW que compõe a dimensão tempo do cubo.

Cada atributo da tabela pode ser um nível de consulta na hierarquia da dimensão. A tabela de dimensão Tempo do DW é composta por dois atributos: *id_tempo* e descrição. O “*id*” sempre será mapeado na hierarquia da dimensão (FIGURA 38) e os demais atributos serão mapeados nos níveis da hierarquia. No caso da dimensão tempo, haverá apenas um nível composto pela descrição de cada ano. Para mapear um nível, basta clicar sobre a hierarquia com o botão direito do mouse, selecionar o campo “*Add Level*” e preencher os parâmetros da tabela de configuração (FIGURA 40).

Level for 'Ano' Hierarchy	
Attribute	Value
name	Ano
description	
table	
column	ano
nameColumn	
parentColumn	
nullParentValue	
ordinalColumn	
type	String
internalType	
uniqueMembers	<input type="checkbox"/>
levelType	TimeYears
hideMemberIf	Never
approxRowCount	
caption	
captionColumn	
formatter	
visible	<input checked="" type="checkbox"/>

FIGURA 40 - Configuração dos parâmetros para criar o nível ano na dimensão tempo.

Após definir o nome do nível, a coluna que referencia o atributo no DW, o tipo do dado e o tipo do nível, através dos campos “name”, “column”, “type” e “levelType”, a dimensão em conformidade Tempo ficou composta por uma hierarquia, um nível e uma tabela (FIGURA 41), pronta para ser utilizada nos cubos do esquema.



FIGURA 41 - Composição da dimensão em conformidade Tempo

O mesmo procedimento foi realizado para mapear as hierarquias, níveis e tabelas das dimensões Pesquisador, Colegiado, Área de Conhecimento e Subáreas (FIGURA 42).

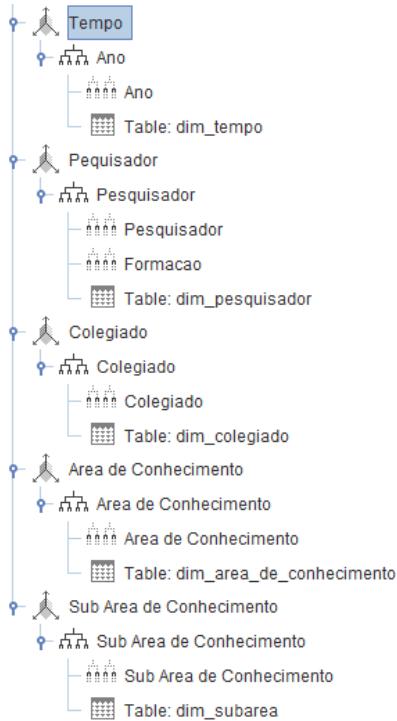


FIGURA 42 - Composição das dimensões Tempo, Pesquisador, Colegiado, Área de Conhecimento e Subáreas de Conhecimento criadas em conformidade com o esquema.

Após criadas, as dimensões em conformidade precisam ser mapeadas nos cubos OLAP. Para isso, clicar com o botão direito do *mouse* sobre o cubo, selecionar a opção “*Add Dimension Usage*” e preencher os parâmetros da tabela de configuração (FIGURA 43), informando o nome da dimensão, a chave estrangeira da consulta e a dimensão em conformidade a qual ela se refere.

Attribute	Value
name	Tempo
foreignKey	id_tempo
source	Tempo
level	
usagePrefix	
caption	
Visible	<input checked="" type="checkbox"/>

FIGURA 43 - Referenciando a dimensão em conformidade Tempo dentro do cubo da Produção Científica.

As dimensões Pesquisador, Colegiado, Área de Conhecimento e Subárea foram mapeadas através do mesmo procedimento realizado para a dimensão Tempo.

Além das dimensões em conformidade, o cubo da Produção Científica possui outras três dimensões específicas, a dimensão Cidade, dimensão Evento e dimensão Tipo de

Publicação. As três dimensões foram mapeadas através do seguinte processo: após clicar com o botão direito sobre o cubo da Produção Científica, foi escolhida a opção “Add Dimension” e os parâmetros da tabela de configuração foram definidos. Após nomear a dimensão, foi preciso adicionar sua hierarquia, seus níveis e a tabela do DW a qual faz referência – mesmo processo realizado para a criação das dimensões em conformidade. Com a criação das dimensões Tipo de Publicação, Evento e Cidade dentro do cubo, no mesmo nível das medidas e da tabela fato (FIGURA 44), o cubo Produção Científica foi mapeado com uma tabela fato, duas medidas e nove dimensões.

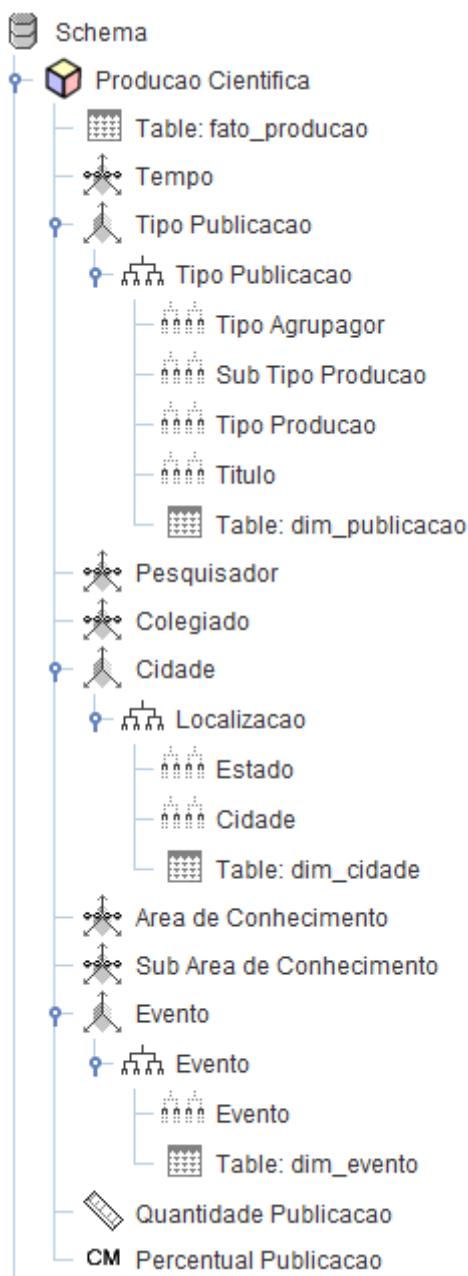


FIGURA 44 - Cubo Produção Científica com todos os componentes mapeados (fato, medidas e dimensões).

Apesar de não possuir dimensões específicas, a criação do cubo Projetos seguiu o mesmo processo de criação do cubo de Produção Científica. Primeiro a tabela fato foi mapeado referenciando a tabela fato Projetos do DW. O cubo de projetos também é formado por duas medidas, medida simples outra calculada, que foram criadas de forma idêntica ao processo realizado no cubo da Produção Científica. O cubo Projetos não possui dimensões específicas, desta forma sua composição conta apenas com as cinco dimensões criadas em conformidade com o esquema. Após mapear as dimensões em conformidade com o cubo de Projetos, repetindo o mesmo processo realizado no cubo de Produção Científica, sua composição final consiste em uma tabela fato, duas medidas e cinco dimensões (FIGURA 45).

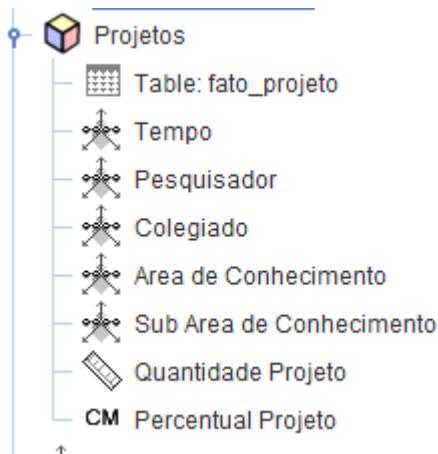


FIGURA 45 - Cubo Projetos com todos os componentes mapeados (fato, medidas e dimensões).

Com os cubos mapeados foi realizado uma verificação da conexão e testado se as dimensões, tabelas e medidas estão mapeadas de forma correta trazendo os mesmos resultados armazenados no DW. Para isso, é realizado uma consulta MDX sobre cada cubo a fim de verificar a quantidade de dados em cada tabela fato mapeada. A consulta MDX é criada a partir da opção “*MDX Query*”, localizada na opção “*New*” do menu “*File*” localizado na parte superior da interface do *Schema Workbench*. Uma nova janela irá aparecer na tela, já com o esquema selecionado no campo “*Schema*”. Ao clicar em “*Connect*” o *Schema Workbench* irá confirmar se a conexão entre o esquema e o DW está acontecendo corretamente através de uma janela informativa (FIGURA 46).

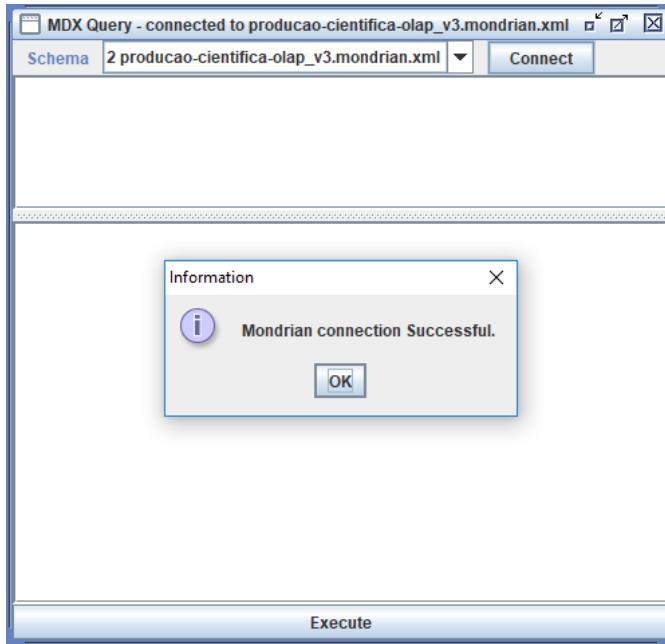


FIGURA 46 - Confirmação da conexão entre o esquema e o DW.

Após confirmada a conexão, pode-se utilizar um *select* para consulta MDX em cada cubo. O *select* utilizado apresenta a quantidade de linhas de cada cubo, através da medida quantidade de cada cubo (FIGURA 47). Após executar a consulta MDX o resultado será o número de linhas identificados na tabela fato mapeada no cubo, este número deve respeitar a mesma quantidade de registros distintos que estão armazenados no DW. No caso da Produção Científica, 1251 registros e no caso dos Projetos, 235 registros.

The image contains two side-by-side screenshots of the MDX Query tool. Both screenshots show the same interface with a title bar 'MDX Query - connected to producao-cientifica-olap_v3.mondrian.xml', a toolbar with 'Schema' selected, and an 'Execute' button at the bottom. The left screenshot shows the following MDX query and its result:

```
select {[Measures].[Quantidade Projeto]} on columns
from [Projetos]
```

Axis #0:
{}
Axis #1:
{[Measures].[Quantidade Projeto]}
Row #0: 235

The right screenshot shows a similar MDX query and its result:

```
select {[Measures].[Quantidade Publicacao]} on columns
from [Producao Cientifica]
```

Axis #0:
{}
Axis #1:
{[Measures].[Quantidade Publicacao]}
Row #0: 1.251

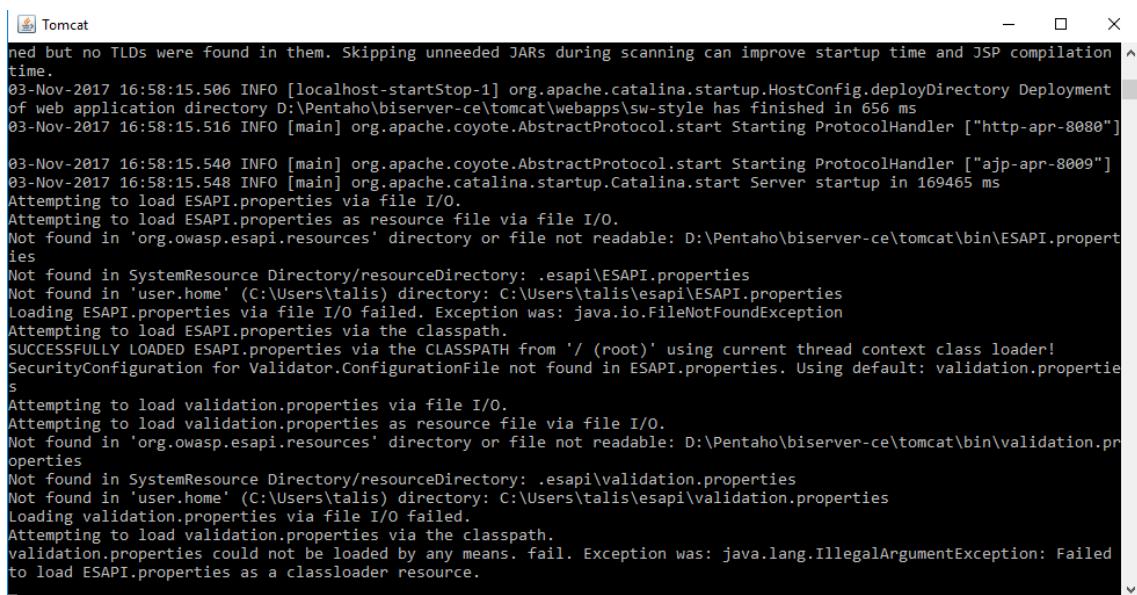
FIGURA 47 - Consultas MDX realizadas nos cubos Produção Científica e Projetos.

Com o esquema conectado e os cubos respondendo as consultas corretamente, foi possível seguir com a publicação do arquivo XML (APÊNDICE 1) na plataforma *Pentaho*. Para isso, acessar o menu “File”, escolher a opção “Publish” e preencher os campos na janela de Publicação do Esquema (FIGURA 48). As três primeiras configurações são referentes ao acesso a Plataforma Analítica da Pentaho que por padrão correspondem a: Server URL: <http://localhost:8080/pentaho>, User: Admin, Password: password. Já no campo “Publish

“Settings” bastando informar o nome da conexão criada entre o esquema e o DW anteriormente. Para que os cubos sejam publicados com sucesso, o *Pentaho BI Server* deve estar inicializado no momento da publicação.

4.5 VISUALIZAÇÃO E CONSULTA DOS DADOS A PARTIR DA PUBLICAÇÃO DOS CUBOS OLAP

A publicação de cubos OLAP, para consultas através do *Pentaho*, depende da inicialização do servidor *Pentaho*. Com o *Pentaho BI Server* instalado e configurado, para iniciar o servidor basta executar o arquivo com a extensão .bat chamado “star-pentaho” que se encontra na pasta biserver-ce. O servidor será iniciado por meio de uma tela do terminal de comando (FIGURA 48) que deverá permanecer aberta durante todo o processo de consulta.



```

Tomcat
ned but no TLDs were found in them. Skipping unneeded JARs during scanning can improve startup time and JSP compilation time.
03-Nov-2017 16:58:15.506 INFO [localhost-startStop-1] org.apache.catalina.startup.HostConfig.deployDirectory Deployment of web application directory D:\Pentaho\biserver-ce\tomcat\webapps\sw-style has finished in 656 ms
03-Nov-2017 16:58:15.516 INFO [main] org.apache.coyote.AbstractProtocol.start Starting ProtocolHandler ["http-apr-8080"]
03-Nov-2017 16:58:15.540 INFO [main] org.apache.coyote.AbstractProtocol.start Starting ProtocolHandler ["ajp-apr-8009"]
03-Nov-2017 16:58:15.548 INFO [main] org.apache.catalina.startup.Catalina.start Server startup in 169465 ms
Attempting to load ESAPI.properties via file I/O.
Attempting to load ESAPI.properties as resource file via file I/O.
Not found in 'org.owasp.esapi.resources' directory or file not readable: D:\Pentaho\biserver-ce\tomcat\bin\ESAPI.properties
Not found in SystemResource Directory/resourceDirectory: .esapi\ESAPI.properties
Not found in 'user.home' (C:\Users\talis) directory: C:\Users\talis\esapi\ESAPI.properties
Loading ESAPI.properties via file I/O failed. Exception was: java.io.FileNotFoundException
Attempting to load ESAPI.properties via the classpath.
SUCCESSFULLY LOADED ESAPI.properties via the CLASSPATH from '/' (root) using current thread context class loader!
SecurityConfiguration for Validator.ConfigurationFile not found in ESAPI.properties. Using default: validation.properties
Attempting to load validation.properties via file I/O.
Attempting to load validation.properties as resource file via file I/O.
Not found in 'org.owasp.esapi.resources' directory or file not readable: D:\Pentaho\biserver-ce\tomcat\bin\validation.properties
Not found in SystemResource Directory/resourceDirectory: .esapi\validation.properties
Not found in 'user.home' (C:\Users\talis) directory: C:\Users\talis\esapi\validation.properties
Loading validation.properties via file I/O failed.
Attempting to load validation.properties via the classpath.
validation.properties could not be loaded by any means. fail. Exception was: java.lang.IllegalArgumentException: Failed to load ESAPI.properties as a classloader resource.

```

FIGURA 48 - Inicialização do Servidor *Pentaho*

O *Pentaho BI Server* é o *front-end*¹ de interação com o usuário e dispõe de dois tipos de acesso conhecidos como *Pentaho User Console* – PUC e *Pentaho Administration Console* – PAC. Para acessar o PUC, ambiente das consultas OLAP, basta digitar no navegador de

¹ Termo geral utilizado para referenciar interfaces.

internet o endereço: *localhost:8080/pentaho*, com o servidor *Pentaho* iniciado, e escolher qual tipo de acesso desejado (FIGURA 49).



FIGURA 49 - Painel de Acesso do usuário.

Para realizar as consultas OLAP de maneira mais simples optou-se pela utilização do plugin *Saiku Analytics* - um cliente *web* que se conecta com o servidor OLAP permitindo realizar consultas aos cubos publicados anteriormente através de uma interface intuitiva *look and feel*.

O *download* do *Saiku Analytics* foi realizado diretamente do site do fabricante. Aós baixar o *plugin*, ele foi descompactado dentro da pasta *biserver-ce* da *Pentaho* (FIGURA 50). Para realizar esta ação, o servidor *Pentaho* foi interrompido.

Disco Local (D:) > Pentaho > biserver-ce > pentaho-solutions > system			
Nome	Data de modificação	Tipo	Tamanho
dtd	07/04/2016 15:18	Pasta de arquivos	
google	07/04/2016 15:18	Pasta de arquivos	
hibernate	07/04/2016 15:18	Pasta de arquivos	
jackrabbit	22/09/2017 19:25	Pasta de arquivos	
karaf	22/09/2017 19:23	Pasta de arquivos	
kettle	07/04/2016 15:18	Pasta de arquivos	
logs	22/09/2017 19:38	Pasta de arquivos	
marketplace	03/11/2017 16:57	Pasta de arquivos	
mondrian	07/04/2016 15:18	Pasta de arquivos	
monitoring-eventbus-plugin	07/04/2016 15:18	Pasta de arquivos	
osgi	07/04/2016 15:18	Pasta de arquivos	
pdi-pur-plugin	07/04/2016 15:18	Pasta de arquivos	
pentaho-cdf	07/04/2016 15:18	Pasta de arquivos	
pentaho-cdf-dd	07/04/2016 15:18	Pasta de arquivos	
pentaho-jpivot-plugin	07/04/2016 15:18	Pasta de arquivos	
pentaho-pdi-platform-plugin	07/04/2016 15:18	Pasta de arquivos	
quartz	07/04/2016 15:18	Pasta de arquivos	
reporting	07/04/2016 15:18	Pasta de arquivos	
saiku	24/04/2016 10:59	Pasta de arquivos	
simple-jndi	07/04/2016 15:18	Pasta de arquivos	
smtp-email	07/04/2016 15:18	Pasta de arquivos	
sparkl	07/04/2016 15:18	Pasta de arquivos	
test-plugin-perspective	07/04/2016 15:18	Pasta de arquivos	
ui	07/04/2016 15:18	Pasta de arquivos	

FIGURA 50 - Local de extração do plugin *Saiku Analytics*.

Com o *Pentaho BI Server* e o plugin *Saiku Analytics* configurados e o esquema *Mondrian* publicado através da ferramenta *Scheme Workbench*, foi iniciado o processo de análise dos dados através do PUC. Na tela inicial do PUC, basta clicar no botão *Create New* e em seguida escolher a opção *Saiku Analytics* (FIGURA 51).

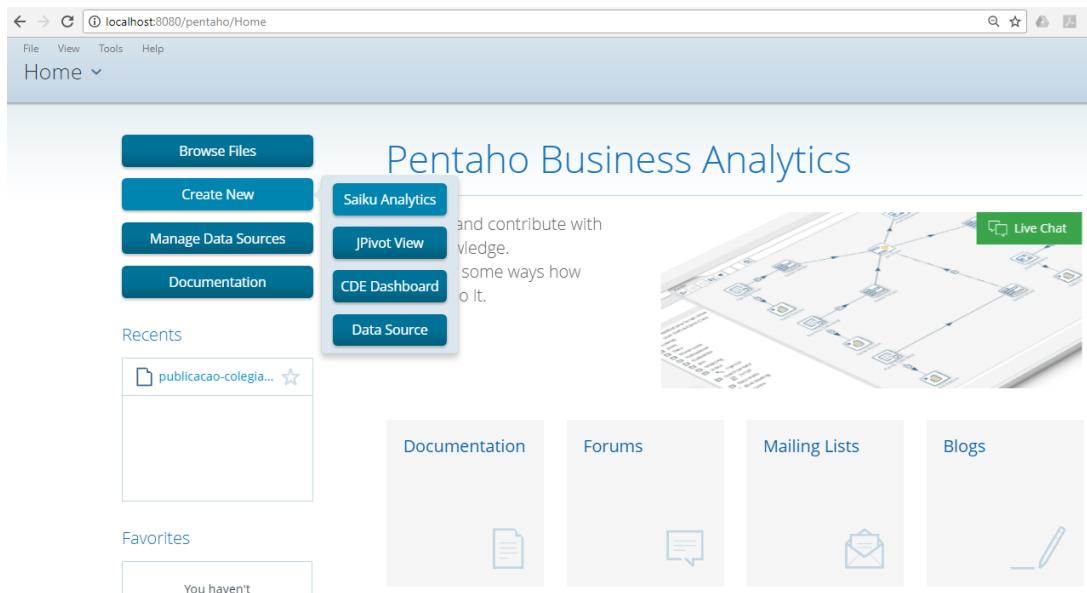


FIGURA 51 - Criando consultas no esquema OLAP através do *Saiku Analytics*.

Na tela inicial do *Saiku Analytics* ao selecionar o campo “*Create a new query*” uma área de trabalho aparecerá composta por quatro campos: (i) menu de acesso; (ii) na coluna do lado esquerdo da tela é onde o usuário pode selecionar o cubo, a medida e as dimensões que pretende consultar; (iv) na parte central da tela é onde aparecerá o resultado da consulta através dos parâmetros inicializados no campo (i); e (iv) na coluna do lado direito da tela é onde o usuário define o formato de visualização da consulta.

A partir desta interface o usuário pode realizar diversas consultas navegando através das dimensões do cubo selecionado, a cada interação do usuário uma tabela contendo o resultado da busca é apresentada ao lado direito da tela (FIGURA 52).

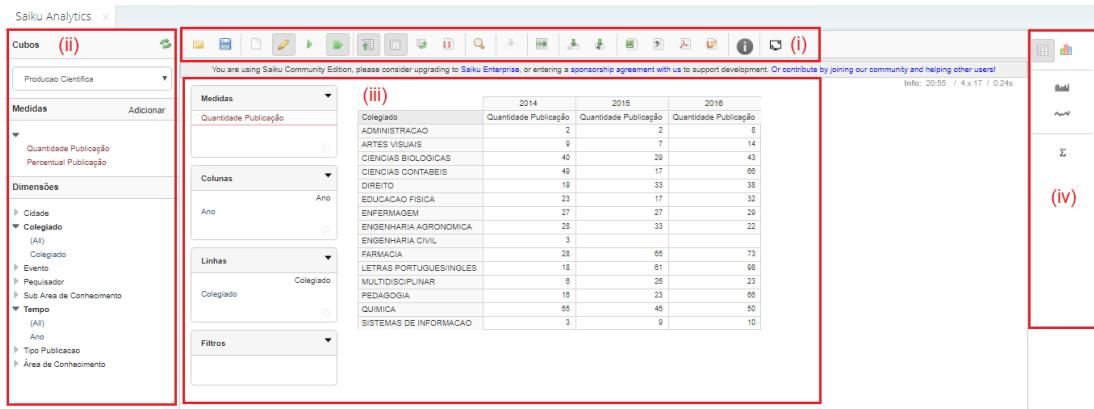


FIGURA 52 - Consulta da quantidade de publicações de colegiado por ano.

Além do formato em tabela, o *Saiku Analytics* também permite que o usuário visualize as informações através de gráficos. São treze modelos de gráficos que podem ser aplicados nas consultas e exportados nos formatos PDF, JPEG e PNG. Para visualizar as consultas em formato de gráfico, basta selecionar o botão que ilustra um gráfico de barras no lado direito da tela (FIGURA 53).

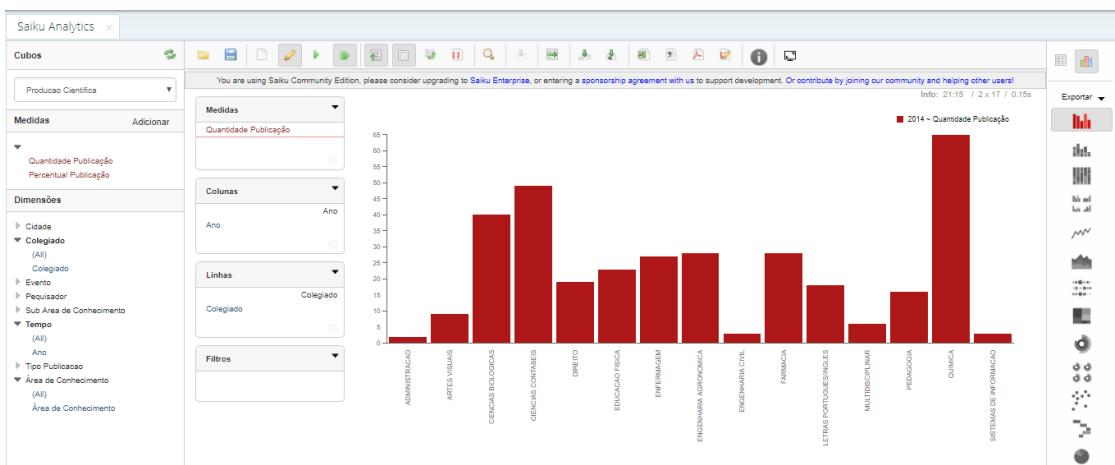


FIGURA 53 - Consulta da quantidade de publicações de colegiado por ano representada em um gráfico de barras.

5. ANÁLISE E DISCUSSÃO DOS RESULTADOS

O quadro 1 apresenta os tipos de produções científicas adaptados da plataforma Lattes do CNPq utilizados como indicadores para a análise da produção científica dos docentes do IFPR campus Palmas neste trabalho.

QUADRO 1 - Tipos de Produções Científicas adaptados do CNPq.

Tipo Agrupador da Produção	Tipo da Produção	Subtipo da Produção
Produção Artística Cultural	Artes Cênicas, Artes Visuais	Pintura
Produção Biblio gráfica	Artigos Publicados Em Periódicos	Completo
Produção Biblio gráfica	Trabalho Publicado Em Anais De Eventos	Resumo, Completo, Resumo Expandido
Produção Biblio gráfica	Capítulo De Livro Publicado	
Produção Biblio gráfica	Livro Publicado	
Produção Biblio gráfica	Texto Em Jornal Ou Revista	
Produção Técnica	Assessoria E Consultoria	Consultoria
Produção Técnica	Curso De Curta Duração Ministrado	Aperfeiçoamento, Extensão, Especialização
Produção Técnica	Desenvolvimento De Material Didático Ou Institucional	
Produção Técnica	Programa De Rádio Ou Tv	Mesa Redonda, Programa, Entrevista
Produção Técnica	Apresentação De Trabalho E Palestra	Congresso, Seminário, Conferência Ou Palestra, Simpósio
Evento	Participação Em Eventos	Seminário, Congresso, Simpósio, Exposição, Encontro

Os projetos do COPE estão cadastrados em sua base como projetos de ensino, pesquisa, extensão e inovação e estes indicadores são utilizados para a análise da produção dos docentes do IFPR campus Palmas neste trabalho.

Os dados foram analisados por meio das técnicas da estatística descritiva, com a determinação da quantidade, porcentagem e média de produção por colegiado.

Os resultados do estudo estão apresentados por meio das seguintes seções: a primeira seção descreve sobre a quantidade de produção científica por colegiado; a segunda seção apresenta a produção científica por colegiado dos trabalhos publicados em anais de eventos; a terceira seção apresenta a quantidade de produção científica por área de conhecimento e por fim apresenta-se a quantidade de projetos por colegiado do IFPR campus Palmas.

5. 1. PRODUÇÃO CIENTÍFICA DE ACORDO COM O CURRÍCULO LATTES

A tabela 1 apresenta a quantidade da produção científica dos docentes do IFPR campus Palmas, realizada no período de janeiro de 2014 à dezembro de 2016, separado pelo tipo agrupador da produção.

TABELA 1 - Produção Científica dos docentes do IFPR campus Palmas

Pesquisador	Tipo Agrupador	Todos os anos			Pesquisador	Tipo Agrupador	Todos os anos		
		2014	2015	2016			2014	2015	2016
DOCENTE 001	PRODUCAO BIBLIOGRAFICA	17	37	46	DOCENTE 069	PRODUCAO BIBLIOGRAFICA	2		2
	PRODUCAO TECNICA		2	1		PRODUCAO TECNICA		1	
	EVENTO		1	1		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL					OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO					EVENTO			0
		17	40	48			2	1	2
DOCENTE 002	PRODUCAO BIBLIOGRAFICA	37	2	35	DOCENTE 070	PRODUCAO TECNICA		2	2
	PRODUCAO TECNICA	1				PRODUCAO BIBLIOGRAFICA		1	
	PRODUCAO ARTISTICA/CULTURAL					PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO					OUTRO TIPO DE PRODUCAO			0
	EVENTO					EVENTO			0
		38	2	35			0	3	2
DOCENTE 003	PRODUCAO BIBLIOGRAFICA		13	23	DOCENTE 071	PRODUCAO BIBLIOGRAFICA		5	
	PRODUCAO TECNICA		2	31		PRODUCAO TECNICA			0
	PRODUCAO ARTISTICA/CULTURAL					PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO					OUTRO TIPO DE PRODUCAO			0
	EVENTO					EVENTO			0
		0	15	54			0	5	0
DOCENTE 004	PRODUCAO BIBLIOGRAFICA	22	16	18	DOCENTE 072	PRODUCAO BIBLIOGRAFICA		2	2
	PRODUCAO TECNICA	4	1			PRODUCAO TECNICA			1
	PRODUCAO ARTISTICA/CULTURAL					PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO					OUTRO TIPO DE PRODUCAO			0
	EVENTO					EVENTO			0
		26	17	18			0	2	3
DOCENTE 005	PRODUCAO BIBLIOGRAFICA	22	14	18	DOCENTE 073	PRODUCAO BIBLIOGRAFICA		3	1
	PRODUCAO TECNICA	4	1			PRODUCAO TECNICA		1	
	PRODUCAO ARTISTICA/CULTURAL					PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO					OUTRO TIPO DE PRODUCAO			0
	EVENTO					EVENTO			0
		26	15	18			0	4	1
DOCENTE 006	PRODUCAO BIBLIOGRAFICA	19	13	12	DOCENTE 074	PRODUCAO BIBLIOGRAFICA		1	3
	PRODUCAO TECNICA	1	3			PRODUCAO TECNICA		1	
	OUTRO TIPO DE PRODUCAO			1		PRODUCAO ARTISTICA/CULTURAL			0
	EVENTO			1		OUTRO TIPO DE PRODUCAO			0
	PRODUCAO ARTISTICA/CULTURAL			0		EVENTO			0
		20	16	14			0	2	3

DOCENTE 007	PRODUCAO BIBLIOGRAFICA	16	16	11	43	DOCENTE 075	PRODUCAO TECNICA	1	2		3
	EVENTO	2		2	4		PRODUCAO BIBLIOGRAFICA		2		2
	PRODUCAO TECNICA				0		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		18	16	13	47			1	4	0	5
DOCENTE 008	PRODUCAO BIBLIOGRAFICA	2	9	15	26	DOCENTE 076	PRODUCAO TECNICA			2	2
	PRODUCAO TECNICA	3	7	10	20		PRODUCAO BIBLIOGRAFICA		1	1	2
	OUTRO TIPO DE PRODUCAO		1		1		EVENTO		1	1	
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	EVENTO				0		OUTRO TIPO DE PRODUCAO				0
		5	17	25	47			0	1	4	5
DOCENTE 009	PRODUCAO BIBLIOGRAFICA	17	6	4	27	DOCENTE 077	PRODUCAO BIBLIOGRAFICA		1	3	4
	PRODUCAO TECNICA	9	8	3	20		PRODUCAO TECNICA			1	1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		26	14	7	47			0	1	4	5
DOCENTE 010	PRODUCAO TECNICA		5	21	26	DOCENTE 078	PRODUCAO BIBLIOGRAFICA			3	3
	PRODUCAO BIBLIOGRAFICA		5	12	17		PRODUCAO TECNICA			2	2
	EVENTO		2	1	3		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				1		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		0	12	35	47			0	0	5	5
DOCENTE 011	PRODUCAO BIBLIOGRAFICA	16	14	11	41	DOCENTE 079	PRODUCAO TECNICA		2		2
	PRODUCAO TECNICA	1			1		PRODUCAO BIBLIOGRAFICA		1		1
	PRODUCAO ARTISTICA/CULTURAL				0		EVENTO		1		1
	OUTRO TIPO DE PRODUCAO				0		PRODUCAO ARTISTICA/CULTURAL				0
	EVENTO				0		OUTRO TIPO DE PRODUCAO				0
		17	14	11	42			0	4	0	4
DOCENTE 012	PRODUCAO BIBLIOGRAFICA	13	7	12	32	DOCENTE 080	PRODUCAO BIBLIOGRAFICA	1	2		3
	PRODUCAO TECNICA		2	3	5		PRODUCAO TECNICA	1			1
	EVENTO		1		1		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		13	10	15	38			2	2	0	4
DOCENTE 013	PRODUCAO BIBLIOGRAFICA	19	6	6	31	DOCENTE 081	PRODUCAO BIBLIOGRAFICA			3	3
	PRODUCAO TECNICA		1		1		PRODUCAO TECNICA			1	1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		19	7	6	32			0	0	4	4
DOCENTE 014	PRODUCAO BIBLIOGRAFICA	15	1	11	27	DOCENTE 082	PRODUCAO TECNICA			2	2
	EVENTO	1		3	4		PRODUCAO BIBLIOGRAFICA			1	1
	PRODUCAO TECNICA				0		PRODUCAO ARTISTICA/CULTURAL		1		1

	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		16	1	14	31				0	1	3
DOCENTE 015	PRODUCAO TECNICA		11	7	18	DOCENTE 083	PRODUCAO TECNICA	1		1	2
	PRODUCAO BIBLIOGRAFICA		5	3	8		PRODUCAO BIBLIOGRAFICA			1	1
	EVENTO			1	1		EVENTO	1			1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
		0	16	11	27				2	0	2
DOCENTE 016	PRODUCAO BIBLIOGRAFICA	1	7	7	15	DOCENTE 084	PRODUCAO BIBLIOGRAFICA			3	3
	PRODUCAO TECNICA	4	6	2	12		PRODUCAO TECNICA			1	1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		5	13	9	27				0	0	4
DOCENTE 017	PRODUCAO BIBLIOGRAFICA	9	4	8	21	DOCENTE 085	PRODUCAO BIBLIOGRAFICA			3	1
	PRODUCAO TECNICA	5		1	6		PRODUCAO TECNICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		14	4	9	27				0	3	1
DOCENTE 018	PRODUCAO TECNICA	2	2	10	14	DOCENTE 086	PRODUCAO TECNICA			2	2
	PRODUCAO BIBLIOGRAFICA	1	4	2	7		PRODUCAO BIBLIOGRAFICA			1	1
	EVENTO	2	2	1	5		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		5	8	13	26				0	0	3
DOCENTE 019	PRODUCAO BIBLIOGRAFICA	9	1	3	13	DOCENTE 087	PRODUCAO BIBLIOGRAFICA			2	2
	PRODUCAO TECNICA	6	3	2	11		PRODUCAO TECNICA			1	1
	EVENTO	1	1		2		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		16	5	5	26				0	2	1
DOCENTE 020	EVENTO	2	4	6	12	DOCENTE 088	PRODUCAO BIBLIOGRAFICA			2	1
	PRODUCAO TECNICA	4	2	1	7		PRODUCAO TECNICA				0
	PRODUCAO BIBLIOGRAFICA	2	3		5		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		8	9	7	24				0	2	1
DOCENTE 021	PRODUCAO TECNICA		2	14	16	DOCENTE 089	PRODUCAO BIBLIOGRAFICA			3	3
	PRODUCAO BIBLIOGRAFICA		3	4	7		PRODUCAO TECNICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		0	5	18	23				0	0	3
DOCENTE 022	PRODUCAO BIBLIOGRAFICA		6	6	12	DOCENTE 090	PRODUCAO BIBLIOGRAFICA			3	3
	PRODUCAO TECNICA		3	5	8		PRODUCAO TECNICA				0

	EVENTO	2	2			PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL			0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO			0		EVENTO			0
		0	11	11	22			0	0 3 3
DOCENTE 023	PRODUCAO BIBLIOGRAFICA	7	4	4	15	DOCENTE 091	PRODUCAO TECNICA		2 2
	PRODUCAO TECNICA	2	2	2	6		PRODUCAO BIBLIOGRAFICA		1 1
	EVENTO	1			1		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		10	6	6	22			0	0 3 3
DOCENTE 024	PRODUCAO BIBLIOGRAFICA	10	1	6	17	DOCENTE 092	PRODUCAO BIBLIOGRAFICA		1 2 3
	PRODUCAO TECNICA		2	1	3		PRODUCAO TECNICA		0
	EVENTO			1	1		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		10	3	8	21			0	1 2 3
DOCENTE 025	PRODUCAO BIBLIOGRAFICA	3	4	12	19	DOCENTE 093	PRODUCAO TECNICA	2	2
	PRODUCAO TECNICA	1			1		PRODUCAO BIBLIOGRAFICA		1 1
	EVENTO	1			1		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		5	4	12	21			2	0 1 3
DOCENTE 026	PRODUCAO BIBLIOGRAFICA	5	1	7	13	DOCENTE 094	PRODUCAO TECNICA		2 2
	PRODUCAO TECNICA		2	5	7		PRODUCAO BIBLIOGRAFICA		1 1
	EVENTO	1			1		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		6	3	12	21			0	0 3 3
DOCENTE 027	PRODUCAO BIBLIOGRAFICA	4	2	11	17	DOCENTE 095	PRODUCAO BIBLIOGRAFICA	3	3
	PRODUCAO TECNICA		2	1	3		PRODUCAO TECNICA		0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL		0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO		0
	EVENTO				0		EVENTO		0
		4	4	12	20			3	0 0 3
DOCENTE 028	PRODUCAO BIBLIOGRAFICA	1	5	6	12	DOCENTE 096	PRODUCAO BIBLIOGRAFICA		2 2
	PRODUCAO TECNICA		3	4	7		PRODUCAO TECNICA		1 1
	EVENTO			1	1		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		1	8	11	20			0	2 1 3
DOCENTE 029	PRODUCAO TECNICA		9	3	12	DOCENTE 097	PRODUCAO TECNICA	1	1
	PRODUCAO BIBLIOGRAFICA	2	2	1	5		PRODUCAO BIBLIOGRAFICA		1 1
	EVENTO		1	1	2		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0

		2	12	5	19			0	1	1	2
DOCENTE 030	PRODUCAO BIBLIOGRAFICA	11	4	2	17	DOCENTE 098	PRODUCAO TECNICA			2	2
	PRODUCAO TECNICA		1		1		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		11	5	2	18				0	0	2
DOCENTE 031	PRODUCAO BIBLIOGRAFICA	8	10		18	DOCENTE 099	PRODUCAO TECNICA			1	1
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA	1			1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		8	10	0	18				1	0	1
DOCENTE 032	PRODUCAO BIBLIOGRAFICA	6	2	7	15	DOCENTE 100	PRODUCAO TECNICA		1		1
	PRODUCAO TECNICA		2		2		PRODUCAO BIBLIOGRAFICA			1	1
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		6	4	7	17				0	1	1
DOCENTE 033	PRODUCAO TECNICA	1	4	7	12	DOCENTE 101	PRODUCAO BIBLIOGRAFICA	1		1	2
	PRODUCAO BIBLIOGRAFICA		1	4	5		PRODUCAO TECNICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		1	5	11	17				1	0	1
DOCENTE 034	PRODUCAO TECNICA		3	9	12	DOCENTE 102	PRODUCAO BIBLIOGRAFICA		2		2
	PRODUCAO BIBLIOGRAFICA		1	3	4		PRODUCAO TECNICA				0
	EVENTO			1	1		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		0	4	13	17				0	2	0
DOCENTE 035	PRODUCAO BIBLIOGRAFICA	6	3	3	12	DOCENTE 103	PRODUCAO BIBLIOGRAFICA			2	2
	PRODUCAO TECNICA	4	1		5		PRODUCAO TECNICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		10	4	3	17				0	0	2
DOCENTE 036	PRODUCAO BIBLIOGRAFICA	11		3	14	DOCENTE 104	PRODUCAO BIBLIOGRAFICA			2	2
	PRODUCAO TECNICA		1		1		PRODUCAO TECNICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		11	1	3	15				0	0	2
DOCENTE 037	PRODUCAO TECNICA	6	2	2	10	DOCENTE 105	PRODUCAO TECNICA	2			2
	PRODUCAO ARTISTICA/CULTURAL		1	1	2		PRODUCAO BIBLIOGRAFICA				0
	EVENTO		2		2		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO BIBLIOGRAFICA				0		OUTRO TIPO DE PRODUCAO				0

	OUTRO TIPO DE PRODUCAO			0		EVENTO				0
		6	5	3	14			2	0	0
DOCENTE 038	PRODUCAO BIBLIOGRAFICA			7	7	DOCENTE 106	PRODUCAO BIBLIOGRAFICA			2
	PRODUCAO TECNICA		3	3	6		PRODUCAO TECNICA			0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO			0
	EVENTO				0		EVENTO			0
		0	3	10	13			0	0	2
DOCENTE 039	PRODUCAO BIBLIOGRAFICA		2	5	7	DOCENTE 107	EVENTO			1
	PRODUCAO TECNICA		1	5	6		PRODUCAO TECNICA			0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO BIBLIOGRAFICA			0
	OUTRO TIPO DE PRODUCAO				0		PRODUCAO ARTISTICA/CULTURAL			0
	EVENTO				0		OUTRO TIPO DE PRODUCAO			0
		0	3	10	13			0	0	1
DOCENTE 040	PRODUCAO BIBLIOGRAFICA		4	2	6	DOCENTE 108	PRODUCAO BIBLIOGRAFICA			1
	EVENTO		2	3	5		PRODUCAO TECNICA			0
	PRODUCAO TECNICA		1		1		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO				0		EVENTO			0
		0	7	5	12			0	0	1
DOCENTE 041	PRODUCAO BIBLIOGRAFICA	5	2	4	11	DOCENTE 109	PRODUCAO TECNICA			1
	EVENTO	1			1		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO TECNICA				0		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO				0		EVENTO			0
		6	2	4	12			0	0	1
DOCENTE 042	PRODUCAO BIBLIOGRAFICA	2	2	7	11	DOCENTE 110	EVENTO			1
	PRODUCAO TECNICA				0		PRODUCAO TECNICA			0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO BIBLIOGRAFICA			0
	OUTRO TIPO DE PRODUCAO				0		PRODUCAO ARTISTICA/CULTURAL			0
	EVENTO				0		OUTRO TIPO DE PRODUCAO			0
		2	2	7	11			0	0	1
DOCENTE 043	PRODUCAO BIBLIOGRAFICA	2	1	6	9	DOCENTE 111	PRODUCAO TECNICA		1	1
	PRODUCAO TECNICA		2		2		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO			0
	EVENTO				0		EVENTO			0
		2	3	6	11			0	1	0
DOCENTE 044	PRODUCAO BIBLIOGRAFICA		5	1	6	DOCENTE 112	PRODUCAO TECNICA		1	1
	PRODUCAO TECNICA		2	1	3		PRODUCAO BIBLIOGRAFICA			0
	EVENTO		1		1		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO				0		EVENTO			0
		0	8	2	10			0	1	0
DOCENTE 045	PRODUCAO BIBLIOGRAFICA	1	2	4	7	DOCENTE 113	PRODUCAO BIBLIOGRAFICA			1
	PRODUCAO TECNICA		2	1	3		PRODUCAO TECNICA			0

	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		1	4	5	10			0	0
DOCENTE 046	PRODUCAO TECNICA	3	2	5	DOCENTE 114	PRODUCAO TECNICA		1	1
	PRODUCAO BIBLIOGRAFICA	1	1	3		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		4	3	3	10			0	1
DOCENTE 047	PRODUCAO BIBLIOGRAFICA		7	1	8	DOCENTE 115	PRODUCAO BIBLIOGRAFICA		1
	PRODUCAO TECNICA		2		2		PRODUCAO TECNICA		0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL		0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO		0
	EVENTO				0		EVENTO		0
		0	9	1	10			0	1
DOCENTE 048	PRODUCAO BIBLIOGRAFICA	2	3	1	6	DOCENTE 116	PRODUCAO TECNICA		0
	PRODUCAO TECNICA	1	2		3		PRODUCAO BIBLIOGRAFICA		0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL		0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO		0
	EVENTO				0		EVENTO		0
		3	5	1	9			0	0
DOCENTE 049	PRODUCAO BIBLIOGRAFICA		3	5	8	DOCENTE 117	PRODUCAO TECNICA		0
	EVENTO		1		1		PRODUCAO BIBLIOGRAFICA		0
	PRODUCAO TECNICA				0		PRODUCAO ARTISTICA/CULTURAL		0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO		0
	OUTRO TIPO DE PRODUCAO				0		EVENTO		0
		0	4	5	9			0	0
DOCENTE 050	PRODUCAO BIBLIOGRAFICA		6	2	8	DOCENTE 118	PRODUCAO TECNICA		0
	PRODUCAO TECNICA			1	1		PRODUCAO BIBLIOGRAFICA		0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL		0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO		0
	EVENTO				0		EVENTO		0
		0	6	3	9			0	0
DOCENTE 051	PRODUCAO TECNICA	3	1	2	6	DOCENTE 119	PRODUCAO TECNICA		0
	PRODUCAO BIBLIOGRAFICA	1			1		PRODUCAO BIBLIOGRAFICA		0
	OUTRO TIPO DE PRODUCAO			1	1		PRODUCAO ARTISTICA/CULTURAL		0
	EVENTO			1	1		OUTRO TIPO DE PRODUCAO		0
	PRODUCAO ARTISTICA/CULTURAL				0		EVENTO		0
		4	1	4	9			0	0
DOCENTE 052	PRODUCAO BIBLIOGRAFICA	3		6	9	DOCENTE 120	PRODUCAO TECNICA		0
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA		0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL		0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO		0
	EVENTO				0		EVENTO		0
		3	0	6	9			0	0
DOCENTE 053	PRODUCAO BIBLIOGRAFICA			5	5	DOCENTE 121	PRODUCAO TECNICA		0

	PRODUCAO TECNICA		2	2		PRODUCAO BIBLIOGRAFICA			0
	EVENTO		2	2		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL			0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO			0		EVENTO			0
		0	0	9				0	0
DOCENTE 054	PRODUCAO BIBLIOGRAFICA		5	3	DOCENTE 122	PRODUCAO TECNICA			0
	PRODUCAO TECNICA			0		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		0	5	3				0	0
DOCENTE 055	PRODUCAO BIBLIOGRAFICA	3		3	DOCENTE 123	PRODUCAO TECNICA			0
	EVENTO		1	1		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO TECNICA			0		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL			0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO			0		EVENTO			0
		3	1	4				0	0
DOCENTE 056	PRODUCAO BIBLIOGRAFICA			5	DOCENTE 124	PRODUCAO TECNICA			0
	PRODUCAO TECNICA	2		1		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		2	0	6				0	0
DOCENTE 057	PRODUCAO BIBLIOGRAFICA		2	4	DOCENTE 125	PRODUCAO TECNICA			0
	PRODUCAO TECNICA		1	1		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		0	3	5				0	0
DOCENTE 058	PRODUCAO TECNICA	1	1	2	DOCENTE 126	PRODUCAO TECNICA			0
	PRODUCAO BIBLIOGRAFICA			1		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0
		1	2	4				0	0
DOCENTE 059	PRODUCAO TECNICA	1	1	3	DOCENTE 127	PRODUCAO TECNICA			0
	PRODUCAO BIBLIOGRAFICA			1		PRODUCAO BIBLIOGRAFICA			0
	EVENTO			1		PRODUCAO ARTISTICA/CULTURAL			0
	PRODUCAO ARTISTICA/CULTURAL			0		OUTRO TIPO DE PRODUCAO			0
	OUTRO TIPO DE PRODUCAO			0		EVENTO			0
		1	1	5				0	0
DOCENTE 060	PRODUCAO BIBLIOGRAFICA		3	2	DOCENTE 128	PRODUCAO TECNICA			0
	PRODUCAO TECNICA			2		PRODUCAO BIBLIOGRAFICA			0
	PRODUCAO ARTISTICA/CULTURAL			0		PRODUCAO ARTISTICA/CULTURAL			0
	OUTRO TIPO DE PRODUCAO			0		OUTRO TIPO DE PRODUCAO			0
	EVENTO			0		EVENTO			0

		0	3	4	7			0	0	0	0
DOCENTE 061	PRODUCAO TECNICA			3	3	DOCENTE 129	PRODUCAO TECNICA				0
	EVENTO			3	3		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO BIBLIOGRAFICA			1	1		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		0	0	7	7			0	0	0	0
DOCENTE 062	PRODUCAO BIBLIOGRAFICA			4	4	DOCENTE 130	PRODUCAO TECNICA				0
	PRODUCAO TECNICA			2	2		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		0	0	6	6			0	0	0	0
DOCENTE 063	PRODUCAO BIBLIOGRAFICA	1		3	4	DOCENTE 131	PRODUCAO TECNICA				0
	PRODUCAO TECNICA			2	2		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		1	0	5	6			0	0	0	0
DOCENTE 064	PRODUCAO BIBLIOGRAFICA	5		1	6	DOCENTE 132	PRODUCAO TECNICA				0
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		5	0	1	6			0	0	0	0
DOCENTE 065	EVENTO			3	3	DOCENTE 133	PRODUCAO TECNICA				0
	PRODUCAO BIBLIOGRAFICA			2	2		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO TECNICA			1	1		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		0	0	6	6			0	0	0	0
DOCENTE 066	PRODUCAO BIBLIOGRAFICA	2	3	1	6	DOCENTE 134	PRODUCAO TECNICA				0
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0
	OUTRO TIPO DE PRODUCAO				0		OUTRO TIPO DE PRODUCAO				0
	EVENTO				0		EVENTO				0
		2	3	1	6			0	0	0	0
DOCENTE 067	EVENTO			6	6	DOCENTE 135	PRODUCAO TECNICA				0
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO BIBLIOGRAFICA				0		PRODUCAO ARTISTICA/CULTURAL				0
	PRODUCAO ARTISTICA/CULTURAL				0		OUTRO TIPO DE PRODUCAO				0
	OUTRO TIPO DE PRODUCAO				0		EVENTO				0
		0	0	6	6			0	0	0	0
DOCENTE 068	PRODUCAO BIBLIOGRAFICA	3		2	5	DOCENTE 136	PRODUCAO TECNICA				0
	PRODUCAO TECNICA				0		PRODUCAO BIBLIOGRAFICA				0
	PRODUCAO ARTISTICA/CULTURAL				0		PRODUCAO ARTISTICA/CULTURAL				0

	OUTRO TIPO DE PRODUÇÃO				0		OUTRO TIPO DE PRODUÇÃO					0
	EVENTO				0		EVENTO					0
		3	0	2	5				0	0	0	0
								TOTAL POR ANO	439	471	726	

A tabela 2 apresenta uma análise estatística descritiva referente a quantidade mínima e máxima, a média, a variância e o desvio padrão da produção científica dos pesquisadores do IFPR campus Palmas ano a ano.

TABELA 2 - Estatística descritiva da Produção Científica dos docentes do IFPR campus Palmas por ano

2014		2015		2016	
MIN	0	MIN	0	MIN	0
MÁX	38	MÁX	40	MÁX	48
MÉDIA	3,23	MÉDIA	3,46	MÉDIA	5,34
VARIANCIAS	41,73	VARIANCIAS	29,37	VARIANCIAS	67,32
DESVIO PADRÃO	6,46	DESVIO PADRÃO	5,42	DESVIO PADRÃO	8,2

Apesar de existirem pesquisadores com um número alto de produção, com até 38 produções em 2014, 40 em 2015 e 48 em 2016 a média de produção por pesquisador não passou de 5,34 nos três anos, assim como o desvio padrão que se manteve entre um intervalo de 5,42 e 8,20. Para entender um pouco o motivo do desvio padrão e a média estarem tão abaixo da quantidade máxima de produções por ano recorreu-se aos dados da tabela 1.

De acordo com a tabela 1, é possível levantar algumas informações a respeito da produção científica nos últimos três anos de cada pesquisador, como: 15,44% dos docentes **não têm** produção nos últimos 3 anos; 35,29% dos docentes têm **de 1 a 5** produções nos últimos 3 anos; 17,64% docentes têm **de 6 a 10** produções nos últimos 3 anos; 5,88% dos docentes têm **de 11 a 15** produções nos últimos 3 anos; 7,35% dos docentes têm **de 16 a 20** produções nos últimos 3 anos; e 18,38% têm **mais de 20** produções nos últimos 3 anos.

Este resultado justifica o desvio padrão apresentado na tabela 2, uma vez que, apesar do número máximo de produção por pesquisador tenha sido igual ou maior que 38 nos três anos, apenas 18,38% dos pesquisadores tiveram mais de 20 produções nos últimos 3 anos, enquanto a grande maioria, 35,29% dos pesquisadores se mantém na média anual, possuindo de 1 a 5 produções nos últimos 3 anos.

Conforme apresentado na Figura 54, mais de 75% possuíram menos de 10 produções por ano, nos últimos três anos. Porém, os pontos discrepantes, ou seja, a quantidade de professores com produções muito acima da média tem diminuído com o passar dos anos, isto porque a média da produção por ano de cada professor tem aumentado a cada ano, mesmo que pouco, de acordo com a posição central dos dados (mediana) apresentado.

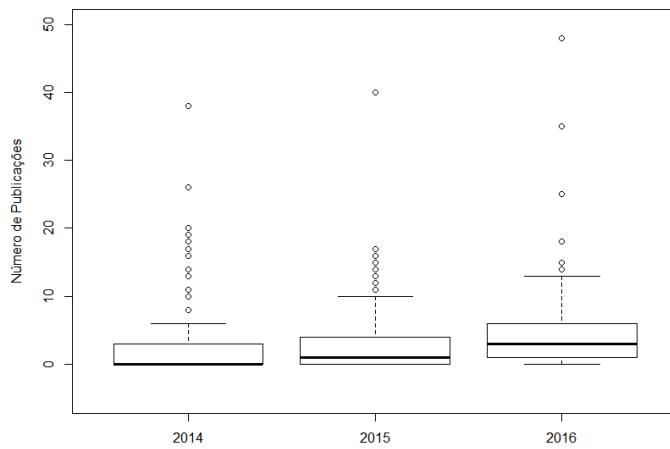


FIGURA 54 - Gráfico Boxplot da Produção Científica dos docentes do IFPR campus Palmas

Outro fato interessante é que apesar de serem identificadas 1251 produções, esta análise apresenta uma somatória de 1636 produções nos três últimos anos. Isso significa que 385 das produções apresentadas possuem mais de um autor.

5. 2. PRODUÇÃO CIENTÍFICA PUBLICADA EM ANAIS DE EVENTOS

A tabela 3 apresenta a relação do tipo de produção científica publicada em anais de eventos segregada em Resumo, Completo e Resumo Expandido e a quantidade realizada em cada curso/colegiado no IFPR campus Palmas, no período de janeiro de 2014 a dezembro de 2016.

TABELA 3 - Quantidade de Produção Científica de acordo com o tipo de trabalho publicado em eventos de cada colegiado

Colegiado	2014			Tot	2015			Tot	2016			Tot	Média Colegiado
	COMP.	RES.	RES. EXP.		196	COMP.	RES.		193	COMP.	RES.	RES. EXP.	
				Quantidade									
ADMINISTRACAO	0	0	0	0	1	0	0	1	4	0	0	4	0,56
ARTES VISUAIS	0	0	0	0	0	0	0	0	2	1	1	4	1,33
CIENCIAS BIOLOGICAS	4	12	15	31	6	17	4	27	7	8	19	34	30,67
CIENCIAS CONTABEIS	9	0	0	9	6	0	1	7	17	2	3	22	12,67
DIREITO	11	1	0	12	10	0	0	10	20	0	0	20	14,00
EDUCACAO FISICA	4	9	10	23	4	5	0	9	2	13	7	22	18,00
ENFERMAGEM	3	17	4	24	8	8	1	17	9	6	1	16	19,00
ENGENHARIA AGRONOMICA	10	10	1	21	16	3	5	24	9	6	2	17	20,67
ENGENHARIA CIVIL	1	0	2	3	0	0	0	0	0	0	0	0	1,00
FARMACIA	2	3	0	5	8	10	2	20	9	5	5	19	14,67

LETROS PORTUGUES/INGLES	3	1	0	4	14	2	1	17	14	9	6	29	16,67
MULTIDISCIPLINAR	1	0	0	1	10	1	3	14	6	0	0	6	7,00
PEDAGOGIA	9	0	4	13	4	1	1	6	24	1	5	30	16,33
QUIMICA	6	20	22	48	7	15	14	36	4	8	19	31	38,33
SISTEMAS DE INFORMACAO	1	0	1	2	3	1	1	5	0	2	3	5	4,00

Os resultados demonstram 648 produções científicas publicada em anais de eventos segregada em Resumo, Completo e Resumo Expandido o que remete à uma média geral de 43,2 produções por colegiado. A maior produção científica média por curso/colegiado é de 38,33 dos docentes do colegiado do curso de Licenciatura em Química, que está vinculado à área de conhecimento de exatas e da terra.

A segunda maior produção científica está relacionada aos docentes do colegiado do curso de Licenciatura em Ciências Biológicas, com uma média de 30,67 produções nos últimos três anos.

Dos três tipos de produção analisado, a modalidade de artigos completos é a que teve um maior número de resultados, representando 44,9% das produções analisadas, seguida pela modalidade de Resumo com 30,6% e Resumos Expandidos com 24,5% (FIGURA 54).

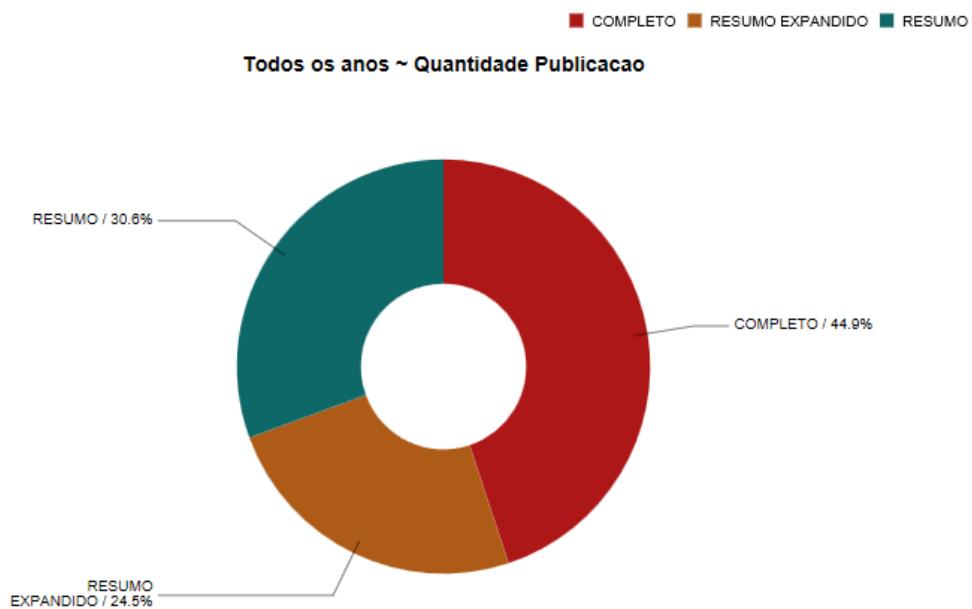


FIGURA 55 - Gráfico das Produções Científicas de acordo com o tipo de publicação

Verificando os dados da estatística descritiva, ano a ano, é possível identificar a média, a quantidade mínima e máxima, a variância e os desvio padrão das publicações realizadas em anais e eventos (TABELA 4). Neste caso, os valores da média e do desvio padrão estão mais aproximados, diferente da análise realizada por pesquisador/docente. Isto acontece pois, nos últimos três anos todos os colegiados tiveram publicações. 33,33% dos colegiados tiveram de

3 a 25 publicações; 33,33% dos colegiados tiveram de **26 a 50** publicações; 20% dos colegiados tiveram de **51 a 75** publicações e apenas 13,33% dos colegiados tiveram **mais de 75** publicações.

TABELA 4 - Análise estatística descritiva das produções científicas por tipo de produção e colegiado .

2014		2015		2016	
Mínimo	0	Mínimo	0	Mínimo	0
Máximo	48	Máximo	36	Máximo	34
Média	13,07	Média	12,87	Média	17,27
Variância	179,93	Variância	105,58	Variância	116,86
Desvio Padrão	13,41	Desvio Padrão	10,28	Desvio Padrão	10,81

5. 3. PRODUÇÃO CIENTÍFICA POR ÁREA DE CONHECIMENTO

A tabela 5 apresenta a quantidade da produção científica dos docentes do IFPR campus Palmas de acordo com a área de conhecimento, realizada no período de janeiro de 2014 à dezembro de 2016.

TABELA 5 - Quantidade da Produção Científica por Área de Conhecimento

Área de Conhecimento	Quantidade Publicação				
	2014	2015	2016	Total	Média
CIENCIAS DA SAUDE	78	109	134	321	107,00
CIENCIAS SOCIAIS APLICADAS	70	52	112	234	78,00
LINGUISTICA, LETRAS E ARTES	27	67	110	204	68,00
CIENCIAS EXATAS E DA TERRA	68	55	60	183	61,00
CIENCIAS BIOLOGICAS	40	29	43	112	37,33
CIENCIAS HUMANAS	16	23	66	105	35,00
CIENCIAS AGRARIAS	28	33	22	83	27,67
MULTIDISCIPLINAR	6	26	23	55	18,33
ENGENHARIAS	3	0	0	3	1,00
Total	336	394	570		

A área de conhecimento Ciências da Saúde detém o maior número de produção nos últimos três anos (321), com uma média de 107 produções por ano. No IFPR campus Palmas, os colegiados que atuam nessa área são dos cursos bacharelado em Enfermagem com 83 produções nos últimos três anos, bacharelado em Farmácia com 166 produções nos últimos 3 anos e curso Licenciatura em Educação Física com 72 produções nos últimos 3 anos.

O segundo maior número de produções pertence à área de Ciências Sociais e Aplicadas, com 234 produções, obtendo uma média de 78 produções por ano. Os colegiados que publicam nessa área, no IFPR campus Palmas, são os cursos de bacharelado em: Administração com 12 produções nos últimos 3 anos, Ciências Contábeis com 132 produções

nos últimos 3 anos e Direito com 90 produções nos últimos 3 anos. A representação gráfica da produção científica por área de conhecimento é apresentada na figura 55.

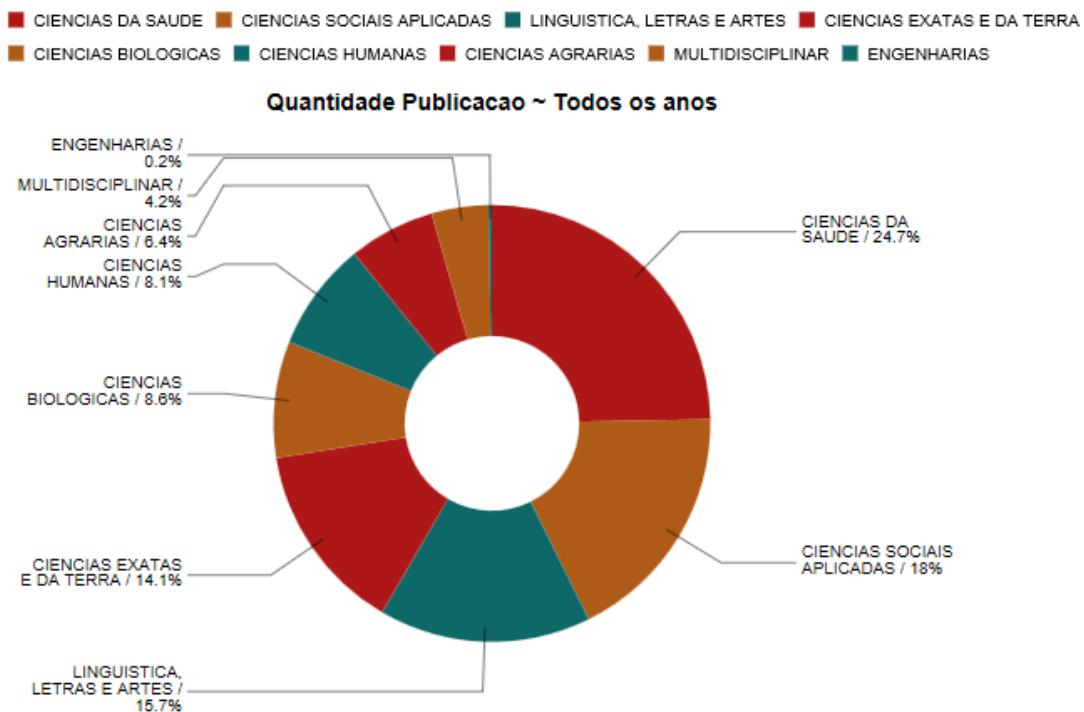


FIGURA 56 - Gráfico da Produção Científica por Áreas de Conhecimento

A tabela 6 apresenta uma análise estatística descritiva das produções científicas organizadas por área de conhecimento e por ano com informações referente ao número mínimo e máximo, média, variância e desvio padrão da produção por área.

TABELA 6 - Análise estatística descritiva das produções científicas por área de conhecimento e ano.

2014		2015		2016	
Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
3	78	0	109	0	134
37,33	43,78	43,78	63,33	63,33	63,33
719,78	887,28	887,28	1924,22	1924,22	1924,22
26,83	Desvio Padrão	29,79	Desvio Padrão	43,87	Desvio Padrão

Analizando ano a ano, é possível verificar que a quantidade mínima de produção chegou a 0 em 2015 e 2016. A área de conhecimento que não obteve produção nesses anos foi a área de Engenharia. No IFPR campus Palmas existe apenas um colegiado que atua nesta área, o colegiado do curso de bacharelado em Engenharia Civil que conta com um número reduzido de professores e desde o ano de 2014 não tem participado dos processos seletivos para abertura de novas turmas. Apesar desta situação, é possível perceber um número significativo de produções científicas por ano, onde de 2014 à 2016 a média de produção atingiu um aumento de 69,65%.

Ainda sobre os dados da produção científica por área de conhecimento, a somatória dos dados é de 1300 produções, o que significa que 49 das produções realizadas contam com a parceria entre duas – ou mais – áreas de conhecimento.

5. 4. PROJETOS DESENVOLVIDOS DE ACORDO COM A BASE DE DADOS DO COPE DO IFPR CAMPUS PALMAS

A tabela 7 apresenta os projetos de ensino, pesquisa, extensão e inovação dos docentes dos colegiados do IFPR campus Palmas, realizada no período de janeiro de 2014 à dezembro de 2016.

TABELA 7- Quantidade de projetos de ensino, pesquisa, extensão e inovação dos docentes dos colegiados do IFPR campus Palmas

Colegiado	Tipo Projeto	2014	2015	2016	total	média/ano
ADMINISTRACAO	EXTENSAO	2	2	3	14	4,67
	PESQUISA	2	2	3		
ARTES VISUAIS	EXTENSAO	2	2	2	9	3,00
	PESQUISA	1	1	1		
CIENCIAS BIOLOGICAS	EXTENSAO	3	1	4	18	6,00
	PESQUISA	2	1	7		
CIENCIAS CONTABEIS	ENSINO	0	0	1	27	9,00
	EXTENSAO	6	4	0		
	PESQUISA	4	10	2		
DIREITO	EXTENSAO	7	0	1	19	6,33
	PESQUISA	0	7	4		
EDUCACAO FISICA	ENSINO	0	0	2	13	4,33
	EXTENSAO	1	0	4		
	PESQUISA	2	2	2		
ENFERMAGEM	EXTENSAO	2	2	3	22	7,33
	PESQUISA	1	8	6		
ENGENHARIA AGRONOMICA	EXTENSAO	3	2	1	36	12,00
	PESQUISA	7	11	12		
ENGENHARIA CIVIL	PESQUISA	0	1	1	2	0,67
FARMACIA	ENSINO	0	0	1	30	10,00
	EXTENSAO	5	6	1		
	PESQUISA	3	5	9		
LETROS PORTUGUES/INGLES	ENSINO	0	0	1	25	8,33
	EXTENSAO	2	6	4		
	PESQUISA	5	5	2		
MULTIDISCIPLINAR	EXTENSAO	0	3	2	14	4,67
	PESQUISA	1	4	4		
PEDAGOGIA	ENSINO	0	0	1	14	4,67
	EXTENSAO	4	2	3		
	PESQUISA	1	3	0		
QUIMICA	EXTENSAO	2	0	5	14	4,67
	PESQUISA	1	1	5		
SISTEMAS DE INFORMACAO	EXTENSAO	4	0	2	14	4,67
	PESQUISA	0	4	4		

O colegiado com o maior número de projetos submetidos ao COPE é do curso bacharelado em Engenharia Agronômica com 36 projetos, sendo 6 projetos de extensão e 30 projetos de pesquisa, somando 13,2% do número de projetos da instituição nos últimos 3 anos, com uma média de 12 projetos/ano.

O segundo colegiado com maior número de projetos submetidos ao COPE é do curso bacharelado em Farmácia, com 30 projetos sendo 17 projetos de Pesquisa, 12 projetos de extensão e 1 projeto de ensino, totalizando 11% dos projetos mantidos pelo COPE, com uma média de 10 projetos/ano. A porcentagem que cada colegiado ocupa na quantidade de projetos mantido pelo COPE pode ser analisado abaixo (FIGURA 57).

■ ADMINISTRACAO ■ ARTES VISUAIS ■ CIENCIAS BIOLOGICAS ■ CIENCIAS CONTABEIS ■ DIREITO ■ EDUCACAO FISICA
 ■ ENFERMAGEM ■ ENGENHARIA AGRONOMICA ■ ENGENHARIA CIVIL ■ FARMACIA ■ LETRAS PORTUGUES/INGLES
 ■ MULTIDISCIPLINAR ■ PEDAGOGIA ■ QUIMICA ■ SISTEMAS DE INFORMACAO

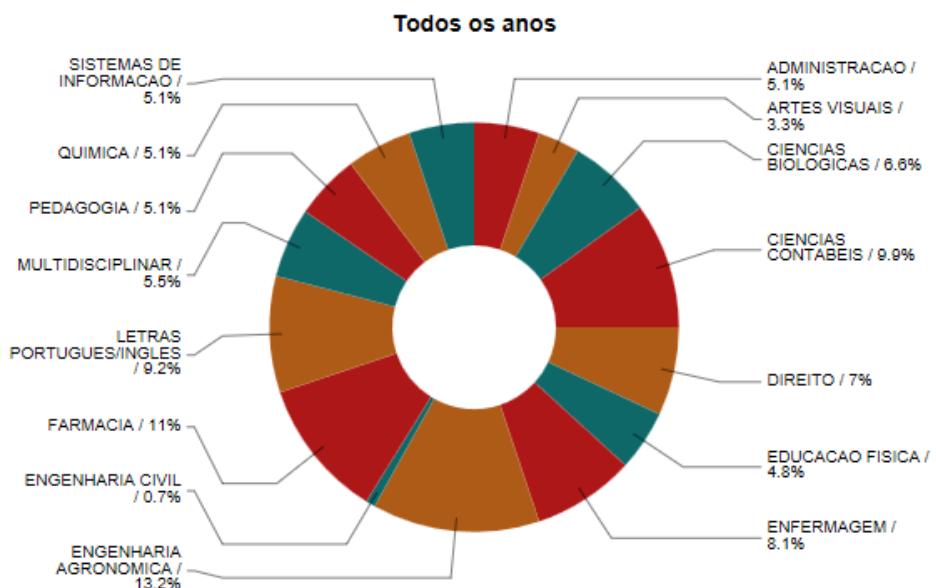


FIGURA 57 - Percentual dos projetos mantido pelo COPE por colegiado.

Dos 15 colegiados que atuam no IFPR campus Palmas, apenas 5 possuem registro de pelo menos 1 projeto de ensino nos últimos 3 anos, são eles: bacharelado em Ciências Contábeis e Farmácia; e licenciatura em Educação Física, Letras Português/Inglês e Pedagogia. Nenhum colegiado possui registro de projetos de inovação nos últimos três anos.

Dos projetos submetidos ao COPE, mais da metade se tratam de projetos de pesquisa, 35,7% são referentes aos projetos de extensão e apenas 2,1% se tratam de projetos de ensino (FIGURA 58).

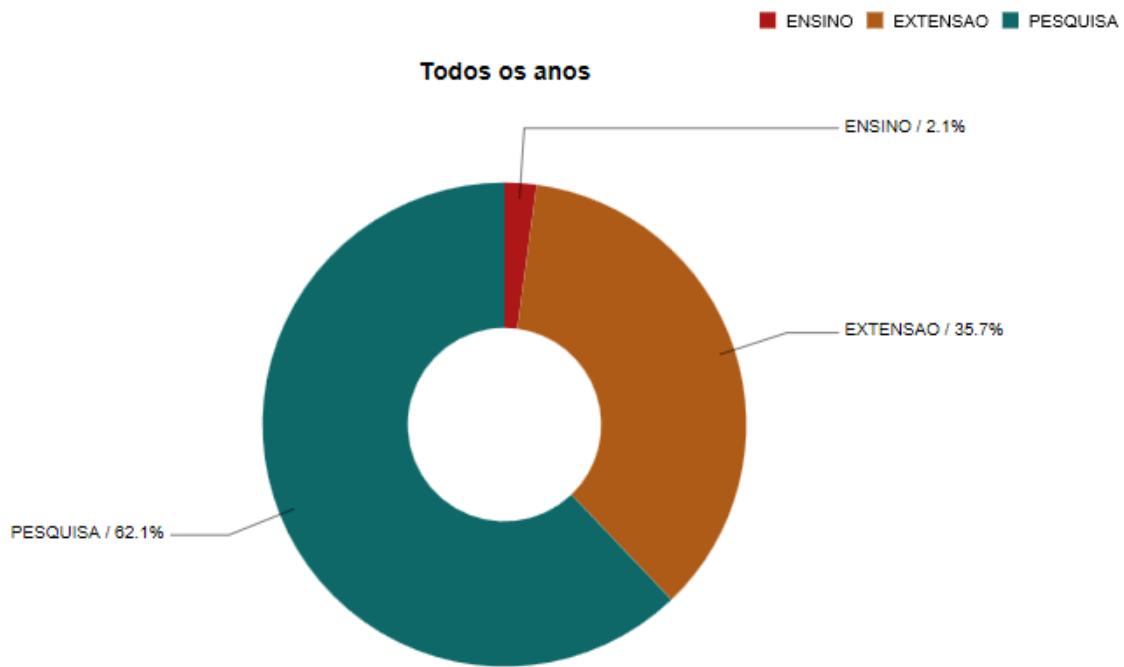


FIGURA 58 - Gráfico dos Projetos mantidos pelo COPE nos últimos 3 anos, por tipo.

A tabela 8 apresenta os dados da análise estatística contendo a quantidade mínima e máxima, a média, variância e desvio padrão dos projetos submetidos ao COPE de janeiro de 2014 à dezembro de 2016.

TABELA 8 - Resultados da estatística descritiva dos projetos submetidos ao COPE.

2014		2015		2016	
MÍNIMO	0	MÍNIMO	1	MÍNIMO	1
MÁXIMO	10	MÁXIMO	14	MÁXIMO	13
MÉDIA	4,87	MÉDIA	6,33	MÉDIA	6,87
VARIÂNCIA	26,44	VARIÂNCIA	48,12	VARIÂNCIA	48,3
DESVIO PADRÃO	5,14	DESVIO PADRÃO	6,94	DESVIO PADRÃO	6,95

Através da tabela 8 é possível perceber um aumento na média de projetos submetidos ao COPE entre os anos de 2014 e 2015. Um dos motivos que pode justificar esse aumento é a chegada de 45 novos docentes na composição do quadro de servidores do campus. De 2015 a 2016 não houveram muitas alterações na quantidade do corpo docente da instituição, por isso os valores mantiveram a mesma média.

Outra informação que pode ser tirada através dos dados apresentados é que dos 235 projetos mantidos pelo COPE, 36 contam com pesquisadores alocados em colegiados diferentes, por isso a somatória dos dados da tabela 7 é igual a 271 projetos.

6. CONSIDERAÇÕES FINAIS

Este trabalho apresentou o perfil da produção científica do IFPR campus Palmas, analisada no período de 2014 a 2016, através do desenvolvimento de um BI que identificou que apesar de existirem pesquisadores com mais de 30 produções no ano a grande maioria dos pesquisadores mantém uma média de 1 a 5 produções ao ano.

Ao analisar as publicações em anais de eventos, foi identificado um número maior de artigos completos em relação aos resumos e resumos expandidos. Quando analisado as produções por colegiado, os pesquisadores do colegiado do curso de licenciatura em Química possuem a maior média dos últimos três anos, porém, essa média não foi suficiente para que a área de conhecimento de exatas e da terra, a qual encontra-se colegiado, liderasse o *ranking*. Ao analisar o coletivo, a área de conhecimento de Ciências da Saúde é quem detém a maior média dos últimos 3 anos com 107 produções/ano.

Também foram analisados os projetos submetidos ao Comitê de Pesquisa e Extensão do IFPR campus Palmas, onde o colegiado com maior número de submissões de projetos ao comitê, nos últimos três anos foi o colegiado do curso bacharelado em Engenharia Agronômica, responsável por 13,2% dos projetos submetidos ao COPE nos últimos 3 anos, com uma média de 12 projetos/ano. De todos os 235 projetos submetidos ao COPE nos últimos 3 anos, destacou-se a inexistência de projetos de inovação e o grande percentual de projetos de pesquisa, somando 62,1% dos projetos.

Além do BI de produção científica, como resultados deste trabalho, cita-se duas publicações em eventos, sendo um artigo completo (APÊNDICE II) e um Pôster (APÊNDICE III) e a elaboração de uma oficina técnica, intitulada “Integrando Dados Através da Ferramenta Gratuita *Pentaho Data Integration*” submetida ao VI SE²PIN – Seminário de Extensão, Ensino, Pesquisa e Inovação do IFPR (APÊNDICE IV).

Durante a realização deste trabalho alguns pontos podem ser destacados. Ao planejar o sistema, uma das preferências era pela utilização de ferramentas que respeitassem as liberdades determinadas pela *Free Software Foundation* e que suas licenças fossem distribuídas sem custo financeiro.

Diante disso, a ferramenta proposta por Mena-Chalco e Junior (2013), nomeada *scriptLattes*, foi a primeira opção de escolha para realizar a extração de informações dos Currículos Lattes disponibilizados na Plataforma Lattes, mantida pelo CNPq. Porém, a partir do ano de 2013, foi implementado o uso de um código de segurança denominado *captcha*,

gerado automaticamente para acessar os Currículos Lattes (FIGURA 59), desta forma, o *scriptLattes* deixou de operar por um determinado tempo.



FIGURA 59 - Código de segurança solicitado para acesso aos Currículos cadastrados na Plataforma Lattes

Atualmente, conforme informações no site do projeto (SCRIPTLATTES, 2017), o *scriptLattes* foi reescrito em uma nova linguagem de programação e voltou a funcionar de forma parcial. Apesar da ferramenta ser capaz de extrair as informações dos currículos da Plataforma Lattes, o processo não é totalmente automatizado. Os usuários devem informar manualmente a identificação de todos os currículos dos quais desejam baixar as informações e a partir daí a ferramenta faz a busca no site do CNPq, porém o usuário deve ainda digitar o código de segurança um a um até finalizar a lista dos currículos desejados. Devido a quantidade de professores relacionados para essa pesquisa, 136 (cento e trinta e seis) no total, e ao tempo que esta ação tomaria, o uso do *scriptLattes* foi descartado.

Através de novas buscas sobre o tema e pesquisas sobre trabalhos similares a este, uma segunda opção foi a utilização do *framework* proposto por Dias (2016). Ao entrar em contato por e-mail com o autor a fim de obter mais informações sobre a disponibilidade da ferramenta, descobriu-se que o *framework* proposto é fruto de uma tese de doutorado e que até a presente data ainda não havia sido defendido e disponibilizado para uso geral.

Por esses motivos, a fim de garantir o cumprimento do cronograma e não comprometer a realização do trabalho, a extração de informações dos Currículos Lattes dos pesquisadores do campus foi realizada através da ferramenta *Stella Experta*, utilizada internamente na instituição e gentilmente disponibilizada para uso neste trabalho.

Outra observação importante é com relação a forma como os projetos são mantidos pelo Comitê de Pesquisa e Extensão do campus. Apesar de todas as informações cadastradas como nome do coordenador e colaboradores dos projetos, título, ano, entre outras informações,

não há nenhuma descrição que aponte quais são as produções resultantes de cada projeto. Desta forma, não foi possível relacionar as publicações extraídas dos currículos Lattes com os projetos mantidos pelo COPE e analisar de maneira efetiva a produção científica do IFPR campus Palmas ligada aos projetos do campus, número de alunos orientados na iniciação científica e impacto na comunidade por exemplo.

Com relação a criação e publicação dos cubos OLAP, apesar da ferramenta *Schema Workbench* não exigir conhecimentos em XML e realizar todo o processo através de sua interface gráfica, é preciso bastante atenção ao mapear as dimensões, tabelas e medidas. Ao mapear a medida que realiza a contagem dos dados, utilizou-se o mesmo nome “quantidade” nos cubos Produção Científica e Projetos. Apesar das medidas serem específicas em cada cubo, apenas o cubo Produção Científica possuía a apresentação dos dados. Somente depois de ter sua medida renomeada é que o cubo Projetos passou a apresentar resultados. Não foi encontrado nenhuma justificativa para este fato nos fóruns e/ou documentação online, porém, alerta-se que ao referenciar duas medidas com o mesmo nome pode gerar complicações na publicação dos cubos, assim como o uso de acentuação no esquema.

Como trabalhos futuros, pretende-se ampliar a análise estatística sobre os dados da produção científica do campus e a partir da aplicação de técnicas de mineração de dados, identificar padrões, tendências e correlações entre os mesmos e automatizar o processo de ETL para realizar novas cargas de dados no DW de maneira automatizada a cada semestre.

REFERÊNCIAS BIBLIOGRÁFICAS

BOUMAN, R.; DONGEN, J. V., **Pentaho Soluções: business intelligence e data armazenamento com Pentaho e MySQL**. Indiana: Wiley Publishing, 2009.

BRASIL, Lei nº 7.596, de 10 de abril de 1987; Decreto Federal 94.664, de 23/07/87.

BRASIL, Lei 11.784, de 22 de setembro de 2008.

BRASIL, Lei nº 11.892, de 29 de dezembro de 2008.

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, **Tabela de Áreas de Conhecimento/Avaliação**. Disponível em: <<http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>>. Acesso em: 05 out. 2017.

CECI, F. **Business Intelligence**: livro digital. Palhoça: UnisulVirtual, 2012. Disponível em: <https://www.researchgate.net/profile/Flavio_Ceci/publication/232882251_Business_Intelligence/links/58d278d1458515b8d287094c/Business-Intelligence.pdf>. Acesso em: 11 mai. 2017.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. Disponível em: <<http://lattes.cnpq.br>>. Acessado em: 02 jun. 2016.

COLAÇO JUNIOR, M. **Projetando sistemas de apoio à decisão baseados em data warehouse**. Rio de Janeiro: Axcel Books, 2004.

COMITE DE PESQUISA E EXTENSÃO – CAMPUS PALMAS, **Regimento Interno do Comitê de Pesquisa e Extensão do IFPR COPE – campus Palmas**, 30 de agosto de 2016. Disponível em: <<http://palmas.ifpr.edu.br/wp-content/uploads/2012/07/Regimento-COPE-Campus-Palmas.pdf>>. Acesso em 02 abr. 2017.

DIAS, D. A. da S. **A Gestão do Conhecimento e o Uso da Tecnologia da Informação em uma Empresa de Engenharia**: um estudo de caso. 94 f. Dissertação (Mestrado em Administração), Universidade FUMEC/FACE, Belo Horizonte, 2012.

DIAS, T. M. R.; et al. Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais**. São Paulo: USP, 2016. Disponível em: <

<https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUEwiXvrOqqa3XAhXDvZAKHWumB4EQFggmMAA&url=http%3A%2F%2Fwww.ebbc.inf.br%2Febbc5%2Findex.php%2Fmain%2Fdownload%2F90&usg=AOvVaw2rxzIxYxCDoNREA9Yq-Pg>. Acesso em: 15 mar. 2017.

GRUPO DE USUÁRIOS PENTAHO BRASIL, **Descrição do Grupo**. Disponível em: <<https://br.groups.yahoo.com/neo/groups/pentahobr/info>>. Acesso em 12 jun. 2017.

HEINZLE, R. **Um modelo de engenharia de conhecimento para sistemas de apoio à decisão com recursos para raciocínio abdutivo**. 251 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento), Universidade Federal de Santa Catarina, Florianópolis, 2010.

INMON, W. H. **Como construir o data warehouse**. Rio de Janeiro: Campus, 1997.

INSTITUTO FEDERAL DO PARANÁ, Portaria 3, de 9 de junho de 2009.

KIMBALL, R. A Dimensional Modeling Manifesto. Kimball Group, 1997. Disponível em: <<http://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto>>. Acesso em: 03 abr. 2017.

KIMBALL, R., ROSS, M. **The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses**. John Wiley & Sons, New York, 1998.

LOH, S. **BI na era do big data para cientistas de dados**: indo além de cubos e dashboards na busca pelos porquês, explicações e padrões. Porto Alegre, 2014.

MACHADO, F. N. R. **Tecnologia e Projeto de Data Warehouse**: uma visão multidimensional – 6ª edição. São Paulo: Érica, 2013.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. In: HAYASHI, M. C. P. I; LETA, L. (Orgs). **Bibliometria e Cientometria**: reflexões teóricas e interfaces. São Carlos: Pedro & João Editores, 2013. p. 109-128.

MONDRIAN. **Documentação Pentaho Mondrian**. Disponível em: <<http://mondrian.pentaho.com/documentation>>. Acesso 09 jun. 2017.

MORAES FILHO, M. S. et al, Desenvolvimento de uma Ferramenta de Business Intelligence para o Ambiente de Ensino a Distância Amadeus: Um Estudo de Caso. **Revista Eletrônica da Estácio Recife**, v.1, n.1. 2015. Disponível em: <<https://reer.emnuvens.com.br/reer/article/view/8/6>>. Acesso em 07 mar. 2017.

MUNARETTO, L. F et al, Produção Científica, Cultura, Artística e Tecnológica: O Caso dos Docentes da UFSM Campus de Frederico Westphalen. **Revista Camine: Caminhos da Educação**, Franca, v. 9, n. 1, p. 91-107, 2017. ISSN 2175-4217.

PENTAHO. Disponível em: <<https://pentaho.com>>. Acesso em 02 jun. 2017.

PENTAHO BRASIL, Pentaho Day. Disponível em: <<http://www.pentahobrasil.com.br/eventos/pentahoday2017>>. Acessado em 02 jun. 2017.

PENTAHO COMMUNITY. Disponível em: <<http://community.pentaho.com>>. Acessado em 02 jun. 2017.

POSTGRESQL. Disponível em: <<https://www.postgresql.org/about>>. Acesso em 02 jun. 2017.

SANTANA, G. A et al, INDICADORES CIENTÍFICOS: Uma Análise da Produção do Programa de Pós-Graduação em Sociologia (PPGS) da UFPE a partir dos currículos da Plataforma Lattes (PL). In: XIV ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, CIÉNCIA DA INFORMAÇÃO E GESTÃO DA INFORMAÇÃO, 14., 2011. **Anais**. Maranhão: Universidade Federal do Maranhão, 2011. Disponível em: <[http://rabc.i.org/rabc.i/sites/default/files/INDICADORES%20CIENT%C3%88DFICOS%20Um a%20An%C3%A1lise%20da%20Produ%C3%A7%C3%A3o%20do%20Programa%20de%20P%C3%B3s-Gradua%C3%A7%C3%A3o%20em%20Sociologia%20\(PPGS\)%20da%20UFPE%20a%20parir%20dos%20curr%C3%A1culos%20da%20Plataforma%20Lattes%20\(PL\).pdf](http://rabc.i.org/rabc.i/sites/default/files/INDICADORES%20CIENT%C3%88DFICOS%20Um a%20An%C3%A1lise%20da%20Produ%C3%A7%C3%A3o%20do%20Programa%20de%20P%C3%B3s-Gradua%C3%A7%C3%A3o%20em%20Sociologia%20(PPGS)%20da%20UFPE%20a%20parir%20dos%20curr%C3%A1culos%20da%20Plataforma%20Lattes%20(PL).pdf)>. Acesso em: 15 mar. 2017.

SCHNEIDER, K., **Experience and Knowledge Management in Software Engineering**. Heidelberg: Springer, 2009.

SEZÕES, C., OLIVEIRA, J., BAPTISTA, M. **Business Intelligence**. Porto: Sociedade Portuguesa de Inovação, 2006.

SILBERSCHATZ, A., KORTH, H. R., SUDARSHAN, S.; tradução Daniel Vieira. **Sistema de Banco de Dados**. 5^a edição. Rio de Janeiro: Elsevier, 2006.

STELA EXPERTA. Disponível em: <<http://site.stelaexperta.com.br>>. Acesso em: 02 jun. 2017.

SOURCEFORCE, **Mondrian**. Disponível em: <<https://sourceforge.net/projects/mondrian/files>>. Acesso em: 20 out. 2017.

TAVARES, R. **Proposta de um sistema de apoio a decisão para controle e gerenciamento agrícola em usinas de açúcar e álcool**. 166 f. Dissertação (Mestrado em Engenharia de Produção) – Escola de Engenharia, Universidade de São Paulo, São Carlos, 2008.

TEOREY, T., LIGHTSTONE, S., NADEAU, T. **Projeto e modelagem de banco de dados**. Rio de Janeiro: Elsevier, 2007.

TURBAN, E. **Business intelligence**: um enfoque gerencial para a inteligência do negócio. São Paulo: Bookman, 2009

APENDICES

APENDICE I - XML do esquema dos cubos OLAP, gerado através da ferramenta *Schema Workbench*.

```
<Schema name="PC IFPR Palmas">
  <Dimension type="TimeDimension" visible="true" highCardinality="false" name="Tempo">
    <Hierarchy name="Ano" visible="true" hasAll="true" allMemberName="Todos os anos" primaryKey="id_tempo">
      <Table name="dim_tempo" schema="public">
        </Table>
      <Level name="Ano" visible="true" column="ano" type="String" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Pesquisador">
      <Hierarchy name="Pesquisador" visible="true" hasAll="true" allMemberName="Todos Pesquisadores" primaryKey="id_pesquisador">
        <Table name="dim_pesquisador" schema="public">
          </Table>
        <Level name="Pesquisador" visible="true" column="nome_pesquisador" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
        <Level name="Formacao" visible="true" column="formacao_pesquisador" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
        </Hierarchy>
      </Dimension>
      <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Colegiado">
        <Hierarchy name="Colegiado" visible="true" hasAll="true" allMemberName="Todos Colegiados" primaryKey="id_colegiado">
          <Table name="dim_colegiado" schema="public">
            </Table>
          <Level name="Colegiado" visible="true" column="nome_colegiado" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
            </Level>
          </Hierarchy>
        </Dimension>
        <Dimension type="StandardDimension" visible="true" highCardinality="false" name="Area_de_Conhecimento">
```

```

<Hierarchy name="Area de Conhecimento" visible="true" hasAll="true"
allMemberName="Todas as Areas" primaryKey="id_area">
  <Table name="dim_area_de_conhecimento" schema="public">
    </Table>
  <Level name="Area de Conhecimento" visible="true" column="nome_area"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
<Dimension type="StandardDimension" visible="true" highCardinality="false"
name="Sub Area de Conhecimento">
  <Hierarchy name="Sub Area de Conhecimento" visible="true" hasAll="true"
allMemberName="Todas as Subareas" primaryKey="id_subarea">
    <Table name="dim_subarea" schema="public">
      </Table>
    <Level name="Subarea de Conhecimento" visible="true"
column="nome_subarea" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
<Cube name="Producao Cientifica" visible="true" cache="true" enabled="true">
  <Table name="fato_producao" schema="public">
    </Table>
  <DimensionUsage source="Tempo" name="Tempo" visible="true"
foreignKey="id_tempo" highCardinality="false">
    <Dimension type="StandardDimension" visible="true"
foreignKey="id_publicacao" highCardinality="false" name="Tipo Publicacao">
      <Hierarchy name="Tipo Publicacao" visible="true" hasAll="true"
allMemberName="Todos as Publicacao" primaryKey="id_publicacao">
        <Table name="dim_publicacao" schema="public">
          </Table>
        <Level name="Tipo Agrupador" visible="true" column="tipo_agrupador"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
          </Level>
        <Level name="Sub Tipo Producao" visible="true" column="subtipo_producao"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
          </Level>
        <Level name="Tipo Producao" visible="true" column="tipo_producao"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
          </Level>
        <Level name="Titulo" visible="true" column="titulo" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
      </Hierarchy>
    </Dimension>
  </DimensionUsage>
</Cube>

```

```

</Dimension>
<DimensionUsage source="Pequisador" name="Pesquisador" visible="true"
foreignKey="id_pesquisador" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Colegiado" name="Colegiado" visible="true"
foreignKey="id_colegiado" highCardinality="false">
</DimensionUsage>
<Dimension type="StandardDimension" visible="true" foreignKey="id_cidade"
highCardinality="false" name="Cidade">
<Hierarchy name="Localizacao" visible="true" hasAll="true"
allMemberName="Todas as Localizacoes" primaryKey="id_cidade">
<Table name="dim_cidade" schema="public">
</Table>
<Level name="Estado" visible="true" column="uf" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
<Level name="Cidade" visible="true" column="nome_cidade" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<DimensionUsage source="Area de Conhecimento" name="Area de
Conhecimento" visible="true" foreignKey="id_area" highCardinality="false">
</DimensionUsage>
<DimensionUsage source="Sub Area de Conhecimento" name="Sub Area de
Conhecimento" visible="true" foreignKey="id_subarea" highCardinality="false">
</DimensionUsage>
<Dimension type="StandardDimension" visible="true" foreignKey="id_evento"
highCardinality="false" name="Evento">
<Hierarchy name="Evento" visible="true" hasAll="true"
allMemberName="Todos os Eventos" primaryKey="id_evento">
<Table name="dim_evento" schema="public">
</Table>
<Level name="Evento" visible="true" column="nome_evento" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
</Level>
</Hierarchy>
</Dimension>
<Measure name="Quantidade Publicacao" column="id_publicacao"
datatype="Integer" aggregator="distinct-count" visible="true">
</Measure>
<CalculatedMember name="Percentual Publicacao por Tipo"
formatString="0.00 %" formula="([Tipo Publicacao].CurrentMember/[Tipo
Publicacao].CurrentMember.Parent)" dimension="Measures" visible="true">
</CalculatedMember>
<CalculatedMember name="Percentual Publicacao por Localizacao"
formatString="0.00 %"
formula="([Cidade].CurrentMember/[Cidade].CurrentMember.Parent)"
dimension="Measures" visible="true">
</CalculatedMember>

```

```

</Cube>
<Cube name="Projetos" visible="true" cache="true" enabled="true">
  <Table name="fato_projeto" schema="public">
    </Table>
    <DimensionUsage source="Tempo" name="Tempo" visible="true"
      foreignKey="id_tempo" highCardinality="false">
    </DimensionUsage>
    <DimensionUsage source="Pequisador" name="Pesquisador" visible="true"
      foreignKey="id_pesquisador" highCardinality="false">
    </DimensionUsage>
    <DimensionUsage source="Colegiado" name="Colegiado" visible="true"
      foreignKey="id_colegiado" highCardinality="false">
    </DimensionUsage>
    <DimensionUsage source="Area de Conhecimento" name="Area de
      Conhecimento" visible="true" foreignKey="id_area" highCardinality="false">
    </DimensionUsage>
    <DimensionUsage source="Sub Area de Conhecimento" name="Sub Area de
      Conhecimento" caption="Sub&#225;rea de Conhecimento" visible="true"
      foreignKey="id_subarea" highCardinality="false">
    </DimensionUsage>
    <Dimension type="StandardDimension" visible="true" foreignKey="id_proj"
      highCardinality="false" name="Tipo Projeto">
      <Hierarchy name="Tipo Projeto" visible="true" hasAll="true"
        allMemberName="Todos os tipos de projeto" primaryKey="id_projeto">
        <Table name="dim_projeto" schema="public">
          </Table>
          <Level name="Tipo Projeto" visible="true" column="tipo" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
          <Level name="Titulo Projeto" visible="true" column="titulo" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
          </Level>
        </Hierarchy>
      </Dimension>
      <Measure name="Quantidade Projeto" column="id_proj" datatype="Integer"
        aggregator="distinct-count" visible="true">
      </Measure>
      <CalculatedMember name="Percentual Tipo Projeto" formatString="0.00 %"
        formula="([Tipo Projeto].CurrentMember/[Tipo Projeto].CurrentMember.Parent)"
        dimension="Measures" visible="true">
      </CalculatedMember>
      <CalculatedMember name="Percentual Projetos por Area" formula="([Area de
        Conhecimento].CurrentMember/[Area de Conhecimento].CurrentMember.Parent)"
        dimension="Measures" visible="true">
      </CalculatedMember>
    </Cube>
  </Schema>

```

APENDICE II - Artigo completo submetido para a VII Contextos e Conceitos – Mostra de Produção Científica do IFPR campus Palmas.

Aplicando Técnicas de Business Intelligence sobre dados da Produção Científica - Um Estudo de caso na Educação

Talita Fatima Paula Monteiro¹, Lauriana Paludo¹

¹Instituto Federal do Paraná – IFPR Campus Palmas

tfpmonteiro@outlook.com, lauriana.paludo@ifpr.edu.br

Resumo: Atualmente as Instituições de Ensino utilizam sistemas de gestão acadêmica e os dados gerados por esses sistemas não tem sido explorados de forma eficaz para a tomada de decisão. A aplicação de técnicas de Business Intelligence (BI) pode auxiliar esse processo. Nesse contexto, o presente trabalho trata da aplicação de técnicas de Business Intelligence sobre dados da produção científica de um campus de um Instituto Federal de Educação do Paraná (IFPR). Trata-se de um estudo de natureza descritiva, com o objetivo de apresentar uma abordagem de modelagem e implementação de um Data Warehouse (DW) para apoio à tomada de decisão. Como resultados obtidos, pode-se citar a revisão da literatura na área de Business Intelligence aplicado em instituições de ensino sobre dados de produção científica, a análise de trabalhos similares desenvolvidos identificando contribuições e o estado da arte na área, a modelagem e a criação de um Data Warehouse de produções científicas, implementado a partir da extração, transformação e carga dos dados do período compreendido entre janeiro de 2014 e dezembro de 2016 e oriundos das bases de dados do Comitê de Pesquisa e Extensão (COPE) da instituição e dos Currículos Lattes dos pesquisadores obtidos através da Plataforma Lattes do CNPq.

Palavras-Chave: Business Intelligence, Data Warehouse, Produção Científica.

1. Introdução

Business Intelligence é o termo utilizado para o processo composto por métodos, métricas, ferramentas e entre outros, que de uma forma genérica recolhem dados, transformam em informações e, sequencialmente, em conhecimento (SANZÕES, OLIVEIRA e BAPTISTA, 2006).

De forma geral, um projeto de BI inicia com o planejamento estratégico da organização identificando todos os seus sistemas e bases de dados e quais tipos de perguntas deseja-se responder com o BI. Em seguida, parte-se para a modelagem do repositório único de dados e informações, chamado de *Data Warehouse*. Com o DW modelado e carregado, é necessário identificar quais as técnicas de análises poderão ser aplicadas para a extração de informações e conhecimentos.

Pode-se dizer que o BI é utilizado para fornecer indicadores por meio de simulações e de cenários com previsões futuras baseado em dados históricos gerados ao longo da existência da organização. Desta forma, a aplicação de técnicas de BI no processo de tomada de decisão estratégica torna-se uma ferramenta importante para transformar estes dados armazenados em informações úteis para as organizações.

Inseridas no mesmo contexto das demais organizações, Instituições de Ensino também necessitam ter recursos para subsidiarem a otimização de seus processos decisórios.

Existem vários trabalhos relacionados a DW aplicados sobre dados educacionais e utilização de indicadores de perfil de produção de seus pesquisadores, como o trabalho de

Santana *et al* (2011) do Programa de Pós-Graduação em Sociologia (PPGS) da Universidade Federal do Pernambuco (UFPE), que através de uma análise qualitativa identificou os principais aspectos e tendências da produção científica do PPGS. Da mesma forma, o conceito e uso de BI foi aplicado através do ambiente open source de Ensino à Distância (EaD) Amadeu, afim de apoiar as decisões estratégicas de uma instituição, possibilitando a simulação da geração de relatórios ad-hoc e dashboards capazes de possibilitar consultas mais rápidas, flexíveis e seguras. (FILHO *et al* 2015).

A instituição utilizada como estudo de caso neste trabalho é um campus de um Instituto Federal de Educação do Paraná. Os Institutos Federais são instituições de educação superior, básica e profissional, pluricurriculares e multicampi, especializados na oferta de educação profissional e tecnológica em diferentes modalidades de ensino. De acordo com a Lei de criação nº 11.892/2008, uma das características e finalidades dessas instituições é realizar e estimular a pesquisa e o desenvolvimento científico tecnológico. A aplicabilidade da pesquisa dentro de Instituições de Ensino traz benefícios a instituição e novas expectativas no desenvolvimento profissional da sociedade, devendo estar presente em todo trajeto e formação do trabalhador.

Para compreender melhor a dinâmica da produção científica, subsidiar e avaliar o planejamento e resultado das políticas voltadas a esse objetivo, instituições e órgãos de fomento à pesquisa incentivam o uso de indicadores para mensurar e guiar os pesquisadores no atingimento de metas previamente delineadas por instituições como a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Nesse contexto, este trabalho apresenta seu foco na aplicabilidade de técnicas de BI sobre os dados de publicações científicas de um campus de um Instituto Federal de Educação escolhido para estudo de caso, apresentando a modelagem e implementação do *Data Warehouse* desenvolvido, de forma que desenvolvedores e interessados na área compreendam quais ferramentas e técnicas de BI foram utilizadas e como o trabalho foi desenvolvido, além de divulgar o mesmo à comunidade *open source*.

O artigo está organizado em mais três seções, onde na seção 2 são descritos os materiais e métodos utilizados para elaboração do mesmo, na seção 3 são apresentados e discutidos os resultados obtidos e na quarta seção, apresentadas as considerações finais.

2. Materiais e Métodos

Para o desenvolvimento deste trabalho optou-se pela utilização de softwares que atendam às quatro liberdades do software livre: executar, modificar, redistribuir e distribuir versões modificadas de um programa, conforme especificado pela *Free Software Foundation* (2017), e que sejam distribuídas de forma gratuita.

Com essa finalidade, a suíte *Pentaho* foi selecionada como plataforma de código aberto para dar suporte ao desenvolvimento do trabalho. A Pentaho é uma empresa fundada em 2014 e integrada ao *Grupo Hitachi* em 2015, líder em integração de dados e análise de negócios, com suporte a diversas implementações de dados que se destaca pela facilidade em poder ser incorporada em qualquer aplicativo de fluxo de trabalho ou modelo de implantação e proporcionar a integração de dados de qualquer fonte através do uso de ferramentas do tipo “arrasta e puxa” (PENTAHO, 2017).

Atualmente, a versão 7.1 das ferramentas Pentaho são distribuídas sobre duas licenças: comercial (*Enterprise*) e comunitária (*Community*). Ambas compostas por oito ferramentas (*Business Analytics Platform*, *Pentaho Data Integration*, *Report Designer*, *Market Place*, *Aggregation Designer*, *Schema Workbench*, *Metadata Editor* e *Hadoop Schims*) que juntas podem atender todo o processo de *Business Intelligence*, desde a extração de dados das mais

diferentes bases até a apresentação de relatórios e gráficos (PENTaho, 2017). A principal diferença entre a versão comercial e comunitária das ferramentas Pentaho está na configuração dos ambientes e no suporte prestado pela empresa àquelas que optam pela licença comercial. Apesar de não contar com o suporte prestado pela empresa, usuários da versão comunitária contam com o apoio de centenas de usuários que participam ativamente em fóruns online e na elaboração de materiais de apoio.

No Brasil, o principal grupo de usuários da comunidade Pentaho conta, atualmente, com 2003 membros que compartilham documentação, recursos e troca de experiência além de promoverem todos os anos o *Pentaho Day*, um evento dedicado à apresentação e troca de experiências sobre desenvolvimento *Open Source* na plataforma Pentaho (GRUPO DE USUÁRIOS PENTAHO BRASIL, 2017), (PENTAHO BRASIL, 2017).

Uma ferramenta da suíte Pentaho é utilizada no desenvolvimento deste trabalho, a *Pentaho Data Integration* (PDI) responsável por realizar a Extração, Transformação e Carga de dados (ETL, do inglês *Extraction, Transformation and Load*).

A metodologia utilizada para o desenvolvimento do trabalho segue três etapas principais (i) extraer dados da base de dados do comitê de pesquisa e extensão do campus e do currículo Lattes dos pesquisadores do campus, (ii) tratar e integrar os dados na área de armazenamento temporário chamada de ODS (*Operational Data Storage*) e (iii) criar um *data warehouse* para persistir os dados de produção científica oriundos das etapas anteriores.

Na Figura 1 são mostrados os componentes da arquitetura da aplicação desenvolvida neste trabalho.

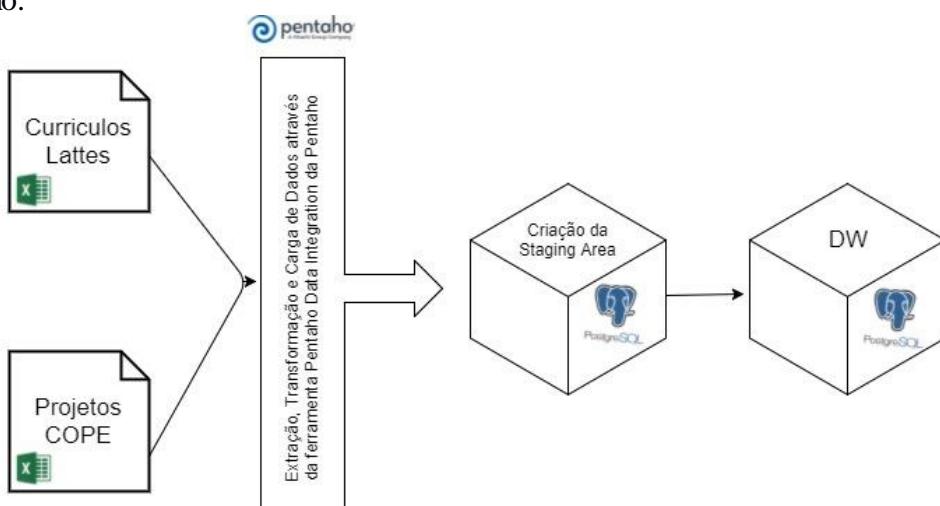


Figura 1 – Arquitetura do DW para o BI do perfil da produção científica.

A extração dos dados da base do COPE foi feita diretamente com a ferramenta PDI. Já a extração dos dados dos Currículos Lattes dos professores lotados no campus foi realizada de forma automatizada pela ferramenta Stela Expert e posteriormente com a ferramenta PDI. Os dados foram tratados, integrados e carregados em uma área intermediária, também conhecida como *Staging Area* ou ODS (do inglês *Operational Data Storage*) com a ferramenta PDI. A ODS é um armazenamento intermediário dos dados com o objetivo de facilitar a integração dos mesmos no ambiente operativo antes da persistência no DW. Sua função é agilizar o processo de consolidação resultando em um melhor desempenho na fase da atualização dos dados (MACHADO, 2013). A ODS, bem como o DW foram implementados com o sistema de banco de dados objeto-relacional *PostgreSQL*.

O *PostgreSQL* foi desenvolvido para todos os principais sistemas operacionais com recursos sofisticados como controle de concorrência de multi-versão (MVCC) e recuperação de ponto-a-tempo. Algumas de suas características são a oferta de suporte pela comunidade;

programas de treinamentos mais econômicos, gerenciáveis e práticos quando comparados com outros bancos de dados proprietários; confiabilidade; disponibilização do código-fonte de forma gratuita, tornando-o extensível; ferramentas de design e administração de banco de dados GUI; entre outros (POSTGRESQL, 2017).

3. Resultados e Discussões

O processo de ETL deste trabalho demorou aproximadamente dois meses para ser concluído. Gerou um esforço muito grande de horas de trabalho devido à diversidade existente em termos de estruturas de dados nas bases dos sistemas utilizados. Os dados primeiramente foram integrados na área intermediária de armazenamento e após a integração e limpeza, foram exportados para o DW.

Para a implementação do DW seguiu-se a abordagem de Machado (2013), realizando duas modelagens: (i) a Entidade-Relacionamento (ER), tradicionalmente conhecida entre os desenvolvedores por ser uma ferramenta que ajuda a análise de requisito de negócio e a essência técnica da arquitetura de DW; (ii) a Modelagem Multidimensional, que é uma técnica de concepção e visualização de um modelo de dados de um conjunto de medidas que descrevem aspectos comuns de negócio.

A Figura 2 abaixo apresenta o modelo ER, representado pelo diagrama ER composto pelas representações de dados: entidade, relacionamentos e atributos. Assim, é possível identificar como acontecerá o relacionamento entre as classes que compõem este trabalho.

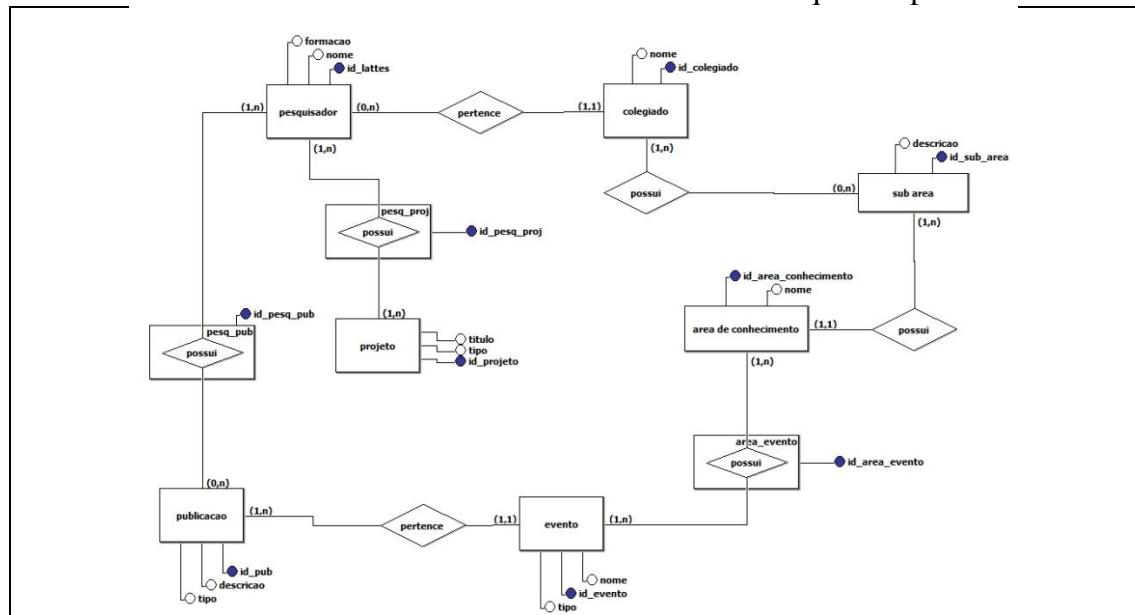


Figura 1 – Modelo ER do DW de produção científica.

Fonte: As autoras.

Neste modelo ER, cada pesquisador está ligado a um colegiado, onde cada colegiado possui uma sub-area de conhecimento, que faz parte da divisão determinada pela CAPES das grandes áreas de conhecimento, podendo ser ciências exatas e da terra, ciências biológicas, entre outras. Além de estarem relacionados com os colegiados, os pesquisadores também possuem projetos – que são mantidos pelo COPE e executados durante parte da carga horária cumprida na instituição, e publicações. As publicações são registradas no currículo Lattes de cada pesquisador e relacionada com o evento onde aconteceu a publicação.

Diferente do ER, o modelo dimensional é composto por uma tabela com várias chaves, chamada de tabela fato que representa o objeto de consulta, e um conjunto de tabelas menores

nomeadas tabelas de dimensão, que determinam a consultas relacionada ao fato. A relação entre a tabela fato e as tabelas de dimensão acontece através de chaves. A chave primária existente em cada tabela de dimensão corresponde exatamente à uma das chaves estrangeiras que compõe a tabela fato. Desta forma, a ligação entre essas tabelas resulta em uma forma parecida com uma estrela, por esse motivo, esse modelo é também conhecido como Modelo Estrela (do inglês *Star Schema*) (KIMBALL, 1998).

A Figura 3 apresenta o modelo dimensional do DW implementado. Neste caso, o fato a ser consultado é a produção científica, representado pela tabela nomeada fato_producao_cientifica, todas as demais tabelas são tabelas de dimensão. Deste modo, será possível consultar a produção científica com cada dimensão, como por exemplo, identificar as produções científicas de uma determinada área de conhecimento no ano de 2015, através das tabelas de dimensões tempo (dim_tempo) e área de conhecimento (dim_area_de_conhecimento); ou ainda qual a média da produção científica de um determinado pesquisador durante os anos de 2014 à 2016, através das dimensões pesquisador (dim_pesquisador) e tempo; e muitas outras consultas utilizando as dimensões propostas em conjunto com as medidas e métricas da tabela fato.

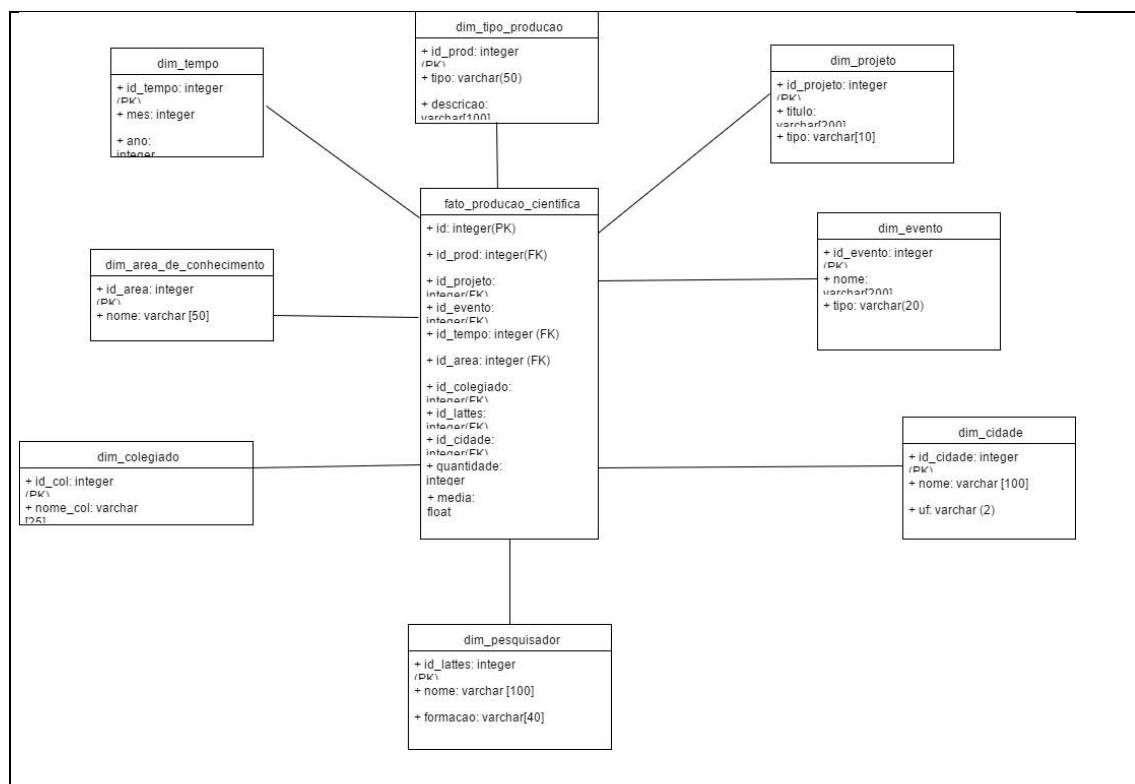


Figura 3 – Modelagem Dimensional do DW através do esquema estrela.
Fonte: As autoras.

A criação das tabelas de dimensão área de conhecimento e sub-áreas de conhecimento foi realizada através da tabela de Áreas de Conhecimento/Avaliação disponibilizada nos formatos PDF e DOC pela CAPES (2017). Após o *download* do arquivo em formato PDF, as informações foram reorganizadas em uma planilha eletrônica e separadas em abas onde a primeira é composta pelo código e descrição das nove grandes áreas de conhecimento e as demais são compostas pelas sub áreas de cada uma das grandes áreas, conforme mostra a Figura 4.

The figure consists of two side-by-side screenshots of Microsoft Excel. The left screenshot shows a table with columns A through E. The first column contains codes (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) and descriptions (e.g., cod capes, area, Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas, MULTIDISCIPLINAR). The right screenshot shows a similar table with columns A through D, listing sub-areas such as Matemática, Probabilidade e Estatística, Ciência da Computação, Astronomia, Física, Química, and GeoCiências.

Figura 4 – Áreas e Sub áreas do conhecimento de acordo com a CAPES

Fonte: As autoras.

A integração dos dados da planilha contendo as áreas e sub-áreas de conhecimento da CAPES para base de dados *PostgreSQL* aconteceu através da ferramenta PDI, conforme ilustrado nas Figuras 5 e 6. O primeiro passo foi realizar a extração e tratamento dos dados da planilha antes de carregá-las na base de dados. Desta forma, a acentuação das descrições foi removida para não gerar erros de consultas, todas as letras foram convertidas para caixa alta (maiúsculas) e a coluna com o código capes foi removida, uma vez que as consultas acontecerão através das descrições ou chaves primárias.

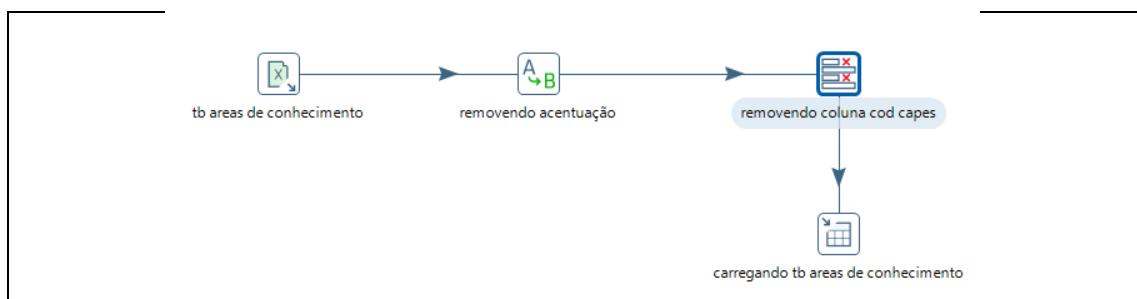


Figura 5 – Extração e limpeza dos dados e criação da tabela de áreas de conhecimento.

Fonte: As autoras.

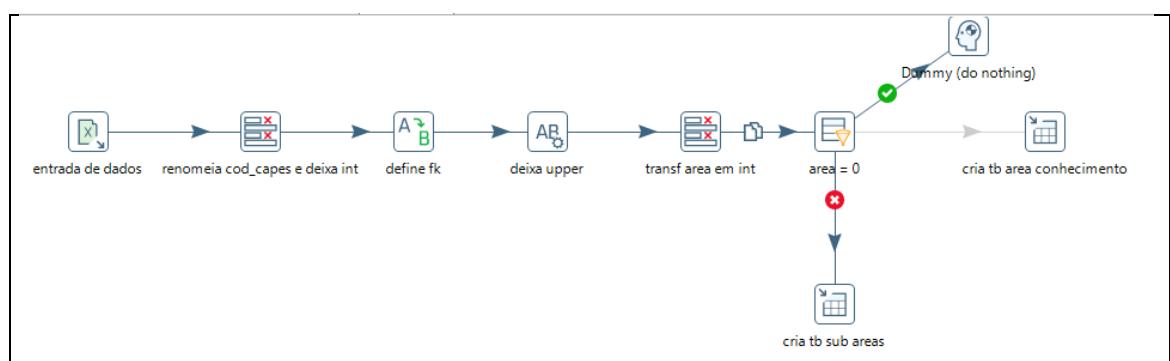


Figura 6 – Extração e limpeza dos dados e criação da tabela sub-área de conhecimento.

Fonte: As autoras.

Com as tabelas área de conhecimento e sub-áreas carregadas, foi possível criar a tabela colegiado contendo todos os colegiados que atuam no campus nos cursos de graduação ofertados. Por se tratar de uma quantidade pequena de informações, 15 (quinze) no total, a criação desta tabela se deu diretamente na base de dados com o uso da linguagem SQL, conforme mostra a Figura 7.

```
bi_cope on postgres@PostgreSQL 9.6
1 CREATE TABLE public.tb_colegiado
2 (
3     id_col bigint NOT NULL ,
4     descricao character varying(40)NOT NULL,
5     sub_area bigint NOT NULL,
6     CONSTRAINT pk_tb_colegiado PRIMARY KEY (id_col),
7     CONSTRAINT tb_colegiado_sub_area_fkey FOREIGN KEY (sub_area)
8 )
9
10 insert into tb_colegiado (descricao, sub_area) values ('ADMINISTRAÇÃO',51), ('ARTES VISUAIS', 74),
11 ('CIÉNCIAS BIOLÓGICAS', 8), ('CIÉNCIAS CONTÁBEIS', 51), ('DIREITO', 50), ('EDUCAÇÃO FÍSICA', 42),
12 ('ENFERMAGEM', 37), ('ENGENHARIA AGRONÔMICA', 43), ('ENGENHARIA CIVIL', 21), ('FARMÁCIA', 36),
13 ('LETRAS PORTUGUÊS/INGLÊS', 73), ('PEDAGOGIA', 69), ('QUÍMICA',6), ('SISTEMAS DE INFORMAÇÃO', 3),
14 ('MULTIDISCIPLINAR', 75)
```

Figura 7 – SQL da tabela colegiado.

Fonte: As autoras.

O próximo passo foi realizar a carga da tabela com informações dos docentes do campus. A relação dos docentes que atuam no campus foi consultada e extraída da plataforma Stela Expertá através de uma busca de Pessoas, utilizando os parâmetros Lotação Institucional, Enquadramento Funcional (Professor do Ensino Básico, Técnico e Tecnológico), Regime de Trabalho (Docente – Dedicação Exclusiva; Docente – 40 horas semanais; Docente – 20 horas semanais) e Vínculo Institucional (Ativo Permanente). Desta forma, obteve-se a relação de todos os professores que atuam no campus Palmas do IFPR enquadrados no regime de trabalho de 40 horas semanais, 20 horas semanais e Dedicação Exclusiva que estejam com o vínculo institucional ativo. O resultado desta busca foi uma planilha eletrônica com 132 registros com as seguintes informações de cada docente: Nome, Endereço do Currículo Lattes, Área de Titulação, Titulação, Instituição da Titulação, Ano da Titulação, Lotação Institucional, Bolsista PQ/DT (CNPq), Vínculo Institucional, Ano de Ingresso na Instituição, Perfil na Instituição, Regime de Trabalho, Titulação Máxima Informada pela Instituição, Data de Atualização do Currículo e 3 (três) Palavras-Chave. A Figura 8 apresenta o resultado da busca na Plataforma Stela Expertá.

A	B	C	D	E	F	G
	Nome	Currículo Lattes	Área de titulação máxima informada no CV-Lattes	Titulação máxima informada no CV-Lattes	Instituição de titulação máxima	Localização institucional
1	Marlei Casturro Mendes Sandi	http://lattes.cnpq.br/9830481482341951		Doutorado	UEM	2016
2	Ricardo Aparecido Pereira	http://lattes.cnpq.br/2122103679450333	Farmácia	Mestrado acadêmico	UNICENTRO	2014
3	Jacob dos Santos Belal	http://lattes.cnpq.br/3325019888451522	Lerias	Doutorado	UNESP	2015
4	Janene Lopes de Souza	http://lattes.cnpq.br/805025470625580	Educação	Especialização	UNASSELVI	2013
5	Laúrana Paludo	http://lattes.cnpq.br/039484682768704	Ciência da Computação	Mestrado acadêmico	UNIVALI	2013
6	Fábio Machado Pluzzenreuter	http://lattes.cnpq.br/07042355276748	Lerias	Doutorado	UFSC	2014
7	Rafael Henrique Santin	http://lattes.cnpq.br/6931202542406017	Educação	Mestrado acadêmico	UEM	2012
8	Christiano Santos Rocha Pitra	http://lattes.cnpq.br/75655912678252	Agronomia	Doutorado	UFPR	2012
9	Vagner Scammi	http://lattes.cnpq.br/940405501395575	Ciência da Computação	Mestrado acadêmico	FACCAMP	2017
10	Lilian Moura de Lima Spagnolo	http://lattes.cnpq.br/387100953903284	Entomologia	Doutorado	UFPEL	2016
11	Diego Benetella Foschiera	http://lattes.cnpq.br/938315648473328	Educação	Especialização	FESL	2017
12	Rodrigo Bento de Almeida	http://lattes.cnpq.br/034784418241586	Farmácia	Mestrado acadêmico	UFPRGS	2003
13	Jussara Isabel Stodmanns	http://lattes.cnpq.br/94536310500533	Engenharia de Produção	Mestrado acadêmico	UFSC	2001
14	Carolina Ribeiro Noronha de Souza	http://lattes.cnpq.br/3417349005593028	Biologia Geral	Doutorado	USP	2014
15	Ana Carolina Vilela de Cavalcão	http://lattes.cnpq.br/634761453262739	Administração	Mestrado acadêmico	UEL	2012
16	Lata Mata Galvão	http://lattes.cnpq.br/0197773008858351	Desenho	Doutorado	UnB	2017
17	Douglas Silva do Prado	http://lattes.cnpq.br/253599795475421	Educação	Especialização	FAMPER	2013
18	Aline Rocha Borges	http://lattes.cnpq.br/43397450738442	Química	Doutorado	UFRGS	2013
19						

Figura 8 – Resultado da Busca por Professores através da Plataforma Stela Expert.

Fonte: As autoras.

A persistência das informações contidas na planilha eletrônica apresentada na Figura 8 para a ODS aconteceu através do PDI. Através do PDI, a relação dos professores foi ordenada de forma crescente e a coluna colegiado foi adicionada, assim, foi realizado o relacionamento de cada docente com o colegiado em que atua. Desta forma, as informações salvas na base de dados foram: Nome, ID Lattes, Colegiado, Área de Titulação, Titulação, Regime de Trabalho e Palavra-Chave, conforme é apresentado na Figura 9.

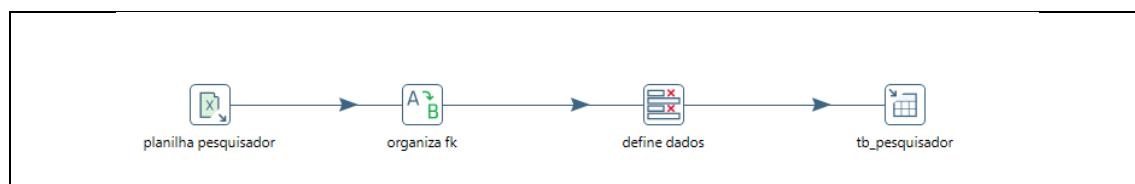


Figura 9 – ETL da tabela professor.

Fonte: As autoras.

De acordo com a modelagem multidimensional, um dos pontos que este trabalho busca identificar é se existem publicações realizadas em outros municípios/estados brasileiros e, a partir daí, apontar quais são eles e com qual frequência essas publicações acontecem. Para isso, foi necessário a criação de uma tabela na base de dados com a relação das cidades brasileiras e seus respectivos estados. Para tanto, utilizou-se a tabela de estados e municípios disponibilizada na internet através do site do IBGE (2017).

Com a planilha eletrônica em mãos, os dados foram organizados em ordem crescente, pelo nome de cada município e carregados para a base através do PDI, conforme mostra a Figura 10.

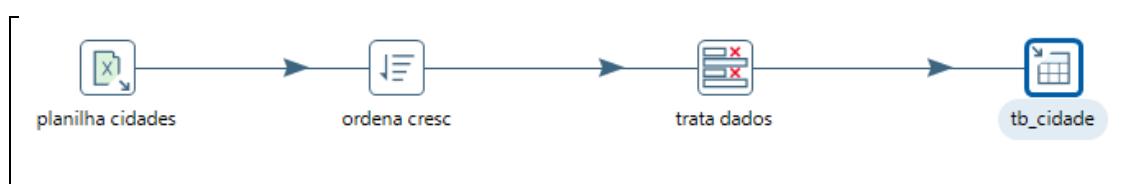


Figura 10 – Extração de dados da tabela do IBGE e persistência dos mesmos na ODS.

Fonte: As autoras.

Os dois últimos processos de ETL acontecem com as tabelas de publicações e projetos, onde:

- As publicações são oriundas do Currículo Lattes de cada docente de acordo com os indicadores de pesquisa proposto pelo CNPq como produção bibliográfica, participação em eventos, produção técnica, produção artística, entre outros. Os dados foram obtidos através da Plataforma Stela Expertá após realizar uma busca por Produções, através dos filtros: Tipo da Produção (Produção Bibliográfica, Evento, Produção Técnica, Outro Tipo de Produção e Produção Artística/Cultural), Ano da Produção, Lotação Institucional do Autor no Período de vinculo institucional do autor. De acordo com essa busca, 1876 agrupamentos foram encontrados referente à produções (bibliográfica, artística e técnicas), entre os anos de 2014, 2015 e 2016, realizadas pelos docentes do IFPR campus Palmas.

- Os projetos são oriundos de uma planilha eletrônica mantida pela instituição, por meio do COPE, contendo informações dos projetos de pesquisa, ensino-aprendizagem, extensão e inovação realizados pelos docentes do campus durante a carga horária de pesquisa que exercem na instituição. Foram extraídos 250 registros desta planilha referente aos projetos cadastrados entre os anos de 2014 a 2016.

O processo de ETL da planilha de publicações e projetos se dividiu em duas fases. Em um primeiro momento foi realizada uma limpeza manual na própria planilha eletrônica, onde os arquivos foram organizados por data e separados por ano (2014, 2015 e 2016). Na planilha de publicações, os nomes dos autores foram corrigidos conforme constam na base de dados, para que o PDI fosse capaz de reconhecer seus pares. Apesar da planilha publicações não trazer informações do local do evento em uma coluna separada, através da coluna ABNT foi possível identificar as cidades onde os trabalhos foram publicados, desta forma criou-se duas colunas separadas para acrescentar os campos Município e UF que foram inseridos manualmente. Na planilha de projetos a padronização dos dados também foi realizada de forma manual ao remover a quebra de linha das colunas, totalizando 250 registros corrigidos.

A segunda fase da transformação, foi importar essas planilhas para o PDI, em transformações diferentes, e realizar a substituição dos nomes dos autores por suas respectivas chaves estrangeiras, bem como o nome do evento e da cidade (no caso da planilha de publicação) e ao final carregar as informações para a base de dados, conforme é apresentado na Figura 11.

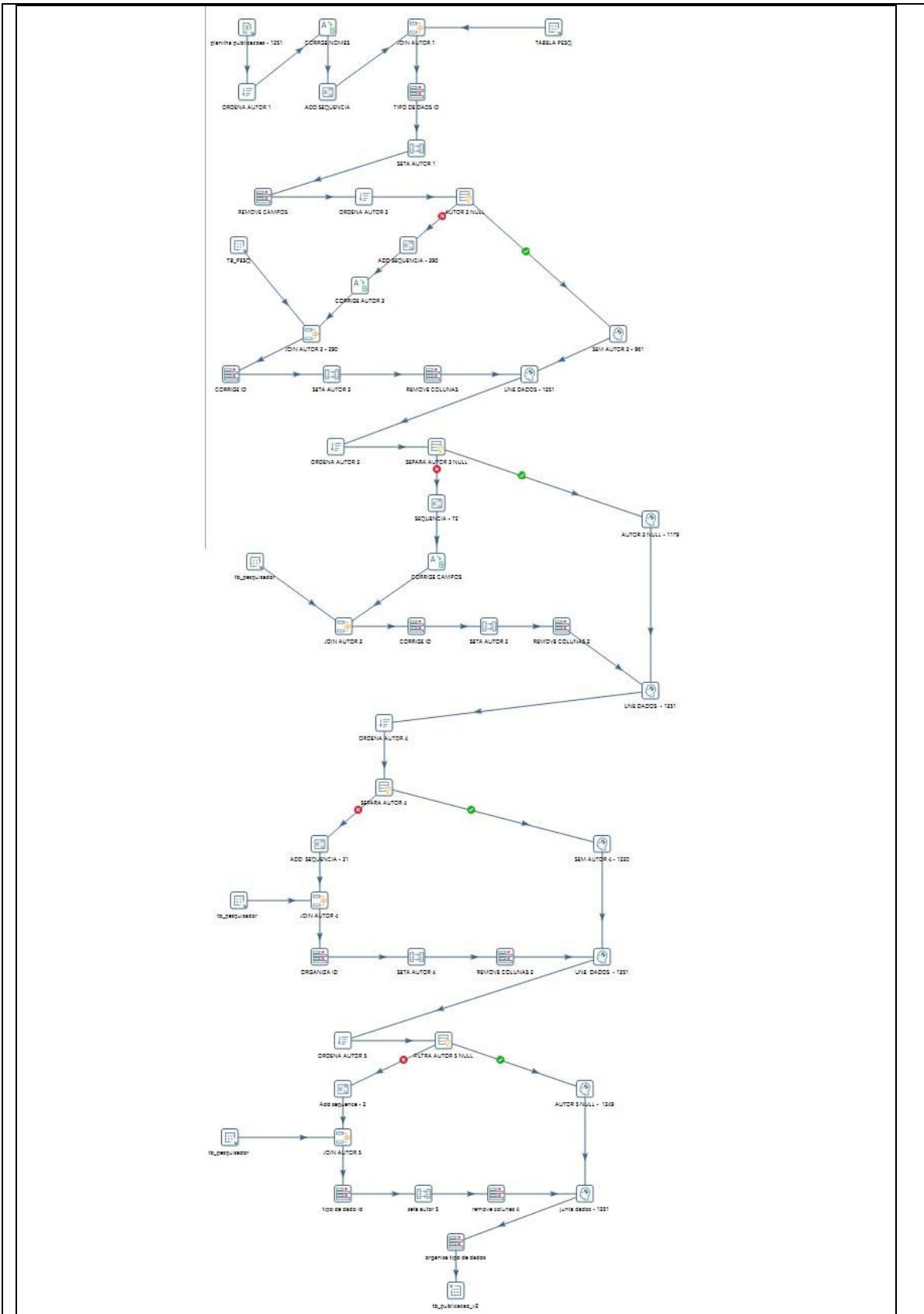


Figura 11 – ETL da tabela publicação.

Fonte: As autoras.

Como resultado das transformações obteve-se a criação da *Staging Area*, conforme o modelo apresentado no diagrama de ER da Figura 2.

Durante a realização desta fase alguns pontos podem ser destacados, além da preferência pela utilização de ferramentas *open source* distribuídas sem custo financeiro. Ao conhecer a ferramenta proposta por Mena-Chalco e Junior (2013), o *scriptLattes*, esta foi a primeira opção para realizar a extração de informações dos Currículos Lattes disponibilizados na Plataforma Lattes, mantida pelo CNPq. Porém, a partir do ano de 2013, a ferramenta ficou inutilizada durante um período devido a utilização do *captcha* (um teste de desafio cognitivo, utilizado como ferramenta *anti-spam*) implementada pela CAPES.

Atualmente, conforme informações no site do projeto (SCRIPTLATTES, 2017), o *scriptLattes* foi reescrito em uma nova linguagem de programação e voltou a funcionar parcialmente. Apesar da ferramenta ser capaz de extrair os currículos da Plataforma, o processo não é totalmente automatizado. Os usuários devem informar a identificação de todos os currículos que desejam baixar as informações e a partir daí a ferramenta faz a busca no site do CNPq, porém o usuário deve digitar a resposta do *captcha* um a um até finalizar a lista dos currículos informados. Devido a quantidade de professores alocados no campus, aproximadamente 150, e ao tempo que esta ação tomaria, esta opção foi descartada.

Através de novas buscas sobre o tema e pesquisas sobre trabalhos similares a este, uma segunda opção foi a utilização do *framework* proposto por Dias (2016). Ao entrar em contato com o autor através de e-mail para obter mais informações sobre a disponibilidade da ferramenta, descobriu-se que o *framework* proposto é fruto de uma tese de doutorado e que até a presente data ainda não havia sido defendido e disponibilizado para uso geral.

Sendo assim, para garantir o cumprimento do cronograma e não comprometer a realização do trabalho, a extração de informações dos Currículos Lattes dos pesquisadores do campus foi realizada através da ferramenta Stella Experta que é utilizada internamente na instituição.

Outra observação importante é com relação a forma como os projetos são mantidos pelo Comitê de Pesquisa e Extensão do campus. Apesar de todas as informações cadastradas como nome do coordenador e colaboradores dos projetos, título, ano, entre outras informações, não há nenhuma descrição que aponte quais são as produções resultantes de cada projeto. Desta forma, não é possível realizar a correlação entre as produções presentes nos Currículos Lattes dos pesquisadores com os resultados dos projetos desenvolvidos dentro da instituição e mantidos pelo COPE.

4. Considerações Finais

Este trabalho apresentou um estudo de caso sobre a aplicação de técnicas de BI sobre dados de publicações científicas de uma instituição de ensino, configurando o desenvolvimento de uma abordagem de *Data Warehouse* de dados educacionais para apoio a tomada de decisão.

O desenvolvimento desta solução mostrou ser um processo bastante complexo devido ao trabalho de extração, transformação e carga dos dados, bem como a integração dos mesmos.

São apresentados como resultados deste trabalho a revisão da literatura na área de BI aplicado em instituições de ensino sobre dados de produção científica, a análise de trabalhos similares desenvolvidos identificando contribuições e o estado da arte na área, a modelagem e a criação do DW de produção científica realizada e o processo de ETL das bases de dados dos Currículos Lattes e do COPE do campus.

Como trabalhos futuros pretende-se desenvolver a camada lógica, não persistida em banco, responsável por manipular e analisar os dados armazenados no DW sob múltiplas perspectivas, o processamento analítico ou cubo OLAP, onde será possível responder questões como: “Qual quantidade de projetos publicados por pesquisadores da área de

ciências exatas no ano de 2016?”, “Qual a média anual de publicação por pesquisadores por colegiado?”, “Quantas publicações teve determinado professor?”, entre outras, identifica ndo o perfil e mapeando quantitativamente a produção científica da instituição.

Referências

BRASIL, Lei nº 11.892, de 29 de dezembro de 2008.

CAPES. Tabela Áreas de Conhecimento/Avaliação. Disponível em: <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>. Último acesso em: 13/08/2017.

DIAS, T. M. R.; et al. Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. Anais... São Paulo: USP, 2016. p. A90

FILHO, M. S. M. *et al*, Desenvolvimento de uma Ferramenta de Business Intelligence para o Ambiente de Ensino a Distância Amadeus: Um Estudo de Caso. Revista Eletrônica da Estácio Recife, v.1, n.1 de 2015.

FREE SOFTWARE FUNDATION. Disponível em: <https://www.gnu.org/philosophy/free-sw.en.html>. Acesso em 13/08/2017.

GRUPO DE USUÁRIOS PENTAHO BRASIL. Disponível em: <https://br.groups.yahoo.com/neo/groups/pentahobr/info>. Acesso em: 06/06/2017.

IBGE. Referências Geográficas Composição por Municípios 2017. Disponível em: http://servicodados.ibge.gov.br/Download/Download.ashx?u=geoftp.ibge.gov.br/organizacao_do_territorio/divisao_regional/divisao_regional_do_brasil/divisao_regional_do_brasil_e_m_regioes_geograficas_2017/tabelas/regioes_geograficas_composicao_por_municipios_2017.xlsx. Acessado em 14/08/2017.

KIMBALL, R. A Dimensional Modeling Manifesto. Kimball Group, 1997. Disponível em: <http://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto> Acesso em 26/04/2017.

MACHADO, F. N. R. Tecnologia e Projeto de Data Warehouse: uma visão multidimensional – 6ª edição. São Paulo: Érica, 2013.

MENA-CHALCO, J. P.; JUNIOR, R. M. C. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. Capítulo de livro: “Bibliometria e Cientometria: reflexões teóricas e interfaces”, p. 109-128. São Carlos: Pedro & João Editores. Maria Cristina Piumbato Innocentini Hayashi e Jacqueline Leta (Orgs.), 2013.

POSTGRESQL. Disponível em: <https://www.postgresql.org/about> Acesso em: 10/06/2017.

PENTAHO. Disponível em: <https://pentaho.com>. Acesso em 04/06/2017.

PENTAHO	BRASIL.	Disponível em http://www.pentahobrasil.com.br/eventos/pentahoday2017 .
<p>SANTANA, G. A <i>et al</i>, INDICADORES CIENTÍFICOS: Uma Análise da Produção do Programa de Pós-Graduação em Sociologia (PPGS) da UFPE a partir dos currículos da Plataforma Lattes (PL). Universidade Federal do Maranhão: XIV Encontro Regional de Estudantes de Biblioteconomia, Documentação, Ciência da Informação e Gestão da Informação, 2011.</p>		
<p>SCRIPTLATTEs. Uma ferramenta para extração e visualização de conhecimento a partir de Curriculos Lattes. Disponível em: http://scriptlattes.sourceforge.net/. Acessado em 07/08/2017.</p>		
<p>SEZÕES, C., OLIVEIRA, J., BAPTISTA, M. Business Intelligence. Porto: SPI – Sociedade Portuguesa de Inovação, 2006.</p>		

APENDICE III - Pôster apresentado no VI SE²PIN – Seminário de Extensão, Ensino, Pesquisa e Inovação do IFPR.



Natureza
Liberdade
SUSTENTA
BILIDADE

INSTITUTO
FEDERAL
Paraná

APLICANDO TÉCNICAS DE BUSINESS INTELLIGENCE SOBRE DADOS DA PRODUÇÃO CIENTÍFICA – UM ESTUDO DE CASO

Autora: Talita R. P. Monteiro / IFPR Campus Palmas (t.pmonteiro@outlook.com)
Orientadora: Lauriana Paludo (lauriana.paludo@ifpr.edu.br)

INTRODUÇÃO

Business Intelligence é o termo utilizado para o processo composto por métodos, métricas, ferramentas e entre outros, que de uma forma genérica recolhem dados, transforma-os em informações e, sequencialmente, em conhecimento (SANZÓES, OLIVEIRA e BAPTISTA, 2006).

Fonte de dados:

Figure 1 – Arquitetura de um BI

Na literatura são encontrados alguns trabalhos relacionados a DW de dados educacionais e utilização de indicadores de perfil de produção de seus pesquisadores, como o trabalho de Santana et al (2011), que através de uma análise qualitativa identificou os principais aspectos e tendências da produção científica do PPGS da UFPE e o de FILHO et al (2015) que possibilita a simulação da geração de relatórios ad-hoc e dashboards de consultas através do ambiente open source de Ensino à Distância Amadeu.

OBJETIVO

Apresentar a modelagem e implementação de um Data Warehouse de dados de publicações científicas de um campus de um Instituto Federal do Paraná escolhido para estudo de caso.

METODOLOGIA

Foram seguidas três etapas principais (I) extrair os dados do COPE e do currículo Lattes dos pesquisadores do campus, (II) tratar e Integrar os dados na área de armazenamento temporário chamada de ODS (Operational Data Storage) e (III) criar um data warehouse de produções científicas do período compreendido entre 2014 a 2016.

Optou-se pela utilização de softwares que atendam às quatro liberdades do software livre: executar, modificar, redistribuir e distribuir versões modificadas de um programa, conforme especificado pela Free Software Foundation (FSF).

RESULTADOS

São resultados deste trabalho: a revisão da literatura na área de BI aplicado em Instituições de ensino, o processo de ETL realizado nas fontes de dados dos currículos Lattes e do COPE do campus, a modelagem e criação do DW de produção científica.

Figura 2 – Metodologia de Trabalho:

Figure 2 – Metodologia de Trabalho

Figura 3 – Modelagem Dimensional de DW:

Figure 3 – Modelagem Dimensional de DW

Figura 4 – ETL: fluxo de dados para criação da Matriz da Matriz de Análise Multidimensional no ambiente de ODS:

Figure 4 – ETL: fluxo de dados para criação da Matriz da Matriz de Análise Multidimensional no ambiente de ODS

CONSIDERAÇÕES FINAIS

O processo de ETL foi complexo e demorou cerca de dois meses para ser concluído devido a diversidade existente em termos de estruturas de dados nas bases dos sistemas utilizados. Os dados extraídos foram primeiramente integrados na área Intermediária de armazenamento e após a integração e limpeza, exportados para o DW. Como trabalhos futuros pretende-se desenvolver a camada lógica responsável por manipular e analisar os dados armazenados no DW sob múltiplas perspectivas, também chamada de processamento analítico ou cubo OLAP, onde será possível identificar o perfil da produção científica do campus.

GESIS

APENDICE IV - Oficina proposta ao VI SE²PIN – Seminário de Extensão, Ensino, Pesquisa e Inovação do IFPR.

Oficina: Integrando Dados Através da Ferramenta Gratuita *Pentaho Data Integration*

Talita Fatima Paula Monteiro, Lauriana Paludo
(tfpmonteiro@outlook.com; lauriana.paludo@ifpr.edu.br)

O processo de **Extração, Transformação e Carga – ETL**, é o processo que inicia com a **Extração** do dado de uma fonte de dados qualquer, seja de um banco de dados, de um arquivo JSON, web service, planilha, etc, passa pela **Transformação**, ou seja, a seleção, tradução, decodificação para um formato diferente, como por exemplo, tornar mais compreensível o significado da letra “M” armazenada em um banco de dados que refere-se ao gênero masculino, ou a letra “F” para feminino e encerra com a **Carga**, que é o processo de armazenar esses dados transformados em uma outra fonte de dados, por exemplo, de uma planilha para um arquivo xml. Geralmente o processo de ETL está vinculado ao desenvolvimento de um *Data Warehouse* - DW, ou armazém de dados históricos de um ambiente corporativo. É utilizado em ferramentas de *Business Intelligence* – BI, que auxiliam os gestores na tomada de decisões e análises gerenciais. O objetivo dessa oficina é conhecer a solução de software livre para extração, tratamento e carga de dados *Pentaho Data Integration*, também conhecida como *Kettle* ou *Spoon*. Essa ferramenta, na versão *community*, atende às quatro liberdades do software livre: executar, modificar, redistribuir e distribuir versões modificadas de um programa. Apresenta uma interface gráfica de fácil uso e é mais rápida que a codificação de rotinas numa linguagem de programação como PHP ou Java. A ferramenta *Pentaho Data Integration* oferece uma diversidade de conectores para realizar conexões com as mais diversas bases de dados, reusabilidade onde uma carga pode ser aproveitada dentro de outras cargas, documentação e uma ativa comunidade de desenvolvedores. Nesta oficina é demonstrado de forma prática o processo de instalação da ferramenta, seus principais componentes e as atividades de extrair dados de uma planilha eletrônica, transformar, tratar campos, remover dados desnecessários, remover registros duplicados e dar a carga, salvando os dados transformados em uma base de dados relacional como o *Mysql* e o *PostgreSQL*. A oficina tem como público alvo estudantes da área de informática ou que tenham interesse em conhecer a ferramenta e novas maneiras de se trabalhar com extração, transformação e carga de dados.