

A pesquisa antropológica em oficinas virtuais: reflexões sobre pesquisa em uma plataforma online de desenvolvimento de software de código aberto¹

Victor Vieira Paulo - UFSC/Santa Catarina.

Palavras-chave: Software, Internet, Antropologia

Introdução

O desenvolvimento de tecnologias relacionadas à Internet tem apresentado forte ênfase na criação de plataformas online voltadas para a promoção e suporte de interações sociais. Diante deste contexto, a rede de computadores tem sido cada vez mais vista como meio para o fomento de colaborações entre indivíduos geograficamente dispersos, sobretudo no que tange a cooperação para realização de atividades de natureza cognitiva. (SMART et al., 2017, p.272). Tentativas de exploração deste potencial tem ocasionado a proliferação de plataformas online que provêm as condições para que seus usuários participem em esforços colaborativos de produção de artefatos informacionais, científicos e tecnológicos. Um exemplo bem conhecido é a enciclopédia global Wikipédia, na qual uma multiplicidade de sistemas sociotécnicos criam as condições para que usuários da rede engajem-se na produção e melhoramento de artigos informativos a respeito de diversas temáticas (SMART et al., 2017, p.274).

O campo de pesquisas antropológicas sobre ciência e tecnologia tem como uma de suas principais tarefas o estudo da prática realizada nos laboratórios de produção de fatos científicos e oficinas de criação de objetos tecnológicos. O fenômeno de realização de tarefas científicas e tecnológicas através de plataformas na Internet indica a emergência de modalidades virtuais de laboratórios e oficinas. Dado que os materiais, atividades e relações constituintes da prática da ciência e tecnologia têm sido realocados para a nuvem, é necessário que os antropólogos que os investigam também passem por um upload.

¹ Trabalho apresentado na 31ª Reunião Brasileira de Antropologia, realizada entre os dias 09 e 12 de dezembro de 2018, Brasília/DF.

O presente artigo parte de uma pesquisa em andamento sobre esta temática. O objetivo aqui é delinear algumas das oportunidades e desafios que estas modalidades de produção de fatos científicos e objetos tecnológicos apresentam à antropologia da ciência e tecnologia, e apontar alguns dos caminhos pelos quais a pesquisa antropológica pode ser estendida a este domínio.

Para realizar este objetivo será examinada uma instância específica do fenômeno em questão: o desenvolvimento de programas computacionais de código aberto. O exame dispensará especial atenção a uma das plataformas online voltadas para o suporte deste tipo de atividade, chamada GitHub, uma vez que esta constitui o “campo” da pesquisa em andamento na qual se baseia este artigo. As próximas páginas apresentarão uma descrição daquilo que se designa por programa computacional e código aberto, bem como da modalidade de desenvolvimento através do qual esse tipo de artefato técnico é criado.

Programas computacionais de código aberto

Um programa computacional é uma sequência de instruções a serem executadas por um computador para realizar uma dada tarefa. Estas instruções, especificadas por programadores através de uso de linguagens formais especializadas, constituem o código fonte do programa.

É comum programas usados em computadores pessoais sejam obtidos na forma de pacotes fechados, contendo uma versão que pode ser executada, mas não modificada, já que o código fonte encontra-se inacessível. Este código é propriedade exclusiva da empresa ou organização desenvolvedora do programa, e apenas ela pode modificar o seu funcionamento, lançando atualizações para os usuários. O desenvolvimento de programas de código aberto constitui uma contraposição direta à esta forma de criação e distribuição de software proprietário. Distribuído sob licenças de propriedade intelectual que garantem acesso ao código fonte, e eventualmente o direito livre apropriação, modificação, e mesmo redistribuição, um programa deste tipo tende a constituir um esforço coletivo. Qualquer pessoa com as competências necessárias pode efetuar modificações ou correções no código fonte do software, enviá-las para os distribuidores originais propondo sua integração no projeto principal, e em alguns casos, a depender

da licença de propriedade intelectual utilizada, até mesmo distribuí-las como uma versão alternativa do programa em questão.

A história dos programas de código aberto é relativamente longa no âmbito da tecnologia da informação, e tem se desenrolado em conjunto com a evolução da Internet. O próprio projeto acadêmico-militar ARPANET, ancestral da rede mundial de computadores criado ainda na década de 60, tinha como uma de suas motivações o estabelecimento de uma rede que facilitasse o compartilhamento de programas e conhecimento entre comunidades de programadores (TUOMI, 2001). Posteriormente, a difusão Internet permitiu a emergência de modalidades de desenvolvimento cooperativo em uma escala até então inédita.

O projeto do sistema operacional Linux constitui um marco na evolução do desenvolvimento de programas de código aberto. Iniciado no começo da década de 90, os primeiros anos deste projeto ocorreram num contexto de massificação do acesso à Internet em computadores pessoais. O desenvolvimento do sistema operacional foi iniciado pelo engenheiro de software finlandês Linus Torvalds, que em 1991 disponibilizou o código fonte de seu programa em um canal de comunicação na Internet, buscando atrair o engajamento de outros programadores no projeto. A aposta foi bem sucedida, e Linus passou a liderar uma crescente comunidade de usuários e colaboradores, organizada através de uma metodologia de desenvolvimento inovadora. Enquanto o modelo de desenvolvimento tradicional vigente no período previa a realização do trabalho por um time restrito de programadores, a comunidade do Linux permitia que qualquer um enviasse correções e novas funcionalidades. Cabia a Torvalds, e posteriormente a uma equipe de desenvolvedores escolhidos por ele, aceitar ou rejeitar as contribuições recebidas e guiar os rumos do desenvolvimento. O desenvolvimento do sistema operacional foi enormemente bem sucedido, passando rapidamente a competir com outros programas do mesmo tipo comercializados à época. Esse sucesso é atribuído em grande parte à constituição de uma metodologia de desenvolvimento capaz de tirar proveito da potencialidade de colaboração em larga escala facultada pela Internet (RAYMOND, 2000).

A criação de um sistema operacional de qualidade provou que o desenvolvimento de programas de código aberto através da Internet configurava uma metodologia adequada para a realização de uma das mais complicadas tarefas da

engenharia de software (RAYMOND, 2000). Diante de tal sucesso, não demorou para que tal modelo de desenvolvimento passasse a ser comparado favoravelmente a várias das práticas consolidadas na indústria de programas computacionais (TUOMI, 2001). A partir daí o desenvolvimento colaborativo de programas de código aberto através da Internet se consolidou. Tal processo teve como um de seus efeitos a emergência de plataformas online que provêm suporte para esta atividade. A pesquisa que embasa este artigo tem por enfoque a maior e mais popular destas plataformas, chamada GitHub.

Plataforma GitHub

GitHub é uma plataforma online que oferece a seus usuários a possibilidade de criação de repositórios para armazenamento do código fonte de programas computacionais em desenvolvimento². Esses repositórios podem ser de acesso público ou privado. Uma das funcionalidades principais do GitHub é permitir que os usuários possam engajar-se no trabalho realizado em repositórios criados por outras pessoas. No caso de um repositório público, qualquer usuário da plataforma pode ter acesso ao código fonte ali armazenado, e por vezes mesmo participar do projeto de desenvolvimento.

É comum que os programadores se refiram à criação de código de programas como *escrita*, e de fato, essa atividade apresenta algumas das mesmas dificuldades vivenciadas na elaboração de outros tipos de textos. Múltiplas versões de um texto emergem antes que ele seja dado por concluído, e torna-se necessário organizá-las. O problema é particularmente evidente quando um texto é desenvolvido pelo trabalho simultâneo de vários autores. Para permitir que vários programadores trabalhem em um mesmo código, a cooperação nos repositórios públicos do GitHub se dá através de um sistema de controle de versionamento chamado Git.

O usuário que pretende contribuir para um repositório público começa por fazer o download do código fonte da versão atual do programa armazenado no repositório

² É importante notar que embora a plataforma GitHub tenha por foco primário o suporte ao desenvolvimento de programas de computador, seus possíveis usos abrangem várias outras atividades. A tecnologia empregada na plataforma permite o armazenamento de diversos tipos de arquivos digitais, abrindo caminho para a hospedagem de conjuntos de dados e documentos. Um dos usos da plataforma, de grande interesse para os estudos da ciência e tecnologia, tem sido a publicização e construção colaborativa de bases de dados científicos (PERKEL, 2016).

para seu computador. Aí, ele pode efetuar as modificações que desejar, e então enviá-las novamente para o repositório, juntamente com uma descrição daquilo que fez. O Git permite que cada repositório tenha administradores, que são capazes de aceitar ou negar as modificações ao código propostas por usuários. Caso o administrador do repositório aceite as modificações enviadas pelo usuário, estas são incluídas a versão atual do código. A partir daí, o próximo usuário deverá obter essa versão modificada para continuar o desenvolvimento. É possível também que o usuário faça uma “cópia” do repositório, criando uma versão alternativa do programa cujos rumos de desenvolvimento podem ser administrados por ele.

Uma das características fundamentais deste sistema é que ele não permite apenas integrar de forma segura o novo código ao já existente, mas também guarda registros de todo o processo. É possível observar cada modificação já enviada, tenha sido ela aceita ou rejeitada, para qualquer arquivo de código fonte do repositório, incluindo informações sobre a autoria da modificação, a data, e alguns comentários do autor explicando aquilo que foi feito. É possível também ter acesso a todas as versões anteriores do código, o que permite observar o funcionamento do programa antes de cada modificação.

Este sistema de versionamento constitui uma ferramenta autônoma, que pode ser usado sem recorrer ao GitHub³. A plataforma GitHub integra o Git com um ambiente que promove a interação entre os programadores, oferecendo diversas funcionalidades para este fim. Modalidades de busca permitem que os usuários encontrem repositórios públicos de seu interesse, ou troquem informações com outros usuários a esse respeito. No âmbito do desenvolvimento, o GitHub faculta uma multiplicidade de ferramentas voltadas para o suporte da cooperação entre os participantes de um projeto de programa computacional. Várias destas ferramentas envolvem formas específicas de comunicação entre os participantes, através dos quais eles podem inteirar-se do trabalho a ser feito, examinar e discutir modificações específicas do código fonte, ou tratar de questões mais amplas sobre os encaminhamentos do projeto. As mensagens trocadas através destas ferramentas costumam permanecer publicamente acessíveis, e podem ser examinadas futuramente.

³ O sistema de versionamento Git precede a plataforma GitHub. De fato, a última foi construída ao redor do primeiro. Git é mais um dos frutos da comunidade de desenvolvimento Linux, sendo desenvolvido sob direção de Linus Torvalds como uma ferramenta para auxiliar o desenvolvimento do sistema operacional.

É comum também que a página de cada projeto público do GitHub referencie uma série de normas, documentos, recursos, e ferramentas relevantes para a realização do desenvolvimento. Muitas vezes os elementos referenciados são externos à plataforma, tais como sites oficiais dos projetos; listas de e-mail; chats; etc. Comumente estes recursos também são publicamente acessíveis, e vários deles permanecem arquivados para consultas futuras. Todos estes históricos e informações, somados àqueles já disponibilizados pela plataforma e pelo sistema de versionamento Git, permitem que um programa possa ser entendido por aqueles que já contribuem ou pretendem começar a contribuir para seu desenvolvimento, e também possibilitam identificar as modificações que foram realizadas no programa e posteriormente causaram erros inesperados.

Para além dos históricos específicos produzidos e mobilizados no contexto do desenvolvimento, GitHub disponibiliza ainda outros dados de natureza quantitativa. As páginas de cada repositório da plataforma contém uma seção que exhibe variadas estatísticas sobre a atividade ali realizada, e as Interfaces de programação de aplicações⁴ oferecidas pela plataforma permitem que grandes volumes de dados sobre os repositórios públicos hospedados sejam coletados de forma automatizada.

Como será visto a seguir, o desenvolvimento colaborativo de programas de código aberto através da Internet e as plataformas online que suportam tal atividade apresentam certas diferenças em relação a alguns processos considerados constituintes da dinâmica de produção de ciência e tecnologia.

Código aberto e caixas-pretas

Os estudos antropológicos de ciência e tecnologia sob a teoria ator-rede tem como corolário a descrição dos processos pelos quais emergem os conteúdos produzidos pela ciência através da análise da prática científica (LATOUR, 1987, p. 21). Essa abordagem pretende ir além dos estudos sociológicos que examinam as instituições científicas mas se recusam a estender à análise social aos conteúdos da ciência, e

⁴ Uma Interface de programação de aplicações, comumente referenciada pela sigla API (application program interface), consiste num dispositivo que permite que funcionalidades disponibilizadas por um programa possam ser acessadas por outro programa sem que este último precise envolver-se com detalhes do funcionamento interno do primeiro.

também àquelas pesquisas que se baseiam apenas no discurso proferido pelos cientistas à respeito de suas práticas (LATOUR & WOOLGAR, 1997, p.19-20). Os defensores do estudo da prática da ciência e tecnologia pontuam como vantagem desta abordagem o enfoque no momento onde os enunciados científicos e objetos tecnológicos ainda não foram plenamente *reificados*⁵, permitindo analisar estes conteúdos em relação ao contexto do qual emergiram, antes que esses contextos sejam apagados e os conteúdos passem a ser tratados como fatos do mundo natural ou objetos plenamente determinados por constrangimentos materiais e técnicos (LATOUR, 1987, p. 21).

O conceito de *caixa-preta* tem sido usado no âmbito da teoria ator-rede para descrever os enunciados científicos e artefatos tecnológicos que já se encontram reificados. A origem do conceito pode ser traçada até a cibernética, e em seu sentido original constitui uma meio para facilitar a apreensão e descrição de sistemas demasiado complexos. A aplicação desta concepção faz com que um setor complexo de um dado sistema passe a ser descrito apenas nos termos do tipo de input que aceita, e no tipo de output que produz, desconsiderando o processo pelo qual inputs são transformados em outputs (LATOUR, 1987, p. 2-4). Em analogia com o conceito da cibernética, a teoria ator-rede postula ser necessário que um fato ou artefato seja reificado em uma caixa preta para reduzir sua complexidade, permitindo que ele seja mobilizado de forma não problemática em outras redes, para além daquela na qual surgiu.

Como quaisquer tecnólogos, os programadores manipulam e produzem caixas-pretas incessantemente. O termo caixa-preta faz parte mesmo do discurso “nativo”, já que nomeia um princípio de design no âmbito da engenharia de software. É desejável que o usuário possa mobilizar um programa para os objetivos que almeja sem que tenha que mergulhar nos meandros operacionais do mesmo. Para atingir este fim, os programadores se esforçam para fazer com que seus programas possam ser utilizados como caixas-pretas, dotando os softwares de interfaces que permitem aos usuários operar a partir de especificações dos tipos de inputs recebidos pela interface, e dos tipos de outputs devolvidos a partir destes inputs (ECK, 2014, p. 135-137).

Os processos de reificação certamente são presentes no desenvolvimento de softwares de código aberto. Em uma interessante análise do desenvolvimento do Linux,

⁵ Uma definição detalhada do conceito de reificação pode ser encontrada em LATOUR & WOOLGAR, 1997, p.63-67; 269-270; 275-277.

o pesquisador finlandês Ilkka Tuomi (2001) demonstrou que a produção de caixas-pretas ocorre em vários níveis na elaboração do sistema operacional. Em primeiro lugar, diversos artefatos tecnológicos produzidos anteriormente - tais como a Internet, os sistemas operacionais da família Unix e o ferramental desenvolvido para o sistema operacional GNU - conformaram uma base estável de tecnologias já reificadas sobre a qual o novo sistema pôde ser construído. Em segundo lugar, os usuários finais do Linux também podem se apropriar deste sistema como uma caixa-preta, mobilizando-o para seus fins sem nunca atentar para os contextos a partir dos quais ele foi produzido, e nem para o funcionamento de seus mecanismos internos. Por fim, o próprio desenvolvimento do Linux exige que os desenvolvedores produzam caixas pretas uns para os outros. Este sistema operacional, como qualquer software complexo, não é desenvolvido em um único bloco, mas sim em vários módulos dotados de funções distintas e trabalhados por comunidades de desenvolvimento diferentes. Como é comum que um módulo seja apropriado como um recurso no desenvolvimento de outro módulo, é desejável que os módulos definam interfaces que facilitem tal apropriação. Em um sistema da envergadura do Linux este processo de redução de complexidade no processo de desenvolvimento torna-se essencial.

A análise de Tuomi evidencia que as caixas pretas desempenham um papel crucial no desenvolvimento do Linux. Há razões para supor que esta não seja uma característica peculiar deste software, sendo na verdade um elemento constitutivo de qualquer outro programa computacional. Entretanto, o ponto de interesse teórico para a antropologia é o fato de que a descrição do autor evidencia algo de particular na dinâmica de reificação presente no desenvolvimento de código aberto. Enquanto na prática científica e tecnológica é costumeiro que os produtores de caixas pretas invistam grande esforço para que elas permaneçam fechadas (LATOUR & WOOLGAR, 1997, 274-278), os desenvolvedores de programas de código aberto fazem questão de garantir que as caixas-pretas que produzem possam ser facilmente abertas.

Como já dito, uma das características singulares da metodologia de código aberto é o fato de que aí encontra-se ausente a clivagem entre os usuários, que não detêm acesso ao código fonte do programa, e os desenvolvedores, capazes de examiná-lo e modificá-lo. É importante ressaltar, contudo, que a liberação do acesso ao código fonte não trata-se de mera concessão aos usuários, sendo pelo contrário condição

de possibilidade do desenvolvimento de código aberto. Os defensores da metodologia de desenvolvimento elaborada por Linus Torvalds argumentam que uma das grandes vantagens que esse modelo tem em relação aos contextos tradicionais de produção de software é o fato de que ele permite engajar os usuários no desenvolvimento, tornando-os co-desenvolvedores (RAYMOND, 2000)⁶. Como o usuário têm acesso ao código fonte, é possível que ao encontrar um erro no funcionamento do programa, ele possa resolvê-lo e enviar a correção para ser integrada na versão principal do software, ou no mínimo consiga produzir um relatório de erro mais preciso, já que informado pela análise do código fonte. Isto supostamente permite que o programa seja escrutinado por muitas pessoas diferentes, fazendo com que os erros logo sejam encontrados e resolvidos (RAYMOND, 2000). Alguns usuários podem até mesmo ir além da mera correção de defeitos, contribuindo com o desenvolvimento de novas funcionalidades para o programa.

O acesso ao código fonte facultado aos usuários de programas de código aberto não é o único fator que faz com que o processo de desenvolvimento possa ser escrutinado. Como já dito, esta modalidade de desenvolvimento de software cresceu entrelaçada à Internet, e em grande parte constitui um aproveitamento das possibilidades produção colaborativa propiciadas por esta tecnologia. A exploração deste potencial requer a criação de uma comunidade de desenvolvimento no ambiente virtual, e um dos fatores fundamentais que viabilizam esse tipo de comunidade é a publicidade do processo de desenvolvimento⁷.

Plataformas como GitHub constituem ambientes criados para suportar a constituição e operação de comunidades de desenvolvimento. GitHub desempenha um

⁶ Obviamente o engajamento de usuários como co-desenvolvedor exige que estes tenham as competências necessárias para examinar, entender e modificar o código fonte. Ou, no mínimo, que tenham o interesse em desenvolver estas competências. Certamente esta condição era presente na emergência do Linux, já que grande parte de seus usuários, ao menos nos primeiros anos, eram também programadores (RAYMOND, 2000). E continua a ser presente em vários dos atuais programas de código aberto (SCACCHI, 2007, p. 246), nos quais boa parte dos desenvolvedores são também usuários. O atesta o fato de que boa parte dos projetos desenvolvidos na plataforma GitHub são ferramentas utilizadas para desenvolvimento de outros programas computacionais.

⁷ É necessário esclarecer que o nível de publicidade não é igual para todos os projetos de desenvolvimento de código aberto levados adiante através da colaboração na Internet, e nem é igual em todas as áreas e períodos de um mesmo projeto. Algumas atividades e decisões concernentes ao desenvolvimento podem ser realizadas em âmbitos de acesso restrito, localizados ou não no ambiente virtual. Contudo um significativo grau de publicidade é um elemento constitutivo dos processos de desenvolvimento de código aberto realizados através da internet (SCACCHI, 2007, p. 246).

papel fundamental na publicidade dos processos de desenvolvimento, no que tange os repositórios públicos. Em primeiro lugar, tal ferramenta constitui um ambiente de desenvolvimento de projetos que os torna acessíveis para qualquer um que detenha uma conexão com a internet. Em segundo lugar, a plataforma provê os meios que permitem que programadores encontrem repositórios de seu interesse, seja através dos mecanismos de busca, seja através dos componentes da plataforma permitem o relacionamento entre usuários (DABBISH et al., 2012, p.1285). Em terceiro lugar, o GitHub mantém vários registros das ações e acontecimentos ocorridos durante o processo de desenvolvimento, e estes históricos mantêm-se publicamente acessíveis nos repositórios públicos⁸.

Se o acesso ao código-fonte é pré-condição para que os usuários possam se tornar co-desenvolvedores, a transparência garantida pela publicidade e historicização do processo de desenvolvimento é necessária para permitir a constituição de uma comunidade de contribuidores em torno do projeto de programa computacional, e a criação e manutenção de uma comunidade de desenvolvedores é tida como uma das condições fundamentais para que um projetos de desenvolvimento de código aberto realize-se (SCACCHI, 2007, p. 269).

Fica evidente então que a própria noção de desenvolvimento de código aberto implica que a caixa preta na qual o programa computacional foi reificado pode, e mesmo deve, ser aberta. O programa não se apresenta ao usuário como acabado, mas como um objeto em constante construção, e o contexto através do qual surge o objeto não é varrido para debaixo do tapete. Entregar a chave do código fonte ao usuário é convidá-lo a integrar a rede através da qual emerge o programa, e tornar público o processo de desenvolvimento através da Internet é garantir o acesso a essa rede. Se a mobilização de processos de reificação é uma característica que o desenvolvimento de programas de código aberto tem em comum com outras modalidades de produção de artefatos tecnológicos, parece ser um traço peculiar deste fenômeno a necessidade de garantir que as caixa-pretas aí produzidas possam ser facilmente abertas.

⁸ Esta ênfase na publicidade e na historicização do desenvolvimento de código aberto vai além do uso do GitHub ou de plataformas similares, e se estende às diversas ferramentas mobilizadas no desenvolvimento. O atestam os canais de comunicação externos às plataformas usados pela comunidade de desenvolvimento, como listas e e-mail e bate-papos, que são também frequentemente historicizados e publicamente acessíveis.

Esta modalidade singular de reificação pode revelar-se um problema de interesse teórico para o estudo da ciência e tecnologia. O presente texto, baseado em uma pesquisa num estágio ainda bastante inicial, limita-se a produzir uma descrição do fenômeno em questão, cuja análise deverá ser aprofundada futuramente. Aqui, nosso objetivo principal é delinear algumas das oportunidades de investigação ofertadas por tal fato.

Antropologia digital além da etnografia

Para Latour, o estudo da prática ciência e do desenvolvimento de tecnologia é uma tentativa de abrir caixas pretas. Ao atentar para o contexto de criação dos fatos e objetos, os praticantes da teoria ator-rede buscam identificar as controvérsias que dividiram cientistas e tecnólogos no processo de emergência do objeto, e descrever como as controvérsias foram sanadas (LATOUR, 1987, p. 4). Em suma, abrir as caixas-pretas permite trazer à tona o contexto de produção de um fato científico ou objeto técnico, e descrever os processos pelos quais ele veio a ser o que é. Contudo, esta tarefa não necessariamente é fácil. Na ciência, o processo de reificação visa garantir que os enunciados e objetos se tornem parte do real, ou seja, que questioná-los se torne por demais custoso (LATOUR & WOOLGAR, 1997, p. 278). O ato de abrir uma caixa preta pode ser encarado como um questionamento da própria realidade amplamente aceita.

A interpretação anteriormente exposta indica que o desenvolvimento de programas de código aberto apresenta um objeto de pesquisa possivelmente promissor para os estudos antropológicos de ciência e tecnologia, já que as próprias características do fenômeno garantem que a abertura de caixas-pretas nesse âmbito seja facilitada por vários fatores. Entretanto, a abordagem desse fenômeno em específico, e dos esforços de produção colaborativa de ciência e tecnologia através da Internet em geral, apresentam também um desafio para a antropologia, já que exigem um arsenal de métodos e técnicas diverso daquele que vem sendo tradicionalmente empregados nas pesquisas desta disciplina.

É certo que a etnografia, método característico da antropologia, tem sido mobilizada no estudo não apenas do objeto tratado aqui (SCACCHI, 2007, p. 280-281),

mas também de diversos outros contextos de interação social propiciados pela Internet. De fato, as reflexões acerca da aplicação das perspectivas teóricas e ferramentas metodológicas da disciplina ao estudo das interações sociais em ambientes virtuais tem constituído um dos motes da chamada ciberantropologia (SEGATA & RIFIOTIS, 2016, p.12-13). Contudo, existem razões para crer que as pesquisas antropológicas dos processos de desenvolvimento de tecnociência através da internet não demandarão apenas uma extensão da etnografia a um novo contexto, mas sua suplementação por um conjunto de métodos e técnicas próprios do mundo digital.

Uma das características definidoras da Internet é o fato de que todas as ações que ocorrem através dela podem ser, e cada vez mais são, registradas. Cliques; movimentos de mouse; requisições de acesso a determinadas páginas; visualizações e compartilhamentos de conteúdos; pesquisas a respeito de informações e produtos. Todos esses atos e várias outros, sejam eles perpetrados por humanos ou por sistemas automatizados, são registrados em bancos de dados, cuja quantidade e volume hoje crescem exponencialmente. Um novo universo sociotécnico tem emergido a partir das técnicas de armazenamento, tratamento e análise de dados (sub)produzidos pelas interações na rede mundial de computadores.

Muitos destes dados podem ser acessados apenas pelas empresas que administram os serviços e plataformas nas quais as ações registradas são realizadas, e tais registros têm se tornado mercadorias valiosíssimas no sistema econômico vigente. Contudo, em determinados contextos, estas informações podem ser visualizados pelos usuários de serviços online, e diversas técnicas de extração permitem que elas sejam capturadas e armazenadas bancos de dados não filiados às empresas proprietárias das plataformas que conformam a paisagem contemporânea da Internet.

Este fato constitui um dos motivos pelos quais os processos de criação colaborativa de programas de código aberto tem capturado o interesse de estudiosos de desenvolvimento de software no âmbito das ciências sociais. O fato de que o desenvolvimento se dá através da Internet, e de forma pública, permite que os registros digitais decorrentes dos processos de desenvolvimento possam ser submetidos à técnicas automatizadas ou semi-automatizadas de captura, tratamento e análise de dados, tais como mineração de dados textuais, análise estatística e aprendizado de máquinas (SCACCHI, 2007, p. 281).

A princípio tal situação não parece encorajadora para a antropologia. As grandes massas de dados produzidas a cada instante pelos bilhões de agentes conectados à Internet parecem colocar-se fora do âmbito de análise da disciplina, que sempre procedeu primariamente através de métodos qualitativos. Contudo, um esforço que objetive a integração da representação e análise destes dados à pesquisa antropológica é não apenas desejável, mas necessário. Se uma das características do ambiente virtual é a constituição de registros de tudo que nele acontece, então deixar de analisar o processo pelo qual esses registros são constituídos, e deixar de analisar o ambiente em questão através dos registros que ele produz, significa ignorar um componente fundamental do fenômeno estudado.

Esta questão se estende a qualquer pesquisa antropológica no ambiente virtual, mas revela-se de particular interesse para os estudiosos da ciência e tecnologia, interessados em seguir os traços deixados pelos agentes participantes da rede sociotécnica, e em revelar as complexas agências não-humanas envolvidas nestas redes (LATOUR, 1996). Os dados que proliferam nas redes da Internet anunciam-se como um locus privilegiado para este tipo de inquérito, na medida em que constituem ao mesmo tempo traços do que fizeram os agentes (humanos ou não) participantes da rede, e traços dos mecanismos sociotécnicos subjacentes responsáveis pela produção de tais registros.

O objetivo aqui não é defender o abandono da etnografia em prol de alguma das técnicas de análise quantitativa que emergem em paralelo à expansão desse estranho amontoado de registros. Pelo contrário, trata-se de examinar os meios pelos quais representações e análises destes aglomerados de dados podem ser mobilizadas no interior de uma moldura qualitativa. Nessa perspectiva as preocupações concernentes à garantia do rigor metodológico típicas do ferramental de pesquisa quantitativa não se aplicam. A questão a ser enfrentada diz respeito sobretudo ao encontro dos meios pelos quais os registros podem ser examinados de forma a tornarem-se significativos para o tratamento das questões de interesse da antropologia.

Não cabe proceder uma listagem dos possíveis métodos e técnicas a serem utilizados para o tratamento de registros digitais em pesquisas antropológicas, já que é possível que esses métodos variem de acordo com o campo a ser examinado e as questões teóricas a serem tratadas. Tudo o que se pode dizer com alguma segurança é que é provável que as pesquisas antropológicas que busquem incluir a análise de

registros digitais em seu rol metodológico venham a se beneficiar de técnicas de visualização de dados que têm sido desenvolvidas para tratar das massas de registros digitais emergentes. Afinal a produção de uma visualização gráfica de um conjunto de dados tem por efeito tornar visíveis determinadas possibilidades de interpretação destes dados, permitindo que o pesquisador tome estas interpretações possíveis como pontos de partida de sua análise. Além disso, as representações imagéticas que resultam deste processo podem ser facilmente inseridas no texto antropológico.

Para ilustrar as possibilidades anteriormente citadas, serão listadas algumas das informações oferecidas pelos registros produzidos no desenvolvimento colaborativo de programas de código aberto através da Internet, e por fim será evidenciada uma questão de interesse antropológico cujo tratamento pode ser beneficiado pelo exame destas informações.

O já citado artigo de Ilkka Tuomi (2001) constitui um exemplo interessante de integração de métodos de análise e representação de dados à descrição qualitativa de uma rede sociotécnica. Análises da variação do espaço em disco ocupado pela sucessão de versões do Linux, da quantidade de linhas de código fonte adicionadas e subtraídas ao longo do tempo em diferentes módulos do sistema, e de outros dados similares permitiram ao autor identificar dinâmicas envolvidas na emergência do artefato técnico em questão, e examiná-las à luz da teoria ator rede e da teoria das comunidades de prática. O uso destas informações para demonstrar como os processos de inovação e de estabilização operam desigualmente entre os diferentes módulos que compõe o Linux são particularmente notáveis já que permitem evidenciar um processo de reificação sedimentar, no qual determinados componentes do sistema vão se tornando caixas-pretas na medida em que são estabilizados, e passam a constituir camadas sob as quais repousam os componentes mais novos, ainda em desenvolvimento.

A análise de Tuomi baseou-se sobretudo nos arquivos de código fonte do Linux. O atual contexto do desenvolvimento de programas de código aberto através da Internet oferece oportunidades de análises de dados significativamente mais amplas. Sistemas de versionamento como o Git, plataformas como o GitHub, e interfaces de comunicação como chats online e e-mails produzem uma gama muito maior de registros. E a publicidade destes dados no contexto em questão garante que eles possam ser capturados e analisados por pesquisadores dispostos a fazê-lo.

O sistema de versionamento Git guarda um histórico completo de todas as modificações realizadas no código fonte de um repositório administrado através dele, permitindo saber quais modificações foram feitas, quais os setores do código foram mais modificados, quantas linhas foram adicionadas ou subtraídas. Isto abre caminho para análises das dinâmicas envolvidas no processo do desenvolvimento técnico com um nível de minúcia maior do que o alcançado por Tuomi. Além disso, o exame dos dados do sistema de versionamento permite que as ações dos participantes do processo de desenvolvimento, e mesmo as relações entre essas ações, sejam escrutinadas. Através destes meios, é possível saber que usuários trabalharam em quais setores do projeto, quantas linhas de código foram contribuídas por cada usuário, quais usuários modificaram ou apagaram trechos de código contribuídos por outros usuários, etc.

Os dados da plataforma GitHub, e de outras plataformas similares, apresentam ainda outras possibilidades. Por meio deles, é possível investigar os padrões de comunicação entre os desenvolvedores, a evolução da comunidade de desenvolvedores ao longo do tempo, identificar que propostas de modificações do código fonte geraram mais discussões entre a comunidade, etc. Além disso, técnicas de análise de redes sociais podem ser utilizadas para investigar as relações entre os usuários dentro de um mesmo repositório, ou entre diferentes repositórios (por exemplo: Madey et al., 2002). Por fim, estes registros permitem analisar ainda o processo pelo qual emergem diversas cópias modificadas de um mesmo repositório, que são posteriormente integradas ao ramo de desenvolvimento principal, ou se tornam versões alternativas do programa computacional em questão (ver p.4-5 deste artigo).

Tais informações obtidas através do exame dos dados podem constituir amparos úteis para a descrição do processo da emergência de artefatos técnicos, sobretudo quando examinadas em conjunto com os resultados de uma etnografia digital produzida no mesmo campo de onde foram extraídos os dados. Estes tipo de procedimento configura um caminho potencialmente promissor para levar adiante questões clássicas dos estudos de ciência e tecnologia. Como exemplo, afigura-se a questão da mapeamento das controvérsias que ocorrem no processo de emergência do artefato técnico. Obviamente alguns métodos utilizados para este fim em pesquisas etnográficas presenciais, como o exame dos discursos dos agentes humanos envolvidos no processo de emergência do artefato técnico, podem ser aplicados também no ambiente virtual,

mediante a análise dos registros das comunicações dos desenvolvedores em canais de comunicação pela Internet. Contudo, os dados que podem ser obtidos através do sistema de versionamento Git e da plataforma GitHub permitem também chegar até as controvérsias mediante o exame de outras modalidades de traços, característicos do ambiente digital. O escrutínio destes registros através das técnicas apropriados pode facultar a resposta de perguntas como “Que usuário apagou ou modificou o código fonte enviado por outro usuário?” e “Quais propostas de modificação do programa foram aceitas, e quais não foram?”, permitindo talvez entrever como controvérsias se expressam nas práticas cotidianas do desenvolvimento.

A listagem de possibilidades fornecidas pelo exame de registros digitais não se pretende exaustiva, e nem teria como sê-lo. Este artigo não parte do balanço de uma pesquisa já finalizada, mas sim de reflexões a cerca de um experimento em andamento. O teste das possibilidades aqui mencionadas, e o esforço de identificação de outras possibilidades, só poderão ser levados a termo mediante a conclusão de tal experimento, e a efetivação de outros similares.

Cabe ressaltar também que o foco adotado nos processos de desenvolvimento de programas de código aberto em geral, e na plataforma GitHub em particular, não significam que as práticas de pesquisa propostas não possam ser aplicadas em outras instâncias do fenômeno de produção de colaborativa de ciência e tecnologia através da Internet. De fato, embora o desenvolvimento de código aberto e as plataformas que compõem sua infraestrutura constituam um meio bastante apropriado para estas práticas, é certo que elas podem ser utilizadas em alguma medida onde quer que se encontrem registros digitais publicamente acessíveis. Dado que a proliferação de registros constitui atualmente uma característica gerais do ambiente online, e a produção colaborativa através da Internet parece ter como uma de suas condições a publicidade do processo de cooperação, é razoável supor que as condições necessárias para a aplicação dos métodos propostos sejam encontradas em vários contextos análogos ao aqui tratado.

Conclusão

Por muito tempo os antropólogos e demais estudiosos da ciência e tecnologia oriundos das ciências sociais se viram às voltas com o problema do exame de conteúdos científicos, vistos como fatos do mundo natural, e dos objetos tecnológicos, vistos como construtos inteiramente determinados por constrangimentos materiais e técnicos. A natureza e a materialidade pareciam fora do escopo de disciplinas voltadas para o exame das relações sociais e dos processos culturais.

O estudo das práticas realizadas em laboratórios e oficinas forneceu um bom caminho para o tratamento do problema. Através deste expediente, os produtos da ciência e tecnologia puderam ser abordados a partir de seu processo de emergência, antes que fossem reificados e se tornassem fatos e objetos puramente naturais e materiais, dissociando-se do seu contexto de produção. Contudo, este processo ainda assim coloca-se como custoso, na medida em que exige caminhar no sentido contrário do processo de reificação, abrindo caixas-pretas construídas para permanecerem fechadas.

Este artigo trata de uma modalidade de produção de tecnologia que apresenta um quadro bastante distinto do tradicionalmente esboçado nos estudos de ciência e tecnologia. O desenvolvimento colaborativo de programas de código aberto através de plataformas online - e possivelmente também outras modalidades de produção de ciência e tecnologia em laboratórios e oficinas virtuais - tem como um de seus aspectos constituintes a produção de caixas-pretas que podem ser facilmente abertas. O antropólogo que decidir se dedicar a tal objeto de estudo será presenteado pela facilidade de acessar um contexto de emergência de objetos técnicos que não é invisibilizado, e nem trancafiado a sete chaves. Pelo contrário, tal processo de desenvolvimento é facilmente acessível por qualquer sujeito capaz de se conectar à Internet, e encontra-se amplamente historicizado em uma série de registros digitais.

Contudo esta aparente facilidade esconde um desafio. Ao proceder a abertura deste tipo caixa-preta, o antropólogo logo se verá diante de uma série de conteúdos inscritos de forma críptica. Tais conteúdos insinuam conter a resposta para as perguntas que o pesquisador deseja responder, mas rapidamente torna-se evidente que não podem ser acessados e decodificados pelas ferramentas que este carrega consigo. Assoma-se a

necessidade de complementar o ferramental antropológico com outros instrumentos mais adequados à tarefa que se coloca. O presente artigo buscou justamente evidenciar esta necessidade, e indicar alguns dos caminhos pelos quais ela pode ser satisfeita.

Referências:

DABBISH, Laura et al. Social coding in GitHub: transparency and collaboration in an open software repository. In: CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 15, 2012, Bellevue. **Proceedings**. New York: Acm, 2012. p. 1277 - 1286.

ECK, David J. **Introduction to Programming Using Java**. 7. ed. 2014. Disponível em: <<http://math.hws.edu/eck/cs124/downloads/javanotes7-linked.pdf>>. Acesso em: 30 out. 2018.

LATOUR, Bruno; WOOLGAR, Steve. A vida de laboratório: a produção dos fatos científicos. Rio de Janeiro: Relume Dumará, 1997. 310 p. Tradução de: Angela R. Vianna.

LATOUR, Bruno. On actor-network theory: A few clarifications. *Soziale Welt*, v. 47, n. 4, p.369-381, 1996. Disponível em: <<https://www.jstor.org/stable/40878163>>. Acesso em: 30 out. 2018.

_____. **Science in action: how to follow scientists and engineers through society**. Cambridge: Harvard University, 1987. 274 p.

MADEY, Gregory; FREEH, Vincent; TYNAN, Renee. The Open Source software development phenomenon: an analysis based on Social Network Theory. In: Americas Conference On Information Systems, 8., 2002, Dallas. **Proceedings**, Dallas: AMCIS, 2002. p. 1806 - 1813. Disponível em: <<https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1606&context=amcis2002>>. Acesso em: 30 out. 2018.

PERKEL, Jeffrey. Democratic databases: science on GitHub. **Nature**, p. 127-128. 3 out. 2016. Disponível em: <<https://www.nature.com/news/democratic-databases-science-on-github-1.20719>>. Acesso em: 30 out. 2018.

RAYMOND, Eric Steven. **The Cathedral and the Bazaar**. 2000. Disponível em: <<http://www.catb.org/~esr/writings/cathedralbazaar/cathedral-bazaar/index.html>>. Acesso em: 30 out. 2018.

SCACCHI, Walt. Free/Open Source Software Development: Recent Research Results and Methods. In: ZELKOWITZ, Marvin V. (Ed.). **Advances in Computers**. Elsevier, 2007. p. 243-295. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0065245806690050>>. Acesso em: 30 out. 2018.

SEGATA, Jean; RIFIOTIS, Theophilos. Introdução. In: SEGATA, Jean; RIFIOTIS,

Theophilos (Org.). **Políticas etnográficas no campo da cibercultura**. Joinville: Editora Letradágua, 2016. p. 09-20.

SMART, Paul; HEERSMINK, Richard; CLOWES, Robert W.. The Cognitive Ecology of the Internet. In: COWLEY, Stephen J.; VALLÉE-TOURANGEAU, Frédéric (Ed.). **Cognition Beyond the Brain: Computation, Interactivity and Human Artifice**. 2. ed. Springer International Publishing, 2017. p. 251-282.

TUOMI, Ilkka. Internet, Innovation, and Open Source: Actors in the Network. **First Monday**, Chicago, v. 1, n. 6, 08 jan. 2001. Disponível em: <<https://www.firstmonday.org/ojs/index.php/fm/article/view/824/733>>. Acesso em: 30 out. 2018.