Transformações da presença dos jornais portugueses na web (1996-2016):

Correio da Manhã, *Diário de Notícias*, *Expresso* e *Público*

Relatório final de um estudo de caso do projecto "Investiga XXI"

Transformations of the Portuguese newspaper web presence (1996-2016): *Correio da Manhã*, *Diário de Notícias*, *Expresso* and *Público*

Final report of a case study of the project "Investiga XXI"

Diogo Miguel Silva da Cunha (cunhadiogo15@gmail.com)

Bolseiro de Gestão de Ciência e Tecnologia da Fundação para a Ciência e a Tecnologia (SFRH/BGCT/135017/2017)

Responsável científico:

Doutor Daniel Gomes, Arquivo.pt, FFCN-FCT

31 de Julho, 2017

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

FCCN Computação Científica Nacional

Índice

Resumo 3
Palavras-chave 3
Abstract 4
Keywords 4
Agradecimentos 5
1 Introdução 6
2 Problemas de memória no jornalismo 9
3 Problemas metodológicos e desenho de investigação 21
4 Avaliar a preservabilidade dos *websites* actuais dos jornais 42
4.1 Metodologia 42
4.2 Análise de resultados 44
5 Estabelecimento de uma rota de investigação 49
5.1 Metodologia 51
5.2 Análise de resultados 53
6 Análise diacrónica descritiva das páginas principais dos *websites* dos jornais de 1996 a 2016 64
6.1 Metodologia 65
6.2 Análise de resultados 66
7 Conclusões 104
8 Referências bibliográficas 110

Resumo

A transição para uma organização societal na qual a penetração, a influência e mesmo a modelação das tecnologias e dos *media* digitais são crescentes reflectiu-se nos formatos de produção de mensagens jornalísticas e na sua preservação. Investigação recente mostra que as empresas de *media* e os próprios jornalistas não reconhecem a relevância do jornalismo para a memória cultural e não têm suficientes cuidados com os materiais informativos digitalizados e nativos digitais. Neste relatório, são apresentados os resultados de um projecto cujo objectivo foi o de compreender a actual situação portuguesa sobre estes assuntos no âmbito da preservação e utilizar materiais disponíveis no Arquivo.pt, arquivo da web portuguesa, para ampliar o conhecimento sobre o jornalismo português na internet. Procurou-se focar a investigação especificamente nos jornais *Correio da Manhã*, o *Diário de Notícias*, o *Expresso* e o *Público* de 1996 a 2016. Através do cruzamento de conhecimentos adquiridos nas áreas da Filosofia da Tecnologia, dos Estudos Sociais de Ciência e Tecnologia, dos Estudos de Comunicação, Media e Jornalismo, dos Estudos Culturais, dos Estudos da Internet, da Historiografia da Web e do Preservação da Web, desenvolveram e aplicaram-se algumas orientações metodológicas para o estudo desses *websites*. Em termos de resultados concretos, os principais contributos deste projecto consistem numa periodização provisória de vários momentos dos *websites* das versões *online* dos jornais portugueses estudados do ponto de vista de algumas das suas características estéticas e funcionais e numa revisão da direcção transformativa por que, no seu conjunto, estes *websites* passaram. Em geral, os resultados da investigação mostram que com o material disponível e os métodos aplicados se podem abrir caminhos interessantes no âmbito da investigação sobre o jornalismo na era da informação, designadamente para se pensar os modelos de publicação digital e os modos de os investigar.

Palavras-chave

Jornalismo digital, webjornalismo, jornalismo *online*, arquivos da web, Arquivo.pt, preservação de notícias, humanidades digitais.

Abstract

The transition towards a societal organization in which the penetration and influence of numerical representation technologies and digital media was reflected in the production formats of journalistic messages. Recent research shows that news organizations and journalists don't recognize the relevance of their work on memory and don't take sufficient care for digitized and born-digital news materials. In this report, the results of a project whose aim was understanding the current Portuguese situation on those preservation and investigation issues and using the materials available in Arquivo.pt, the Portuguese web archive research infrastructure, for extending our knowledge on the Portuguese journalism in the internet are presented. The research was focused on the newspapers *Correio da Manhã*, *Diário de Notícias*, *Expresso*, and *Público* from 1996 to 2016.Through the cross-fertilization of knowledge acquired in the fields of Philosophy of Technology, Social Studies of Science and Technology, Communication, Media and Journalism Studies, Cultural Studies, Internet Studies, Web Historiography, and Web Archiving, some methodological orientations were developed and applied for the study of those websites. In terms of concrete results, the main contributions of this project consist in a provisional periodization of several moments of the websites of the online versions of the Portuguese newspapers studied from the point of view of some of their aesthetic and functional characteristics and in a revision of the transformative directions through which those websites have passed as a whole. In general, the research results show that with the available material and the applied methods, interesting paths can be found for journalism research in the information age, in particular to think about digital publishing models and ways of investigating them.

Keywords

Digital journalism, webjournalism, online journalism, web archives, Arquivo.pt, news preservation, digital humanities.

Agradecimentos

Gostaria de agradecer à Fundação para a Ciência e Tecnologia pelo financiamento que tornou possível levar a cabo este projecto. Quero também agradecer a todos aqueles que, no âmbito das empresas de *media* estudadas, estiveram disponíveis para responder às minhas questões. Estou em divida para com o Doutor Daniel Gomes pelos seus contributos para a preservação e investigação da web portuguesa, pelo seu conhecimento sobre a internet, pela sua abertura ao diálogo interdisciplinar com as ciências sociais e com as humanidades, pelo seu sentido de serviço público, pela sua disponibilidade e pela sua boa vontade. Deixo também uma palavra de agradecimento a todas as pessoas que conheci na Fundação para a Computação Científica Nacional e que, cada um à sua maneira, contribuíram para que eu pudesse realizar este trabalho – especialmente, ao Engenheiro Daniel Bicho, ao Engenheiro Fernando Melo, ao Engenheiro João Gomes, ao Engenheiro João Nobre, à equipa de secretariado, à equipa de recursos humanos, à equipa de comunicação pública e ao grupo do estúdio virtual. Um agradecimento especial aos meus colegas Diogo Duarte e Ricardo Basílio, que este projecto contribua para criar caminhos para a amizade.

1 Introdução

A transição para uma organização societal na qual a penetração, a influência e mesmo a modelação das tecnologias e dos *media* digitais são crescentes reflectiu-se nos formatos de produção de mensagens jornalísticas e na sua preservação. Num contexto de crescente complexidade do ecossistema de *media*, pequenas e grandes empresas de *media* criaram *websites* de notícias e versões *online* dos jornais de que eram proprietárias, onde colocaram os seus materiais impressos depois de os digitalizarem e materiais nativos digitais. Embora existam projectos internacionais atentos à preservação e ao estudo do conjunto diverso formado por esses materiais, investigações recentes mostram que as empresas de *media* e os jornalistas não só não reconhecem a relevância de seu trabalho para a memória cultural (Zelizer, 2008), como não têm suficiente atenção e cuidados para com esses materiais e respectiva preservação (Carner, McCain, & Zarndt, 2014; Presnell & Morris, 2017).

Neste relatório, apresentam-se os resultados de um projecto cujo objetivo foi o de entender a actual situação portuguesa sobre os problemas de preservação e investigação e de utilizar os materiais disponíveis no Arquivo.pt, arquivo da web portuguesa, para enriquecer o conhecimento acerca do jornalismo português na e para internet. As perguntas de que se partiu foram as que se seguem. O que caracteriza as culturas e as políticas de preservação dos *websites* dos jornais portugueses? Como mudaram efectivamente os *websites* dos jornais portugueses ao longo do tempo? E que implicações culturais e significados sociais têm tais culturas, políticas e mudanças? Para responder a essas perguntas, desenvolveu-se e aplicaram-se novas estratégias metodológicas resultantes do cruzamento de contributos das áreas da Filosofia da Tecnologia, dos Estudos Sociais de Ciência e Tecnologia, dos Estudos de Comunicação, Media e Jornalismo, dos Estudos Culturais, dos Estudos da Internet, da Historiografia da Web e da Preservação da Web. Concretamente, estabeleceu-se como meta a avaliação da preservabilidade dos *websites* em questão, descrição da sua mudança estrutural, a sugestão de tendências de desenvolvimento específicas e o confronto com periodizações provenientes de investigações prévias.

Uma vez que o Arquivo.pt permite a pesquisa apenas a partir de 1996 e tem um período de embargo correspondente a um ano, delimitou-se, logo à partida, o material a estudar, ao período que vai de 1996 até 2016. E, dada a quantidade e diversidade de jornais portugueses com *websites*, seleccionaram-se apenas alguns jornais utilizando como critérios a longevidade dos *websites*, a constância das suas actualizações, o significado nacional dos jornais (edições e circulação), a proximidade geográfica das suas infraestruturas e a diversidade da propriedade. Os jornais selecionados foram os seguintes:

- *Correio da Manhã*: abreviado como CM, é um jornal diário português estabelecido em 1979, propriedade da empresa de *media* portuguesa Cofina e com sede em Lisboa, publicado em formato tabloide, líder habitual de vendas entre os diários, o seu primeiro endereço na web foi registado em 1995, e o seu *website* permanece *online*, sendo actualizado recorrentemente;
- *Diário de Notícias*: abreviado como DN, é um jornal português estabelecido em 1864, propriedade da empresa de *media* portuguesa Global Media Group (anteriormente Controlinveste) e com sede em Lisboa, publicado em formato tabloide, é geralmente classificado como um jornal de qualidade, teve o seu primeiro endereço na web registado em 1995, e o seu *website* permanece *online*, sendo actualizado regularmente;
- *Expresso*: é um jornal semanal português estabelecido em 1973, propriedade da empresa de *media* portuguesa Impresa e com sede em Lisboa, publicado em formato berliner, geralmente é classificado como um jornal de qualidade e é o líder habitual de vendas entre os semanários, o seu primeiro endereço na web foi registado em 1997, e o seu *website* permanece *online*, sendo atualizado com regularidade;
- *Público*: é um jornal português estabelecido em 1990, propriedade da empresa de retalho portuguesa Sonae e com sede em Lisboa, publicado em formato tabloide, geralmente classificado como um jornal de qualidade, teve o seu primeiro endereço na web registado em 1995, e o seu *website* permanece *online*, sendo atualizado regularmente.

Este estudo de caso faz parte de um projecto de investigação maior na área de Comunicação e Gestão de Ciência intitulado “Investiga XXI”, que é, por sua vez, uma iniciativa do Arquivo.pt, um projecto liderado pela Fundação para Computação Científica Nacional, unidade da Fundação para a Ciência e Tecnologia. Os principais objectivos do “Investiga XXI” eram os de reunir investigadores de várias áreas científicas, providenciar formação no contexto das Tecnologias da Informação e da Preservação da Web e fomentar a utilização dos materiais do Arquivo.pt nas suas investigações (Arquivo.pt, 2016).

Quando o projecto de investigação sobre as transformações dos *websites* de jornais foi proposto, o propósito era apenas o de usar os materiais do Arquivo.pt para sustentar uma análise diacrónica. No entanto, depois da formação em Preservação da Web, após as discussões dos estudos de caso aprovados no âmbito do “Investiga XXI”, depois da revisão da literatura realizada e posteriormente ao confronto informado com o material arquivado, tornou-se claro que tal projecto de investigação implicaria uma maior compreensão dos problemas e das metodologias em jogo. Assim sendo, começa por se apresentar neste relatório um enquadramento histórico e teórico dos problemas e das definições em causa, seguindo-se uma discussão metodológica constitutiva do desenho de investigação e os três principais passos da investigação concreta realizada, primeiro a avaliação da preservabilidade das versões *online* dos jornais estudados, segundo o estabelecimento de uma rota de investigação para o prossecução efectiva da análise dos materiais, e terceiro a análise diacrónica elaborada a partir de uma análise de imagens. Este relatório procura reflectir a revisão de literatura contínua na consecução deste projecto, os dados empíricos considerados, a sua análise e crítica e também considerações obre a própria experiência acerca das questões que emergiram na realização do projecto. Por consequência, são de duas ordens as conclusões que se tiram. Relativamente aos dados concretos estudados, podemos seguir algumas linhas mais ou menos estáveis de desenvolvimento com expressão tecnológica, especificamente sobre os aspectos estético-funcionais em mudança. Relativamente à obtenção dos dados, podemos levantar questões acerca das características dos arquivos da web e de uma série de questões emergentes na discussão de tais assuntos.

2 Problemas de memória no jornalismo

O jornalismo e os arquivos são instituições sociais modernas. Embora o jornalismo tenha antecedentes culturais, sociais e tecnológicos na Antiguidade Clássica e no Renascimento (Pizarroso Quintero, 1996) e ainda que diferentes história e culturas nacionais possam estar na base de modelos diferenciados de jornalismo (Chalaby, 2003; Schudson, 2007; Williams, 2006), o jornalismo com a função comunicacional de fornecer às pessoas a informação necessária para orientarem e coordenarem a sua vida em sociedade emergiu no contexto das transformações sociais que, como as grandes migrações e a urbanização, enformam a industrialização e a profissionalização da actividade jornalística, processo que se cristalizou no século XIX (Carey, 1977; Fenton, 2010; Fidalgo, 2006; Garcia, 2009a; Garcia, 2009b; Hartley, 1996; Kovach & Rosenstiel, 2004; Ruellan, 1993; Schudson, 2003). Com a hegemonia do modelo americano de um jornalismo baseado em factos (Williams, 2006), o jornalismo comunicacional-informativo aparece como uma actividade profissional numa sociedade de massa, com um modo de produção industrial e com meios de comunicação de massa. E a notícia e a reportagem tornam-se os géneros recorrentes do jornalismo profissional.

Algo semelhante pode ser dito sobre os arquivos. Nas chamadas sociedades tradicionais, a memória é um fenómeno de síntese temporal de ordem neurofisiológica, social e cultural (Assmann, 2008). A memória fica situada aos níveis do cérebro, da linguagem e de objectos e práticas mnemónicas. Neste contexto, a memória segue aquilo a que Pierre Nora chamou a dialéctica dinâmica da lembrança e do esquecimento (Nora, 1989). A aceleração temporal moderna criou uma necessidade de instituições próprias para a promoção do não-esquecimento. Tudo deveria ser arquivado. Essa é a lógica moderna do arquivo. Conceitos e manuais, museus e arquivos são alguns dos *loci* de cristalização da lembrança a que Nora deu o nome de "lugares de memória" (*ibid*.).

Embora a divisão do trabalho entre jornalistas e historiadores reenvie o primeiro para um trabalho centrado no presente, o jornalismo enquanto tal tem um papel fundamental na manutenção da memória. Barbie Zelizer considera mesmo que o

jornalismo é "uma instituição de registo mnemónico", "uma das principais instituições de recordação e de lembrança da sociedade contemporânea" (Zelizer, 2008: 85). A relação do jornalismo com a memória tem pelo menos três dimensões importantes:

- o envolvimento dos jornalistas com o passado, porque estes fazem uso da sua experiência como profissionais, mas também como pessoas em geral, para tomarem decisões em termos das suas narrativas e porque utilizam eventos considerados como parte do passado para enquadrar questões consideradas actuais (como no uso de analogias históricas) ((Zelizer, 2008);
- o envolvimento dos jornalistas com o presente, pois é como trabalhadores do presente que a divisão do trabalho os permite definir, mas nesse trabalho sobre o presente os jornalistas procuram elaborar registos que armazenam a memória de diferentes períodos concebidos como presente, o que associa o jornalismo a toda a vida quotidiana (Kitch, 2008), facto muito evidente na mobilização de imagens, sobretudo de tipo fotográfico, para dar testemunho de eventos (Tirohl, 2000; Zelizer, 2012);
- e a espécie de "fórum" formada pelo trabalho jornalístico, onde se cruzam as memórias das pessoas, das instituições e das organizações da sociedade, colocando-se assim dentro do próprio processo de recordação da vida social e cultural (Kitch, 2008).

Este triplo trabalho de manutenção da memória realizado pelo jornalismo integra a função comunicacional moderna do jornalismo. Este trabalho é comunicacionalmente relevante porque as conexões entre eventos históricos e os registos de vários presentes possibilitado pelo trabalho jornalístico são conexões de comunicação e registos de comunicação, quer dizer, são formas de uma sociedade num certo período se orientar e coordenar.

A sociedade contemporânea é confrontada com um fenómeno que tem vindo a ser entendido como uma crise generalizada das instituições modernas com consequências para o jornalismo e a memória. As tecnologias e os *media* digitais estão num lugar especial nas narrativas sobre essa crise, particularmente numa convergência entre os discursos de hiperbolização ou abandono dos pilares da modernidade e a ideia de uma sociedade de comunicação e informação (Vattimo, 1992; Finnemann, 2002). No

decorrer do século XX, a confluência entre as transformações tecnológicas nas tecnologias de informação, nas telecomunicações e nos meios de comunicação e o processo conhecido como "globalização" conferem à comunicação, à informação, aos meios de comunicação e às redes uma relevância societal singular, que se reflete, além do mais, na institucionalização do campo dos Estudos de Comunicação, Media e Jornalismo nos Estados Unidos da América (Proulx, 2001). Nestes processos, o nexo causal é tão complexo que não se trata apenas de desenvolvimentos tecnológicos darem lugar a um certo modo de ser da sociedade. Na verdade, o próprio modo como as sociedades começam a projectar futuros possíveis ligados à informação tem um impacto tremendo na orientação do desenvolvimento social (Sousa, 2006).

Depois de Jon Katz ter escrito em 1994 que os jornais desapareceriam em 10 anos, as empresas de *media* apressaram-se em criar lugares para publicarem os seus materiais noticiosos *online* (Nerone & Barnhurst, 2001). As tecnologias e os *media* digitais estão na base de um processo no qual os jornalistas se envolvem na especialização de um jornalismo *na* e *para* a internet. Esta especialização traduz uma gigantesca mudança nas empresas de *media* e na actividade dos jornalistas, com uma reorganização do fluxo de trabalho da redacção, das práticas profissionais, da cultura organizacional, dos modelos de negócios (especialmente as relações com o mercado publicitário) e dos próprios géneros e formatos jornalísticos (Boczkowski, 1999; Boczkowski, 2004; Pavlik, 2000; Ward, 2002; Kawamoto, 2003).

Num primeiro momento, as empresas de *media* reproduzem na internet os conteúdos das suas publicações originais. Durante a década de 1990, há um modelo de publicação "fac-símile" ou "shovelware", ou seja, os *websites* apresentam os textos dos jornais impressos, transmitem as ondas das rádios ou as notícias da televisão (González, 2000; Pavlik, 2001). Numa fase posterior, o modelo de publicação torna-se "adaptado", o *layout* é adequado ao próprio *website* e são introduzidas hiperligações e peças multimédia, mas os conteúdos são semelhantes aos originais dos outros meios (*ibid*.). No final da década de 1990, dá-se uma grande ruptura com o modelo do shovelware, adoptando-se a abordagem "web-first" ou "online first" para o fluxo de trabalho da redacção, ou seja, a produção de informação jornalística fica centrada nos *websites*, que devem ser actualizados constantemente durante vinte e quatro horas por dia

(Grabowicz, 2014). Neste contexto começa a surgir uma nova fase do jornalismo na e para a internet, a do modelo de publicação "digital", no âmbito do qual os *websites* das publicações começam a ter layouts próprios, utilizam-se links, os utilizadores podem comentar, e as notícias de última hora têm um papel central (González, 2000). Finalmente, as publicações procurariam utilizar as principais características da web para elaboração das narrativas jornalísticas, abrindo-se um "modelo multimédia" de publicação (González, 2000; Pavlik, 2001). Hoje, encontramos na internet publicações enquadráveis em todos estes diferentes modelos de publicação, o que permite argumentar que estas fases não são sucessivas, mas coexistentes no tempo (Zamith, 2013).

No contexto português, o modelo de shovelware dissemina-se durante uma primeira fase do jornalismo na e para a internet, uma fase que decorre entre 1995 e 1998 e a que Helder Bastos dá o nome de "implementação" (Bastos, 2010; Bastos, 2011). Trata-se de um período no qual as empresas de *media* existentes (por vezes, designadas por *legacy media*) criam os *websites* para as suas publicações. É neste período que os jornais aqui estudados registam os seus primeiros endereços na web: o publico.pt em Maio de 1995 (começando a edição do *Público* a ser colocada na web em Setembro desse ano), o correiomanha.pt em Dezembro de 1995 (embora a edição do CM apenas seja colocada na web em Março de 1998), o dn.pt em Dezembro de 1995 (sem referência a data de colocação da edição na web na bibliografia consultada), e o expresso.pt em Janeiro de 1997 (tendo a edição do *Expresso* começado a ser posta *online* em Julho de 1997). A partir de 1999 começa uma fase de "expansão" na qual os jornais portugueses, em concordância com o que se passaria no resto da Europa e nos Estados Unidos da América, aparecem exclusivamente *online* e as empresas de *media* apostam no reforço dos seus segmentos *online* (*ibid*.). Por exemplo, nesse ano o *Público* anunciou que a área digital de negócios se tornaria uma área com produtos específicos e com serviços pagos. Sublinhe-se que entre os serviços propostos estava a possibilidade de consulta de um arquivo digital do jornal. Segue-se logo a partir de 2001, ainda seguindo a proposta de Bastos, uma fase de "depressão" durante a qual muitos *websites* são fechados e há muitas reduções de custos, que se refletem em muitos níveis (despedimentos, qualidade dos conteúdos, etc.) (*ibid*.). As empresas de *media* existentes, como aquelas

que são proprietárias dos jornais aqui estudados, continuariam a renovar os *websites* que já possuíam ou a lançar e relançar novos *websites* ao longo do tempo. Contudo, a informação sobre estes processos ainda é muito reduzida e encontra-se dispersa por documentos de natureza diversa. Num estudo recente sobre o *Expresso*, procurou-se construir uma periodização destes processos no caso desse jornal recorrendo não só a investigações científicas prévias como também a publicações jornalísticas com referências de datas e aos relatórios societários da respectiva empresa proprietária (Cunha, no prelo).

No quadro de um novo ecossistema de *media*, o jornalismo parece encontrar nas tecnologias novas possibilidades, particularmente devido às características estruturais da internet e da web. As características mais salientes nas discussões no âmbito dos estudos sobre o jornalismo na e para a internet são a interacção entre humano e máquina a que se dá o nome de "interactividade", a combinação textual através de hiperligações a que se chama "hipertextualidade", a associação de meios e de formatos designada como "multimedialidade", a actualização constante e simultânea a que se pode chamar "instantaneidade" e a adaptação de conteúdos e da sua distribuição e apresentação aos hábitos de consumo dos utilizados designada por "personalização" (*vide* Zamith, 2013). Uma outra característica fundamental é aquilo a que no âmbito desses estudos se tem dado o nome de "memória", entendida no sentido da acumulação de conteúdos jornalísticos ao longo do tempo. O uso da noção de memória neste sentido não é uma completa novidade, afinal jornais e revistas são guardados há séculos em arquivos e bibliotecas, e mesmo as cassetes com emissões radiofónicas e televisivas são guardadas há décadas em arquivos dos jornais e em algumas bibliotecas. No entanto, há pelo menos três elementos de novidade: nos *websites* dos jornais os conteúdos podem ser acumulados na sua diversidade mediática, há muito mais material a preservar e a ser possivelmente preservado e podem aceder-se a muitas mais informações (Ward, 2002). Um outra novidade mecionávelé o acesso instantâneo a essas informações. Em síntese, na era informacional a relação do jornalismo com a memória promete ser incorporada em formas digitais que permitem que várias pessoas, em qualquer momento, através de vários meios, acedam a trabalhos jornalísticos publicados nas mais diversas alturas. Como diz Marcos Palácios, tratar-se-ia de uma

"memória múltipla, instantânea e cumulativa" (Palácios, 2003: 84). Neste quadro, a ideia de "arquivo" é a de um banco de dados que armazena o material jornalístico e que permite a sua consulta (por parte de jornalistas para elaborarem novos trabalhos e por parte de leitores para confrontarem trabalhos mais antigos) e o reenvio através de hiperligações de outras peças. Ou seja, a relação entre jornalismo e memória acaba por ser reenquadrada nas próprias características da web.

Ainda que as tecnologias e os *media* digitais pareçam trazer ao jornalismo um grande conjunto de novas possibilidades, inclusivamente os referidos aspectos da sua relação com a memória, também emergem novos problemas sociais. A reflexão sobre o tempo na sociedade actual vê-se logo à partida confrontada com a questão do esquecimento da força de transformação social das dinâmicas estruturais de longo termo, gerando um conjunto de problemas a que se pode dar o nome de "presenteísmo", como as visões de "redução de processos a estados" (Elias, 2008) e de "cesurismo" (Martins, 1996), que conduzem à negligência sobre a continuidade e a cumulatividade dos processos sócio-históricos. Mas a memória e o arquivo, em geral, enfrentam pelo menos dois principais problemas relacionados com a própria natureza da web e com a postura dos utilizadores e de diversas instituições no que diz respeito à preservação. O primeiro problema é o da "fragmentação" ou "deslocalização" ou, ainda, "disseminação" da noção de "arquivo". Segundo Mike Featherstone, a ideia moderna de arquivo está a ser desafiada pelas tecnologias digitais por várias razões. Uma delas diz respeito à transição de uma ideia localizada de arquivo, como um local físico, para um arquivo fluido, processual e dinâmico, como um lugar virtual (Featherstone, 2006). Dá-se uma reorganização espaciotemporal que arrasta o arquivo para o espaço abstracto do ciberespaço, colocando-o sob os regimes de controlo cibernético (Miranda, 1996). A própria categoria de "lugar" torna-se paradoxal quando se fala de arquivo, visto que essas tecnologias o deslocalizam, o disseminam através da experiência (*ibid*.).

Para Featherstone e para Arjun Appadurai, algumas obras da Filosofia francesa contemporânea reflectem a existência de problemas conceptuais que afloram com essa fragmentação do arquivo (Appadurai, 2003; Featherstone, 2000; Featherstone, 2006). Featherstone refere-se a Michel Foucault e Jacques Derrida. Appadurai refere-se apenas a Foucault. O que une as contribuições desses filósofos relaciona-se essencialmente com

uma ideia do arquivo como um princípio organizador da vida moderna. Walter Benjamin (2000) já havia mostrado a fragmentação da ideia de arquivar nas ruas da cidade e nos seus objectos quotidianos, destacando aquilo a que Featherstone dá o nome de "arquitectónica do processo de arquivamento" (Featherstone, 2006: 595). Derrida também o fez, à sua maneira, num famoso texto sobre aquilo que designou como o *mal d'archive* (Derrida, 1995). Usando o modelo psicanalítico freudiano como ponto de partida, Derrida encontrou nessa febre do arquivamento um impulso paradoxal tanto para a repetição quanto para a destruição, dispondo tanto de pulsões de vida como de morte. A concepção de *mal d'archive* sintetiza-se na ideia de que o modelo arquivístico de preservação incorpora necessariamente a noção de destruição. Ou seja, é a finitude radical das coisas que alimenta o desejo de arquivar. Preserva tudo antes de que tudo desapareça. Lembra-te antes de que te esqueças. Como diz Michel Melot, o arquivo legitima uma sociedade que destrói cada vez mais seus objectos (*apud* Miranda, 1996). Numa outra óptica, Appadurai considera que as possibilidades colocadas pelas tecnologias de envio de textos, fotografias e sons e pela existência de *websites* pessoais, *websites* de armazenamento e arquivos digitais fazem com que a ideia de arquivo se esquive à lógica do Estado e esteja mais próximo da memória popular (Appadurai, 2003). Como se disse, Foucault é um nome importante no caso destas novas concepções de arquivo. Para Featherstone, Foucault reformulou o arquivo deslocando o seu significado de uma estrutura unificada de acesso ao passado para uma dimensão complexa e fluída de reconfiguração desse passado (Featherstone, 2000; Featherstone, 2006). Para Appadurai, Foucault levou o arquivo para além da função de preservação de vestígios de memória colectiva e da circunscrição do Estado-nação (Appadurai, 2003). De acordo com Appadurai, para Foucault, o arquivo tem uma função produtiva antecipadora da memória coletiva e funciona como ferramenta de policiamento, vigilância e governação (*ibid*.). Com a sua "arqueologia", Foucault procurou descrever o discurso no sistema de arquivo do qual faz parte. Embora este trabalho de busca no arquivo não corresponda, para Foucault, à busca de um começo, ele está ainda relacionado com a memória, na medida em que para realizar a descrição arqueológica seria necessário demarcar aquilo a que Foucault dá o nome de "campo enunciativo", um "domínio da memória" composto por enunciados que deixaram de ser admitidos ou discutidos e que deixaram de definir verdades e validades, mas em relação aos quais são estabelecidas ligações de

filiação, génese e transformação (Foucault, 2014: 170). Em Foucault, o próprio arquivo está sempre em torno desse campo. Ou seja, não é uma instituição física, mas antes o "sistema geral de formação e transformação de enunciados", a "lei do que pode ser dito, o sistema que governa o surgimento de declarações como eventos singulares" e é também o que faz com que todas as coisas sejam ditas e compostas ou desaparecem num certo conjunto (*idem*: 177-179). Em última análise, com Foucault o arquivo fica separado da ideia física de um lugar de memória e passa a fluir em relações discursivas (do que se diz, no modo como se diz e no poder expresso no que é dito e no que é interdito), o que se torna interessante na medida em que as tecnologias e os *media* digitais destacam-se de outras tecnologias, entre outros aspectos, pela capacidade de uma certa "dilatação" do discurso e mesmo do discurso informativo (Mancini, 2013).

O segundo problema que a lógica do arquivo enfrenta no mundo contemporâneo é a "dinâmica" e a "efemeridade" da web, que se traduziu num desaparecimento progressivamente acelerado de grande parte do conteúdo publicado na web. Os *websites* mais variados são desactivados devido a uma grande variedade de causas, como a finalização de projectos que estão na base da sua génese, a falta de financiamento para a sua manutenção, etc. Por vezes, certos *websites* não são desactivados, mas são totalmente modificados, também por conta de uma grande variedade de causas, como a venda de domínios, a mudança de negócios, etc. No período de um ano, cerca de 85% das páginas disponíveis na internet desaparecerão ou serão alteradas (Gomes & Silva, 2006). Vinton Cerf, um dos pioneiros da internet, afirmou que mesmo os documentos e as imagens guardados em computadores podem desaparecer devido à obsolescência de *hardware* e *software*, não deixando registo do século XXI, o que tornaria a era informacional uma espécie de "Idade das Trevas Digital" (Ghosh, 2015). Em suma, na web é publicada uma quantidade imensa de dados que não é preservada e que, consequentemente, não pode ser recuperada e consultada por quem por ela tiver algum interesse, o que se traduz também numa perda de informação sobre o presente para as gerações futuras (que entenderão este presente como passado). Abre-se uma "lacuna de conhecimento sobre o presente para as gerações futuras" (Costa, Gomes & Silva, 2016: 1).

A esfera do jornalismo enfrenta dificuldades semelhantes, ainda que tenha as suas próprias especificidades. Para começar, o jornalismo, tanto como uma forma social de conhecimento quanto como um actividade profissional, vê-se atravessado pela narrativa da crise estrutural internacional há décadas. Algumas das tendências dessa crise seriam as seguintes: muitos jornais têm fechado ou têm-se dedicado exclusivamente a versões digitais, há uma redução na circulação e leitura de publicações, há uma falta de definição de modelos de negócio, a mediatização de acontecimento já não é exclusiva dos jornalistas, algumas fontes impõem-se no espaço mediatizado, há uma falta de auto-regulação por parte da classe jornalística, há muitas dificuldades no acesso à profissão, há uma precarização geral da profissão, impõem-se lógicas produtivas orientadas para o lucro das empresas, etc. Estas tendências têm conduzido à reorganização das lógicas de produção, difusão e consumo de informação. Nesse contexto têm-se salientado vários problemas, como a garantia de autonomia e liberdade dos jornalistas na aplicação dos seus critérios de selecção e redacção de notícias, a garantia de separação entre conteúdos informativos e conteúdos publicitários e a garantia de objectividade e mediação jornalísticas e, portanto, da própria base que legitima eticamente a profissão. Muitas dessas tendências e destes problemas têm sido observados em diversos trabalhos internacionais. Deve salientar-se que a crise que emerge no campo do jornalismo não deve ser reduzida a uma mudança de instrumentos e consequente adaptação económica ou vice-versa. Como argumenta Jeffrey Alexander, há todo um quadro cultural de codificação das tecnologias e das relações económicas que transformam as mudanças de tipo material numa profunda crise do jornalismo, uma crise compreensível em termos da profissão, do mercado e da sociedade em geral (Alexander, 2015). A singularidade das tecnologias digitais nesta crise diz respeito à já referida dilatação dos domínios da informação (Mancini, 2013), no âmbito dos quais o jornalismo era pivô, e também à associação da actividade jornalística a uma indústria dos *media* virada para o mercado digital de informação e consequente alteração ao nível dos valores que alimentam o jornalismo, colocando-o muitas vezes ao serviço do lucro das empresas de *media* (Cohen, 2002; Mateo et al., 2010; McChesney, 2013; McManus, 1994; Scott, 2005).

Tendências muito semelhantes têm sido encontradas no âmbito do jornalismo em Portugal. Dois extensos trabalhos de recolha de depoimentos dos jornalistas portugueses mostram que os profissionais reconhecem vários dos problemas que perpassam a transformação do campo dos *media* em Portugal, as empresas de *media* e mais especificamente os jornais, os jornalistas e o jornalismo (Lopes, 2015; Rebelo, 2011). Apesar da existência de questões nacionais, regionais e económicas específicas, várias investigações têm dado relevância às questões levantadas pelas tecnologias e pelos *media* digitais, pelas práticas que tornam possíveis, mas também pelos constrangimentos que provocam a diversos níveis, particularmente pelas transformações tecnoculturais, pelo emergente quadro do novo capitalismo, digital e imaterial, e pelo crescendo correspondente de problemas de ordem ética e deontológica (Camponez, 2004; Fidalgo, 2016; Garcia, 2009a; Garcia, 2009b; Garcia, Marmeleira & Matos, 2014; Gomes, 2012; Matos, Baptista & Subtil, 2017; Sá, Peixinho & Camponez, 2012). José Nuno Matos, Carla Baptista e Filipa Subtil escreveram muito recentemente que "Embora não tenha sido imune à regressão da economia, a actual crise do jornalismo é anterior a 2008. A sua origem é inseparável da emergência do suporte digital e da consequente possibilidade de acesso gratuito à informação ou da dispensabilidade dos jornais como o meio primordial de comunicação com o público" (Matos, Baptista & Subtil, 2017: 10).

No âmbito da revisão de literatura realizada, o que parece ser mais interessante sublinhar no contexto geral dos problemas levantados é a existência de uma intersecção entre duas crises apanhadas na crise da modernidade e das suas instituições: a crise no jornalismo e a crise da memória. Tanto o jornalismo como a memória nas suas formas modernas vêem-se envoltos no contexto tardo-moderno de uma sociedade informacional cujas lógicas produtivas, os meios e s discursos subjacentes os introduzem numa outra crise. A intersecção entre a crise no jornalismo e a crise da memória gera, pois, a possibilidade de falarmos de um contexto particular de problemas, o qual ainda está pouco documentado. Trata-se do contexto dos problemas de memória no jornalismo, resultante, então, da intersecção, por um lado, de um conjunto de fenómenos entendidos como uma crise mundial e estrutural, muito documentada, de jornalismo no contexto da informatização progressiva das indústrias de *media* e, por

outro lado, de um outro conjunto de fenómenos também entendidos como crise – também ela global e cada vez mais estudada, pelo menos nas áreas disciplinares de Engenharia Informática, Ciências da Informação, Historiografia da Web e Preservação da Web – na memória cultural, no quadro do processo contemporâneo generalizado de mudança tecnocultural.

Logo à partida, como nota Carlos Camponez, as novas condições em que o jornalismo opera reflectem-se numa perda de memória através da "juvenilização" das redações (Camponez, 2012), o que em Portugal parece ser o cenário dominante, dada a tendência marcante de "rejuvenescimento" da profissão nas redacções dos jornais portugueses (Garcia, 2009b). Perde-se a memória no sentido de que se desconhecem os contextos para colocar os acontecimentos em perspectiva assegurando a qualidade informativa. Mas o problema é mais amplo e mais complexo. Devido à relação problemática entre acesso e preservação no que concerne aos jornais serem fontes primárias de acesso a informação, os jornais enfrentam aquilo a que Jenny Presnell e Sara Morris recentemente chamaram uma "crise histórica dos jornais" (Presnell & Morris, 2017). Ainda que os segmentos digitais das empresas de *media* e os suportes digitais, com diversos formatos de publicação possíveis, dominem a industria dos *media* e o jornalismo, parece não existir uma preocupação com a preservação dos materiais digitais. Num trabalho recente, Dorothy Carner, Edward McCain e Frederick Zarndt conduziram uma investigação sobre a preservação das notícias digitais com uma componente empírica interessante (Carner, McCain & Zarndt, 2014). A principal conclusão dos autores é clara relativamente ao modo como os jornalistas entendem a sua missão: "A missão principal dos produtores de notícias é a de criar notícias, não a de as preservar" (*idem*: 19). Carner, McCain e Zarndt referem-se a uma "lacuna de preservação", no sentido de um período temporal no qual os materiais digitais desaparecem porque não foram preservados por alguma organização (*ibid*.). Para além desta dificuldade, a recolha e preservação dos materiais digitais depara-se ainda com problemas como a complexidade cultural e jurídica de preservação de materiais nas redes sociais e nos agregadores de notícias, a ausência de depósito digital legal, o bloqueio (intencional ou não) de agentes virtuais de recolha de materiais digitais, a desconfiança das empresas de *media* em relação às organizações que efectuam a

recolha dos materiais digitais, a muito acelerada actualização dos *websites* noticiosos e mesmo a fragilidade das estruturas físicas (como discos rígidos) onde se fazem *backups* dos materiais digitais. Sendo que a web é hoje um domínio fundamental para a história das notícias, porque ""providencia um recurso crítico para o exame da natureza da mudança dos jornais" e porque fornece um "registo para compreender o que realmente existia", o "sentido de mudanças prévias" e o "contexto para entender a mudança futura" (Weber, 2017: 99), torna-se urgente compreendermos o cenário com que nos deparamos no jornalismo contemporâneo.

Os problemas de memória no jornalismo são um fenómeno social que merece a atenção dos investigadores da área, principalmente porque conduzem ao levantamento de questões acerca da execução de uma das tarefas mais importantes no contexto da função comunicacional-informativa do jornalismo, a de deixar um vestígio histórico das culturas particulares em que está inscrito. Se, como nos habituou o modelo americano, o jornalismo pode colocar pessoas espacialmente dispersas em conversa, não se deverá manter o mesmo sentido de conversa ao longo do tempo e não apenas do espaço? Se o jornalismo tem como função fornecer às pessoas a informação para se orientarem e coordenarem, não será também relevante ressalvar o valor histórico dessa informação? Não se poderá argumentar no sentido de que aqueles que vivem hoje e as gerações futuras poderão precisar daquilo que hoje se poderia considerar como presente para se orientarem e coordenarem no seu próprio tempo, ainda por vir? É precisamente uma ideia semelhante que pode ter fundamentado a existência de espaços próprios para jornais nas bibliotecas e nos arquivos públicos e que terá estado na origem da invenção das hemerotecas. Se assim for, poderemos estar perante uma certa urgência de repensar a situação digital dos jornais na componente da sua preservação. O projecto conjunto entre o Donald W. Reynolds Journalism Institute's Journalism Digital News Archive e o Internet Archive sinaliza a relevância de tal ser feito. Num conjunto de conferências designadas como Dodging the Memory Hole, esses arquivos têm-se debruçado sobre o acesso e a preservação dos materiais jornalísticos digitais. No âmbito dessas conferências afirmam os organizadores que a perda desses materiais é "a ameaça mais urgente da memória cultural hoje" (RJI, 2017). Face a isto, cabe-nos agora procurar compreender a situação portuguesa.

3 Problemas metodológicos e desenho de investigação

Na secção anterior procurou mostrar-se como as tecnologias e os *media* digitais trazem à relação entre jornalismo e memória novas possibilidades, mas também novos problemas de diversas ordens (culturais, económicos, políticos, sociais, tecnológicos), os quais – no seu conjunto – nos rementem para uma relação problemática do jornalismo com a memória. Contudo, ao contrário do que poderá parecer à primeira vista, o material jornalístico digital não está completamente perdido. Isto deve-se a um progressivo reconhecimento público e privado da importância de preservação e reprodução de páginas web que poderiam desaparecer rapidamente. É significativo que desde 2003 a UNESCO defenda a relevância do digital como património cultural. Num documento sobre a preservação de tal património, a UNESCO considera o material digital como património com especificidades técnicas que "abrange recursos culturais, educacionais, científicos e administrativos, bem como informações técnicas, legais, médicas e outros tipos de informações criadas digitalmente ou convertidas em forma digital a partir de recursos analógicos existentes" (UNESCO, 2004: 75). Pensando em textos, bases de dados, ficheiros de som, imagem e vídeo e páginas web inteiras, a UNESCO considera que o material digital corre riscos de se perder devido a vários factores, como a efemeridade dos dados, a rápida mudança de *hardware* e de *software* e lacunas legislativas sobre o assunto, de tal modo que se torna urgente garantir a proteção e a preservação do material digital como património. Também em 2003, foi estabelecido na Bibliothèque nationale de France o International Internet Preservation Consortium (IIPC), cuja missão seria a de "adquirir, preservar e disponibilizar conhecimentos e informações acessíveis da Internet para as gerações futuras de todos os lugares, promovendo a troca global e as relações internacionais" (IIPC, s.d.).

No contexto de crescente reconhecimento de preservação dos materiais digitais, têm surgido iniciativas de preservação bastante diversas. Um pouco por todo o mundo, incluindo em Portugal, constituíram-se colecções em arquivos digitais associados a bibliotecas e hemerotecas com o objectivo de digitalizar diversos tipos de documentos, incluindo jornais. Um dos mais fortes programas de digitalização de jornais impressos foi criado também em 2003 nos Estados Unidos da América. Trata-se do National Digital Newspaper Program, uma parceria entre o National Endowment for the Humanities e a

Library of Congress com o objectivo de providenciar acesso digital a jornais impressos com relevância histórica (Library of Congress, s.d.). Em Portugal, foi criada em 2005 a Hemeroteca Digital como uma iniciativa da Hemeroteca Municipal de Lisboa com o objectivo de contruir uma "biblioteca digital de jornais e revistas caídos em domínio público", tratando-se especificamente de "títulos de publicações periódicas, com destaque para as colecções digitais de periódicos do fundo local e histórico, completadas com fichas históricas de apresentação dos jornais e revistas, raridades bibliográficas relacionadas com a imprensa escrita, e bibliografia de referência para o estudo e consulta do acervo bibliográfico da HML [Hemeroteca Municipal de Lisboa]" (Hemeroteca, s.d.). No mesmo contexto, têm também surgido uma série de directrizes para a preservação do material jornalístico digital, de que é um exemplo interessante o trabalho levado a cabo por Katherine Skinner e Matt Schultz (2004).

Os arquivos da web são um tipo específico de iniciativa de preservação de material digital, distinguindo-se de outras iniciativas principalmente pela capacidade de irem para além da digitalização de material físico produzido por meios analógicos, preservando e reproduzindo páginas web inteiras e uma gama alargada de materiais que seriam nativos digitais. Os arquivos da web podem ser entendidos como "um tipo especial de bibliotecas digitais", porque incluem documentos impressos digitalizados e também diferentes tipos de ficheiros audiovisuais e multimédia, sendo que a maior diferença dos arquivos da web seria estaria no facto de estes arquivos terem uma quantidade de dados tal que "excede a organização e o gerenciamento tradicionais das bibliotecas digitais típicas" e também no facto de se proceder, precisamente pela quantidade de dados e correspondente necessidade de tempo para trabalho manual, a métodos de "indexação automática" dos materiais arquivados (Costa, 2014: 19). O fluxo de trabalho habitual dos arquivos da web caracteriza-se pela aquisição (por via da digitalização de materiais analógicos, de recolha automática de materiais digitais ou por oferta de terceiros), armazenamento, pesquisa, apresentação e preservação dos materiais digitais (Costa, 2014). A recolha automática é realizada através de *software* especializado, seguindo algoritmos desenvolvidos pelos responsáveis de operação dos arquivos da web (habitualmente, bibliotecários ou engenheiros informáticos). As

políticas de aquisição dos arquivos da web são muito variáveis, sendo que cada um segue a sua própria perspectiva sobre o que deve ou não ser preservado.

Um dos primeiros e um dos mais conhecidos arquivos da web é o americano Internet Archive, criado em 1996. Este arquivo preserva livros, fotografias, vídeos e páginas web e reúne mais de 5 milhões de gigabytes. Ainda que este arquivo da web, como outros, seja anterior aos anos 2000, foi apenas nesta última década no contexto já referido de crescente reconhecimento da importância da preservação dos materiais digitais que se multiplicaram por todo o mundo iniciativas de Preservação da Web (Wikipedia, s.d.a). Em 2014, o conjunto dessas iniciativas tinha guardado mais de 17 milhões de gigabytes, estando a maioria integrada no IIPC (Costa, 2014). O Arquivo.pt é uma dessas iniciativas. Este arquivo é apresentado como "uma infraestrutura de investigação que permite pesquisar e aceder a páginas da web arquivadas desde 1996" com o principal objectivo de preservar a "informação publicada na Web para fins de investigação", servindo para "pesquisar informação do passado que já não se encontra disponível na Web" e "fornecer recursos de investigação, por exemplo, nas áreas da História, Sociologia ou Linguística" (Arquivo.pt, s.d.). Enquanto o Internet Archive recolhe a web global, mas apenas parcialmente algumas das webs nacionais, como a portuguesa, e apenas permite pesquisa por endereço das páginas web, o Arquivo.pt faz recolhas exaustivas da web portuguesa e permite a pesquisa por endereço, mas também por termo, e permite também a utilização de computação automática de dados.

Assim, para além das questões colocadas anteriormente sobre as tecnologias e os *media* digitais e a relação entre jornalismo e memória, as mesmas tecnologias oferecem também aos investigadores simultaneamente novas possibilidades e novos problemas. Do ponto de vista das ciências sociais, esta convergência de questões emergentes da mudança cultural-económica-tecnológica não é uma mera circunstância nem uma simples conjuntura favorável. A própria convergência entre objectos, temas e problemas tem significados que são importantes para os cientistas sociais. Esses significados encontram-se no nexo de um processo em curso de informacionalização das sociedades, no âmbito do qual as tecnologias e os *media* digitais e os valores que incorporam são protagonistas. A percepção da relevância dos significados deste processo para o desenvolvimento deste projecto não foi imediata. Quando a proposta

para o projecto foi elaborada, a concepção de partida era a seguinte: utilizar materiais do Arquivo.pt como o fazem há décadas humanistas e cientistas sociais, sobretudo historiadores, utilizando materiais de arquivos físicos para dar forma a uma investigação com questões concretas sobre a mudança social a partir dos contributos da História e da Sociologia da Comunicação, especificamente utilizando os materiais de arquivo para sustentar uma análise diacrónica explicável das mudanças estruturais desses *websites*. Durante a formação em Preservação da Web dada pela equipa de informáticos especialistas em Preservação da Web do Arquivo.pt, ao longo das interacções com o material digital preservado no Arquivo.pt e por impulso da revisão de literatura realizada, a investigação sobre as transformações que se procurara inicialmente começou a colocar muitas mais questões do que aquelas que se poderiam ter colocado inicialmente, sobretudo acerca das questões levantadas pela técnica na contemporaneidade.

Em conjunto, a formação, os materiais e os trabalhos prévios tornaram bem clara a necessidade e a responsabilidade dos cientistas sociais para com a sociedade de considerarem o sentido dos problemas que tratam no quadro amplo das relações sociais e dos processos sociais que estão na sua origem ou, pelo menos, que os situam em certos contextos. Neste quadro, afloraram varias questões designadamente acerca de pressupostos a respeito da realidade, a sua natureza e a sua organização (ontológicos), pressupostos sobre como podemos conhecê-las (gnosiológicos), particularmente sobre como podemos fazê-lo cientificamente (epistemológicos), pressupostos acerca dos métodos através dos quais podemos alcançar tal conhecimento (metodológicos) e também pressupostos acerca do modo como compreendemos e construímos uma narrativa que enquadre tal realidade, tais conhecimentos e tais métodos na sociedade em geral (económicos, culturais, históricos, políticos e sociais).

Reconhecer o carácter multidimensional e, portanto, também social, das tecnologias e dos *media* digitais e do seu desenvolvimento permite ir para além de uma visão tecnicista ingenuamente determinista ou ingenuamente neutral subscrevendo uma perspectiva processual e crítica. Na sua vida quotidiana, os seres humanos regulam a vida através de uma naturalização de uma diversidade de factos em que se envolvem, o que é facilmente assinalável na esfera dos costumes, como no caso dos horários de

acordar e de adormecer ou do que se come e quando se come. Esta naturalização é parte comum do processo de constituição da realidade social em geral (Berger & Luckmann, 2010). No entanto, do ponto de vista da elaboração do conhecimento sistemático sobre esta realidade, compete aos cientistas sociais entrar em ruptura para com os próprios mecanismos sociais que naturalizam estes factos e compreender estes factos inserindo-os na dinâmica relações e dos processos sociais que os constituem e no âmbito dos quais podem ser discutidos. Por conseguinte, estes factos são percebidos como "factos" ou "fenómenos sociais totais", ou seja, como factos que combinam tudo o que é constitutivo da vida social das sociedades, expressando simultaneamente estruturas e instituições sociais, formas de produção, distribuição e consumo e fenómenos estéticos e morfológicos (Mauss, 2005), ou como "relações" e "acções sociais", isto é, como um conjunto de comportamentos que se engendram intersubjectivamente, decorrendo do sentido atribuído pelos diferentes agentes envolvidos na acção e da correspondente orientação reciproca dos seus actos (Weber, 1997). Neste contexto, a desnaturalização pode ser entendida como uma operação científico-social de ruptura para com a ordem do senso-comum pré-constituída (Silva, 1986), de distanciamento em relação às ordens fundadas por forças emocionais que vinculam agentes a estruturas específicas (Elias, 1997) e ainda de ruptura para com as estruturas dos poderes dominantes nessas ordens, inclusivamente das formas de conhecimento dominantes, como o conhecimento científico, particularmente o científico natural (Santos, 2010). No quadro de investigação sociológica sobre questões tecnológicas, isto significa que a tecnologia – com todos os tipos de aspectos sociais que fazem parte da tecnologia e de que a tecnologia faz parte – deve ser "desfatalizada", no sentido de que as dúvidas, os receios, as ansiedades e a impotência que assolam as pessoas que pensam no desenvolvimento tecnológico devem encontrar resposta no questionamento dos valores, dos objetivos e dos enredos que estão em jogo naquilo que se entenda por tal desenvolvimento (Martins *apud* Urbano & Guarda, 2001; Garcia et al., no prelo), mas também "desindividualizada", no sentido de que embora as tecnologias tenham abertura para uma diversidade de usos, o seu desenho constitui um estimulo inicial para práticas que, assim sendo, são pré-determinadas (Garcia et al., no prelo; ver também Winner, 1980). Numa fórmula sintética, pode referir-se a Primeira

Lei da Tecnologia enunciada pelo historiador da tecnologia Melvin Kranzberg: "A tecnologia não é nem boa nem má; também não é neutral" (Kranzberg, 1986: 545).

É porque a tecnologia não é neutra que as ciências sociais devem promover uma reflexão e uma ampla discussão fundamental para a compreensão do nosso contexto tecnológico. Levando isso em consideração, o trabalho sobre os jornais torna-se um trabalho reflexivo sobre as práticas de investigar as transformações dos jornais. É na dialéctica entre essas duas formas de saber que se estabelece a base desta investigação. Isto significa que um estudo sobre o material digital dos jornais – especialmente se esse estudo for sobre *websites* de jornais, jornais digitais ou qualquer tipo de material jornalístico nativo digital – não deve ignorar o quadro social altamente complexo a partir do qual essa questão pode surgir, isto é, as suas condições sociais. A questão central no contexto dessa reflexão tem que ver com as tecnologias e os *media* digitais, pois é no quadro do desenvolvimento que tornou possíveis e efectivos novos usos e respectivos problemas em torno de tais objectos técnicos que convergem as problemáticas que dão forma, sustento e legitimidade ao trabalho que se apresenta, visto que o projecto científico presente resulta, por um lado, do que de novo se apresenta no jornalismo por impulso das tecnologias e dos *media* digitais e, por outro lado, do que de novo se apresenta nas ciências por força dessas mesmas tecnologias nas formas não só dos novos arquivos como também dos modos de conhecer e de investigar a eles associados. Como é possível que os jornais tenham materiais digitais? Como é possível que o arquivo da Web armazene materiais digitais de jornais? Como é possível que as ciências sociais utilizem ferramentas digitais? Estas são as questões que se entrelaçam.

Estas questões podem ser pensadas pelo menos em dois âmbitos distintos neste projecto científico. O primeiro diz respeito à relação entre as culturas científicas e epistémicas dos engenheiros e dos cientistas sociais, particularmente às formas diferentes de olhar para os produtos tecnológicos. As tecnologias e os *media* digitais têm produzido muitos dados nos últimos anos e têm surgido perspectivas tecnicamente correspondentes para tratar esses dados. A estes domínios que lidam com grandes dados tem-se dado genericamente o nome de "Big Data". Embora não haja um consenso definitivo, o conceito de "Big Data" decorre de um desenvolvimento tecnológico das bases de dados, graças a capacidades conjugadas de computação, design e

armazenamento, permitindo que um conjunto de dados possa ter simultaneamente três características: elevado "volume" (à escala dos terabytes ou superior), alta "velocidade" (próxima do "tempo real") na sua produção e grande "variedade" nos tipos de conteúdos (Kitchin, 2014). Outras características podem ser referidas, nomeadamente: a "exaustividade" do escopo dos dados e o "detalhe", "flexibilidade" e natureza "relacional" dos dados (*ibid*.).

O mundo dos grandes dados tem dividido pensadores e cientistas, particularmente no caso das ciências sociais. A perspectiva geral encontra uma síntese possível na ideia de "convergência" de perspectivas, métodos e técnicas que o National Research Council apresenta como caminho para a relação das ciências da vida com a Física, a Química, a Matemática, as ciências sociais, as ciências computacionais e as engenharias (National Research Council, 2014) e que tem a sua contraparte na noção de "cultura de convergência" no âmbito das ciências sociais, uma concepção segundo a qual no contexto de uma "transmediação" generalizada se cruzam os *media*, as empresas e as pessoas a um nível global (Jenkins, 2006). Alguns entusiastas do Big Data consideram que a acumulação de dados conduz a uma reforma completa do sistema científico, que se reconverteria complementarmente num sistema técnico de procura de correlações entre dados. O caso mais famoso de alguém que defenda esta posição é o do tecnólogo Chris Anderson, que sustenta que "com dados suficientes, os números falam por si mesmos", porque, para o Anderson, uma grande quantidade de dados analisados através de ferramentas estatísticas substituem a causalidade, os modelos coerentes, as teorias unificadas e a explicação científica, de tal modo que a ciência teria algo a aprender com a Google (Anderson, 2008).Esta posição e congéneres levantam quatro grandes problemas:

- com base na suposição de que o grande problema epistemológico do Big Data é em si mesmo o crescimento de dados supõe-se que a solução reside em tecnologias de gestão desses dados, quando o problema epistemológico está, na verdade, em como descobrir "pequenos padrões" nos grandes dados e, portanto, a solução não é simplesmente técnica, mas sobretudo epistemológica, sobre como fazer e responder a questões, como descobrir esses padrões (Floridi, 2012);

- as ciências movidas por impulso dos dados e da análise computacional estariam focadas naquilo a que era costume chamar "ciência aplicada" e em raciocínios e métodos puramente indutivos e a-teóricos (McFarland, Lewis & Goldberg, 2016), podendo renovar um indutivismo ingénuo na forma de um "empirismo de base tecnológica", não reconhecendo o enquadramento e as limitações teóricas, metodológicas e instrumentais da actividade científica e vendo ingenuamente significado em dados sem sentido se não tiverem um contexto (Mazzocchi, 2015), Rob Kitchin referiu-se concretamente a quatro falácias do empirismo "dataísta": a ideia de que o Big Data pode oferecer a totalidade de um certo domínio da realidade, a ideia de que não existe necessidade de estabelecer teorias, modelos ou hipóteses, a ideia de que os dados falos por eles mesmos sem viés humanos e a ideia de que o significado do que estiver em causa numa investigação transcende as especificidades cognitivas de um dado domínio do saber (Kitchin, 2014);
- essas ciências teriam uma acção colonizadora das ciências em geral em função dos interesses da indústria global, porque a produção de conhecimento seria expropriada da reflexão filosófica, histórica e ética ficando totalmente dependente da sua aplicabilidade técnica, da criação correspondente de patentes e da sua adequação aos mercados, alguns autores referem-se a um acoplamento da ciência com a tecnologia conducente a uma profunda alteração do *ethos* da ciência moderna (Garcia, 2007; Garcia & Martins, 2009);
- por seu turno, as ciências sociais seriam colonizadas pelas engenharias, dado que perderiam para os engenheiros o monopólio sobre os dados sociais e sobre a análise desses dados (McFarland, Lewis & Goldberg, 2016), neste quadro têm-se manifestado leituras e discussões sobre a crise das ciências sociais na qualidade de ciências empíricas (*vide* Crompton, 2008; Savage & Burrows, 2007; Webber, 2009).

Não podemos permitir que os arquivos da web se desenvolvam como um "mercado de futuros tecnológicos"

Não aceitando argumentos cépticos e dogmáticos e reconhecendo o processo histórico em curso de que os cientistas sociais também fazem parte e onde a informática

e o digital têm uma relevância crescente, vários autores têm sugerido perspectivas para recolocar o Big Data numa relação com as ciências sociais. Nos Estudos de Comunicação, Media e Jornalismo têm-se procurado renovar as perspectivas sobre a investigação em jornalismo. Especificamente, tem-se reflectido também no uso de Big Data e na reflexão sobre esse uso por parte de jornalistas na execução das suas investigações e também por parte dos investigadores de ciências sociais que tomam os produtos jornalísticos como objecto de estudo (*vide* Karlsson & Sjøvaag, 2016a; Karlsson & Sjøvaag, 2016b; Lewis & Westlund, 2015; Mahrt & Scarkow, 2013). Existem outras posições, algumas muito extremadas. Logo à partida, o facto de as ciências sociais terem a capacidade reflexiva para pensar os seus objectos mas também os seus métodos e as suas técnicas no quadro dos mundos sociais que estudam faz com que possam contribuir para a percepção do significado do Big Data e do seu lugar na estrutura social (Mützel, 2015). Alguns exemplos de perspectivas:

- "Teoria social maquievélica": neste caso, pensa-se a teoria social como enformação do método a partir das problemáticas tratadas e do quadro teórico de referência, confrontando o mundo no pensamento (Frade, 2016);
- "Ciência social sinfónica": a sociologia teria um papel na modelação do Big Data promovendo uma reflexão abdutiva focada na interacção co-constitutiva entre dados, método e teoria, envolvendo o Big Data na própria teoria social (Halford & Savage, 2017);
- "Ciência social forense": a partir desta perspectiva prevê-se uma síntese ou hibridação entre perspectivas dedutivas e indutivas e teoria e evidencia por meio de, ao mesmo tempo, explorações teoricamente enquadradas dos dados e procura de teorias e de explicações alternativas a partir dos próprios dados (McFarland, Lewis & Goldberg, 2016).

Ainda que a discussão sobre o estatuto de Big Data dos arquivos da web ainda esteja por fazer de forma sistemática, o que se afirmou sublinha a necessidade de uma atitude de prudência por parte dos investigadores que utilizem os arquivos da web, atitude que se traduz na conjugação entre forças de enquadramento histórico e forças de sistematização teórica. Assim, as questões referidas previamente revelam-se também num segundo âmbito, que diz respeito ao modo como as investigações em

arquivos da web podem ser levadas a cabo, como têm sido concretamente realizadas e como têm sido pensadas. Ainda que os arquivos da web possam não ter a quantidade massiva de dados que são produzidos em grandes estruturas tecnológicas internacionais de projectos de outras áreas, como no caso da European Organization for Nuclear Research, os dados que esses arquivos produzem são suficientes para o seu tratamento ser encaminhado para os domínios da automatização de processamento de dados, o que acontece logo de raiz devido ao que se disse sobre os processos de recolha e indexação de material digital. Portanto, para utilização destes dados, os cientistas sociais enfrentam uma relação semelhante com a cultura de engenharia que no caso do engajamento com o Big Data: há processos automatizados e, concomitantemente, ocultos em relação à percepção das pessoas sem compreensão técnica desses processos, portanto, para compreensão dos processos, a única alternativa dos cientistas sociais está na formação dada por e do diálogo estabelecido com os técnicos envolvidos nesses processos. No caso específico da investigação em arquivos da web, o conhecimento adquirido durante este contacto parece contribuir mais para o bom trabalho dos cientistas sociais do que contribuiria um estranhamento excessivo. Sabendo como funcionam os arquivos da web tecnicamente, o cientista social pode tomar decisões relativamente informadas sobre a sua investigação, pode também fazer uma crítica informada dos métodos disponíveis e pode ainda procurar fazer uma utilização cientificamente sustentável e culturalmente crítica do material tecnológico disponível. As possibilidades de articulação cognitiva/intelectual que daí advêm são fundamentais no trabalho reflexivo das ciências sociais.

Num projecto semelhante ao "Investiga XXI" realizado no Reino Unido, vários cientistas sociais procuraram utilizar material de arquivos da web de forma relativamente experimental e levantaram questões metodológicas que pareciam irresolúveis, como o paradoxo de haver poucos dados preservados, mas os arquivos da web estarem repletos de dados, a dispersão do conjunto desses dados ou a dificuldade de encontrar significados sociais na massa demasiado variável desses dados (*vide* BUDDAH, s.d.; Cowls, 2017). Num dos trabalhos realizados no âmbito desse projecto, o investigador apresentava problemas como o significado não histórico, mas informático das anomalias no material arquivado, a inutilidade imediata dos dados em si mesmos e

uma certa incompreensão em relação aos processos de recolha da web (Millward, s.d.). Não seria diferente este tipo de diagnóstico se o trabalho de investigação em arquivos da web estivesse engajado num enquadramento das ciências sociais? Tal como podem contribuir para perceber o crescendo de dados nos seus mundos sociais, os cientistas sociais podem também contribuir para a investigação nos arquivos da web através do seu enquadramento analítico e crítico (Winters, 2017).

Segundo Federico Nanni, há duas formas de considerar os materiais dos arquivos da web como fonte primária para compreender a mudança histórica (Nanni, 2017):

- no âmbito da perspectiva dos Estudos de Media e dos Estudos de Internet procura-se contextualizar e compreender mudanças de layout, estrutura, conteúdo e usos da web;
- no âmbito da perspectiva historiográfica propriamente dita, os materiais são tomados como fontes históricas em termos culturais, económicos, políticos e sociais.

Próximo do sentido da primeira perspectiva, Niels Brügger tem apresentado algumas ideias interessantes para se pensar as relações dos Estudos de Media com as Humanidades em geral, a formação a que se tem chamado Humanidades Digitais e os Estudos de Internet (Brügger, 2015). Brügger defende que as Humanidades, as Humanidades Digitais, os Estudos de Media e os Estudos de Internet são áreas íntimas, porque partilham de uma compreensão ao nível de objectos, teorias, métodos e ferramentas analíticas e também ao nível da colaboração e disseminação dos resultados de investigação. Brügger considera que entre essas disciplinas há três nexos de relações de reciprocidade:

- as relações reciprocas entre as Humanidades e as Humanidades Digitais devem-se à crescente preocupação das humanidades com questões levantadas pelo digital e ao repensar de práticas de investigação no contexto das humanidades digitais, designadamente ao uso de métodos informáticos de investigação;
- as relações recíprocas entre os Estudos de Media e os Estudos de Internet resultam de os Estudos de Internet receberem dos Estudo de Media uma tendência de reflexão sobre para a reflexão sobre a Internet enquanto *medium*,

com a sua materialidade especificas e com interacções, conteúdos e instituições específicos e de os Estudos de Media incluírem elementos dos Estudos de Internet, como as diversas formas de meios digitais presentes na internet;

- as relações recíprocas entre as humanidades digitais (e indirectamente as humanidades) e os Estudos de Internet (e indirectamente os Estudos de Media) partem logo de um mesmo domínio onde impera o digital, mas um aumento de proximidade entre ambos seria interessante na medida em que a internet, especialmente a web, é cada vez mais importante para as Humanidades Digitais, porque lhe fornece materiais nativos digitais e renascidos digitais e ferramentas de estudo e meios de colaboração e disseminação de resultados, e na medida em que os Estudos da Internet, por sua parte, poderiam alargar os seus horizontes no contexto de uma tradição humanística.

Noutro texto Brügger defende que o material digital em si mesmo constitui uma força motriz para a transformação das humanidades, no sentido de que esse material se torna cada vez mais importante do ponto de vista do estado actual da sociedade e porque pode ser colectado à distância e podem utilizar-se ferramentas digitais para a pesquisa, filtragem e análise e uma grande diversidade de suportes digitais para a disseminação e debate dos resultados (Brügger, 2016). Brügger salvaguarda, no entanto, que não se deve entender a relação das humanidades com o digital de uma forma determinista e unidireccional dos computadores para o humano. Pelo contrário, defende uma perspectiva dialéctica que tem como ponto de partida a materialidade digital de novos objectos de estudos e, consequentemente, de ferramentas que nos podem auxiliar na compreensão desses objectos e da sua materialidade.

O uso da web como uma fonte de materiais para investigação implica a delimitação dos materiais disponíveis de forma analítica para constituir objetos de estudo (Brügger, 2009). A delimitação analítica implica vários níveis de delimitação dos materiais. O primeiro nível diz respeito à diferenciação entre os vários materiais disponíveis na web em relação a certos elementos. Do mais complexo ao menos complexo, temos os seguintes elementos (variando do menos complexo ao mais complexo, os contextos uns dos outros) (Brügger, 2009; Brügger, 2012):

- a “web como um todo”, a totalidade do material da web;

- a “esfera da web”, ou seja, actividades no contexto da web relacionadas com um tópico específico;
- o “*website*” enquanto colecções de páginas web delimitadas pelo protocolo de transferência de hipertexto;
- a “página web” propriamente dita, isto é, o material principal visto nas janelas do navegador;
- os “elementos web”, como textos, imagens ou vídeos.

A definição dada de “*website*” é insuficiente porque está focada apenas no campo da tecnologia informática. A unidade de um *website*, na esfera da web, depende de dois aspectos: um elemento informático de registro em protocolo, permitindo distinguir um endereço da web em relação a outro, mas também um elemento institucional, como referência de um endereço a uma determinada instituição social. Então, quando um jornal tem um *website* com um determinado endereço e, ao longo do tempo, muda esse endereço, estará mudando de *website*? Tecnicamente, sim. Socialmente, não. A unidade deste duplo fenómeno é dada por uma operação de computador chamada “redireccionamento” ou “encaminhamento”, ou seja, um *website* tecnicamente descontinuado direcciona os usuários para um novo *website* sobre a mesma instituição (Wikipedia, s.d.b). Desta forma, considera-se como um *website* de um dado jornal todos aqueles que estavam conectados entre si na web por operações deste género, o que poderá ou não acontecer automaticamente nos arquivos da web (ao procurarem preservar e apresentar as páginas web como elas teriam sido).

A todos estes elementos acrescenta-se a “página inicial” (*homepage*). Embora este termo seja muito impreciso nos seus usos gerais, será utilizado nesta investigação para nomear a página de entrada de um *website*, onde estão todos os seus elementos da interface gráfica para apoio à navegação.

O segundo nível de delimitação analítica dos materiais relaciona-se com a natureza dos materiais disponíveis. Do ponto de vista ontológico, todos são objectos digitais, no sentido de que são objetos cujas qualidades são representadas na forma de dados e meta-dados (Hui, 2012). No entanto, existem diferentes tipos de objectos digitais entre si. No âmbito da web, podemos falar de três grandes tipos de objetos (Brügger, 2012; Brügger, 2016):

- “digitalizados”, ou seja, material digital inicialmente projetado e desenvolvido como um todo através e para *media* analógicos;
- “nativos digitais”, material digital desde as suas origens;
- “renascidos digitais” (*reborn-digital*), isto é material nativo digital ressuscitado em arquivos da web.

O material renascido digital aparece como um fenómeno específico de arquivos da web. Na verdade, parece ser a base do contributo desses arquivos para a investigação científica, levantando também, todavia, questões particulares. A discussão em torno do material renascido digital permite problematizar os assuntos mais comumente levantados no uso dos arquivos da web, particularmente questões práticas com as quais os investigadores são confrontados. Algumas das questões são as que em seguida se apresenta. Como podemos organizar a enorme quantidade de dados e informações disponíveis nos arquivos da web? É o problema da quantidade de dados e informações. Existem redireccionamentos entre páginas que não são esperadas, o que fazer sobre isso? É o problema da dispersão de dados. Como o processo de preservação é contínuo (podem ser encontrados materiais de backup, etc.), como pode um investigador seleccionar o material para estudar? É o problema da “dinâmica de actualização” (Brügger, 2005). Como podemos saber se o que vemos num arquivo da web é o que estava na web quando esse material ainda não tinha sido descontinuado? É o problema da autenticidade. Finalmente: como podemos confiar nesse material sem ter certeza da sua autenticidade? É o problema complementar da fiabilidade.

Olhe-se para os problemas da quantidade de dados e informações, da dispersão e da dinâmica de actualização. Uma quantidade de dados num arquivo é muito menor do que a quantidade de dados na web, mas ainda corresponde a um grande número de elementos quando alguém propõe analisá-lo um por um. Vários projectos de investigação em arquivos da web foram profundamente alterados por causa da quantidade de dados, tal como se observou no exemplo dado do “BUDDAH project”. Além disso, os *websites* não são publicados com limites de tempo rígidos. Por vezes, nem sequer há ancoragem temporal das publicações nas próprias publicações. Assim, a indicação de uma data e de uma hora nos arquivos da web é, portanto, instituída no próprio processo de preservação, podendo não corresponder uma ancoragem temporal

expectável prévia a esse processo. Isto faz com que os arquivos da web sejam "multitemporais", no sentido de que os *websites* não são arquivados na sua totalidade, existindo *websites*, páginas web e elementos web repetidos e alguns sem continuidade temporal com outros Brügger & Finnemann, 2013). Assim sendo, certas páginas diriam respeito ao mesmo dia, mas teriam apresentações diferentes.

Brügger propôs uma "filologia da web" para resolver esses problemas (Brügger, 2008). O objetivo desta filologia é "examinar as diferenças e as similaridades entre variantes de evidências textuais que são parecidas entre si" (*idem*: 162). Pressupondo que nenhuma das versões arquivadas seja idêntica, o filólogo umas páginas web e respectivos com outras desde que tenham as mesmas indicações temporais de preservação e/ou sejam oriundas de outros arquivos da web (tendo também uma ancoragem temporal idêntica).

Veja-se agora o problema da autenticidade e da fiabilidade. Se no processo de digitalização o material analógico é totalmente guardado nas suas características básicas, os materiais renascidos digitais dos arquivos da web estão sempre incompletos. Os objectos em falta podem ser divididos em dois tipos (Brügger, 2008; Brügger & Finnemann, 2013):

- informações sobre a web *online*, como estatísticas, sistemas de classificação, número de nomes de domínios, usuários, etc.;
- elementos web específicos e elementos funcionais dos *websites*, como imagens, hiperligações, gráficos, etc.

Para Brügger e Niels Ole Finnemann, a incompletude é a característica geral de qualquer arquivo, não apenas de arquivos da web (Brügger & Finnemann, 2013). Utilizando como contraparte a relação entre materiais físicos e a sua digitalização, os autores referem-se ao facto de os materiais dos arquivos da web nem sempre terem um material absolutamente original para se fazer uma comparação e, por conseguinte, perceber o que estaria em falta. É claro que este é um problema historiográfico antigo, porque, por um lado, nem sempre se encontram mais do que cópias de documentos, sem se conhecer o paradeiro dos originais ou sabendo-se que os originais foram destruídos em algum processo histórico, sendo que a sobrevivência ou destruição de

certos documentos podem ser vistos como um fenómeno contingente, no sentido de que são fruto do acaso (*vide* Winters, 2017), e trata-se de um problema antigo porque, por outro lado, há documentos que nunca tiveram originais propriamente ditos, como panfletos de propaganda política, não tendo mesmo autores definidos, ao contrário do caso da configuração literária de um romance (Foucault, data). Para superar o problema da incompletude no caso específico dos arquivos da web, devemos perceber, segundo Brügger e Finnemann que, contrariamente aos materiais digitalizados, os materiais renascidos digitais não são necessariamente duplicados, não são necessariamente cópias de escala de 1: 1, são versões diferentes. Para Brügger e Finnemann, devemos assumir que "a web arquivada é uma versão renascida, única e deficiente e não simplesmente uma cópia do que estava online" (*idem*: 74).

Como renasce? Para Brügger, esta é uma construção que resulta de três razões (Brügger, 2013):

- os responsáveis pela preservação fizeram escolhas técnicas (estratégias de preservação e *softwares* utilizados) que determinaram as versões;
- pode ter havido mudanças nos *websites* por altura da sua preservação e tais mudanças não estão documentadas, portanto, além de incompleto pode haver inconsistências (como as já referidas inconsistências temporais);
- e pode haver problemas arquivísticos muito mais concretos que fazem com que a preservação escape a um plano prévio, como a existência de conteúdos não-preserváveis.

Deve sublinhar-se que não se deve, neste contexto, confundir "arquivamento" com a "preservação" propriamente dita. A última implica não apenas a aquisição e armazenamento de conteúdos, mas a manutenção continuada do acesso a esses conteúdos.

Finalmente, é necessário agora esclarecer como podemos avaliar e seleccionar versões disponíveis nos arquivos da web. O filólogo da web, na visão de Brügger, procura a relação de semelhança e simultaneidade entre as várias versões (Brügger, 2008). O carácter digital do arquivo da web faz com que algumas comparações sejam automatizadas. Para Brügger, no âmbito análise, a versão selecionada para ser

estudada, tendo em conta uma ancoragem temporal que coloca as duas versões comparadas no mesmo momento do sistema, deve ser a versão menos deficiente. Do ponto de vista do renascimento digital, será realmente o "original" (*ibid*.). Esta perspectiva levanta um problema: a versão a que não se tem acesso por estar *offline* e que seria, por assim dizer, a autêntica original poderia deficiente em relação a si mesma. Não se encontrou nenhuma discussão de Brügger sobre este assunto. Este problema coloca-se, por exemplo, quando um arquivo da web arquiva uma página com o erro 404. Neste caso, o pedido de um utilizador não é cumprido e é esse incumprimento que é armazenado. Assim sendo, a página autêntica original é que foi, digamos, menos fiel a si mesma, sendo aquele momento capturado de forma rigorosa pelo arquivo.

Neste contexto, pode adicionar-se aos níveis de delimitação analítica mencionados um outro nível relacionado com diferenciação do que vemos na interface de um arquivo web e, portanto, para a forma como percebemos o seu conteúdo. Seguindo uma vez mais Brügger, pode delimitar-se o material de arquivos da web como objecto de análise em pelo menos três "ambientes" (Brügger, 2008; Brügger, 2009; Brügger, 2010):

- o "ambiente dos *media*" do *website*, que é a internet como um todo em que um *website* está incorporado e que dá a esse *website* a especificação das suas possibilidades e condiciona os seus resultados em termos técnicos, actuando particularmente em níveis de textualidade binária não perceptíveis pelo usuário;
- o "ambiente textual", contexto de elementos significantes com relações semânticas mais complexas do que os fonemas (este ambiente tem duas grandes formas: ou se trata do *software* de navegação e auxilio à navegação (fluxo de informações entre servidores, os protocolos de transferências, as traduções de códigos) ou diz respeito à interface gráfica do utilizador (ambiente de trabalho e respectivas janelas e ícones);
- a "textualidade", que, para Brügger, se refere aos elementos textuais que trabalham os elementos da web, como os elementos escritos, os elementos de som, os elementos da imagem estática, os elementos da imagem dinâmica e também elementos audiovisuais (os últimos não são incluídos por Brügger).

Continuando a seguir Brügger, os elementos que constituem a textualidade podem analisados em duas dimensões:

- através de uma "análise morfológica" o foco é colocado nas características dos elementos individuais;
- através de uma "análise sintática" a atenção é dada à combinação de elementos em numa "composição audiovisual".

Estas análises podem ser pensadas, por sua vez, também em três níveis:

- o nível de "inter-relações semânticas", relacionado a questões de coesão, coerência, gramática, léxico, isomorfismo, etc.;
- o nível de "inter-relações formais", relacionado a questões de forma, como barras, separadores, quadros, tipografia, etc.;
- o nível das "inter-relações fisicamente performativas", relacionadas a questões de ações e movimentos, como rolagem, clique, etc.

Os níveis analíticos referidos podem ser, de uma forma mais geral, diferenciados em dois níveis (Brügger, 2008):

- num "nível imediatamente perceptível", que combina os sinais auditivos e visuais nos quais os usuários estão envolvidos durante o relacionamento com a interface,
- numa "variedade de níveis textuais subjacentes que não são imediatamente perceptíveis", que são as camadas de código e anotações escritas em binários.

O ambiente dos *media* e o da textualidade, ao nível das interações fisicamente performativas, parecem estar muito relacionados com essa variedade de elementos não perceptíveis, mas que, em alguns casos, se tornam visíveis com os comandos dos utilizadores. Assim, embora o ecrã levante problemas de compreensão devido à sua crescente complexificação técnica, particularmente ocultando vários dos seus aspectos aos utilizadores e inibindo assim o seu controlo directo, pode encontrar-se no ecrã ou através do ecrã "um caminho para aquilo que se esconde, mediante uma comunicação crítica e desmistificadora da imagem" (Mendonça, 2010). Para além de uma dimensão técnica, num sentido estrito, os ecrãs têm uma dimensão social e semiótica que se manifesta na noção de "interface desenhada", no sentido de que nela se apresenta e

exibe o mundo como imagem (Pinto-Coelho, 2010). Na sua codificação em mensagens visuais, o mundo é apresentado ao utilizador com formas, disposições, cores, tipografias, cores próprios que podem ser analisados recorrendo a ferramentas da web. Hoje, um *website* tem uma coluna que poderá vir a não ter daqui a um mês, por exemplo. Para alguns autores a ordem cultural imanente nesta apresentação é tal que os conteúdos seriam absorvidos pela própria ideia de interface (Lévy, 1994; Rokeby, 2003).

As questões ao nível da imagem que se levantam têm directamente que ver com o modo de existência do arquivo da web, do Arquivo.pt em particular, como um objecto técnico. As delimitações que têm sido feitas até aqui permitem já ter uma ideia geral relativamente à estrutura técnica de um arquivo da web, mas é necessário perceber que tipo de imagens produz um arquivo desse género e em que medida estas imagens podem ser utilizadas para investigação científica. Quando perguntamos o que são estas imagens sugere-se neste projecto que se dêem três respostas:

- "imagens incorporadas", ou seja, imagens mais isoladas que habitualmente vemos nos *websites*, imagens no sentido de elementos web individuais, como *thumbnails* ou quaisquer formas individualizadas de ilustrações ou fotografias visíveis nos *websites*;
- "*snapshots*", isto é, a expressão gráfica das versões preservadas na interface gráfica dos arquivos da web (no caso do Internet Archive chama-se a essa interface gráfica Wayback Machine), aqui a imagem é uma página web dinâmica como um todo no quadro do arquivo, corresponde à captura de um momento do sistema com todos os elementos web que compõem dada página nesse momento, podendo com eles interagir-se, dado que a interface os arquivos web têm a possibilidade de gerar versões das páginas web como elas aproximadamente terão sido em determinada altura;
- "*browsershots*", quer dizer, a forma estática da expressão gráfica das versões preservadas, é uma espécie de *screenshot* amplo do conteúdo de um *browser*, neste caso seria um *screenshot* produzido a partir dos *snapshots*, uma imagem que reúne todos os elementos web que compõe uma página web sem possibilidades de interacção e apenas com alguns desses elementos visíveis num

certo momento, é uma espécie de conversão imagens dinâmicas em imagens estáticas representativas de um momento muito muito situado dessa dinâmica, alguns arquivos como o Arquivo.pt, têm a opção de fazer este trabalho de conversão em imagem estática a partir de um *snapshot*.

Este último tipo de imagens é altamente paradoxal do ponto de vista ontológico e semiótico: embora tenha uma origem digital não representativa, é uma captura representativa muito próxima das imagens analógicas. Ou seja, um *website* tem páginas web com muitos elementos dinâmicos. No caso de fotos de eventos extra-web, haverá algum grau de representatividade, como ocorre nas imagens resultantes da actividade de fotojornalismo. Mas a transformação produzida no processo informático de conversão numa imagem estática cria uma imagem que é representativa da estrutura do próprio *website* de que essas outras imagens fazem parte. Portanto, a faculdade de transformação da web em imagens dinâmicas e destas em imagens estáticas abre portas a questões sobre o estatuto da representatividade das imagens digitais. Embora não haja aqui espaço para avançar uma reflexão desenvolvida sobre este assunto, fica claro que a materialidade do arquivo, sobretudo na medida em que codifica mensagens visuais, é fundamental no processo de constituição dos objectos de estudo em contexto digital.

***

A incipiência das investigações em ciências sociais utilizando os arquivos da web como infraestruturas de investigação, principalmente tendo em conta o contexto português, faz com que a necessidade de discussão de todos estes assuntos seja muito relevante para a realização dessas investigações – fazendo falta envolver na discussão pessoas de todas as áreas científicas a quem este assunto possa interessar do ponto de vista da construção de teoria e de hipóteses de investigação e do ponto de vista das práticas de investigação.

No contexto desta investigação, optou-se por projectar três grandes etapas de investigação. Na primeira etapa procura fazer-se uma avaliação técnica e sociológica da preservação e do perfil de preservação dos *websites* dos jornais, tendo em conta os seus aspectos informáticos e os quadros sociais e empresariais que lhes servem de contexto.

Na segunda etapa procurou-se elaborar uma *rota de investigação* dentro do Arquivo.pt considerando os materiais aí disponíveis para o estudo dos *websites* dos jornais escolhidos. Finalmente, na terceira etapa apresenta-se o objectivo central deste projecto, o da análise das mudanças de estético-funcionais dos *websites* dos jornais referidos. Cada uma das três etapas de investigação está dividida, por sua vez, em três partes. Numa primeira parte, há uma discussão teórica sobre a relevância do assunto a ser investigado e sobre contextos de investigação prévios que possam enformar a investigação presente. Numa segunda parte, apresentam-se os aspectos metodológicos considerados e a metodologia concreta que será em seguida aplicada numa terceira parte, onde se fará a análise dos resultados e a sua discussão.

4 Avaliar a preservabilidade dos *websites* actuais dos jornais

Para conhecer a situação actual dos jornais estudados e a própria postura das empresas de *media* e dos jornalistas em relação aos seus materiais digitalizados e nativos digitais, é necessário considerar aquilo que realmente preservam, aquilo que simplesmente descartam porque não têm cuidados de preservação e aquilo que eventualmente eliminam intencionalmente. Neste sentido, procurou-se perceber o que era preservado e visível para qualquer pessoa nos *websites* dos jornais e o que teriam as empresas de media e os jornais a dizer sobre isso.

Para pensar a complexidade técnica e social estrutural da preservação dos materiais digitalizados e nativos digitais, é fundamental perceber que no processo de preservação estão em jogo vários aspectos técnicos diferenciáveis, várias pessoas e até mesmo várias culturas, mentalidades e políticas de preservação. Podemos distinguir, pelo menos, por um lado, aquilo que o *website* nos oferece directamente, sem necessidade de suporte exógeno para este *website*, e, por outro lado, aquilo que o *website* não nos oferece e que só podemos conhecer através do diálogo com as pessoas que fazem parte das organizações responsáveis pela (não) preservação ou consultando materiais adicionais que essas organizações possuem.

4.1 Metodologia

Para realização da investigação, considerou-se importante sintetizar esses dois aspectos de complexidade estrutural a que se chamou camadas de preservabilidade. Primeiro, procura-se avaliar as condições de preservação do *website* em termos técnicos. Para esse efeito, recorre-se a uma ferramenta chamada ArchiveReady. É uma ferramenta técnica *online* gratuita que calcula aquilo a que se chama a "arquivabilidade do *website*". Tomando como ponto de partida princípios de preservação de objetos digitais, o ArchiveReady dá uma classificação de arquivibilidade de um *website* (entendido como HTML, Imagens, CSS, JS, Sitemaps) com base na análise de um conjunto de "facetas de arquivabilidade na web", como a "acessibilidade", a "coesão", os "meta-dados" e a "conformidade de padrões" (ArchiveReady, s.d.). Aplicou-se esta ferramenta a cada um

dos *websites* actuais de cada um dos quatro jornais para perceber o seu estado atual de preservação em termos técnicos.

Em segundo lugar, investigações recentes sobre a preservação de notícias digitalizadas e materiais nativos digitais mostram que é relevante definir um conjunto de indicadores para a análise quantitativa e qualitativa das opções de arquivo dos próprios *websites* dos jornais. Embora vários autores falem da função de memória ou preservação dos *websites* de jornais, essa ideia baseia-se num tratamento dos *websites* como bancos de dados com sistemas de consulta específicos. O padrão geral de acesso a esses materiais está centrado na navegação pelo *website* e na possibilidade de "busca simples", colocando termos e/ou frases numa caixa de pesquisa, e "pesquisa avançada", que permite pesquisar por critérios como autor, publicação de data ou secção (Abadal & Guallar, 2010; Nanni, 2013). Alguns *websites* têm a opção de procura em colecções organizadas por edições, temas, datas ou categorias (Abadal & Guallar, 2010; Zamith, 2008). Alguns *websites* permitem pesquisar através de tags associadas a cada artigo, dando acesso por clique a todos os artigos que foram atribuídos a essa tag (Zamith, 2008). Existe também a possibilidade de reutilizar estratégias de pesquisa guardadas de pesquisas anteriores (Abadal & Guallar, 2010). Habitualmente, esses *websites* não possuem ferramentas de pesquisa específicas para imagens, vídeos e outros materiais não-textuais (Nanni, 2013). Uma opção comum é o texto explicativo sobre o uso do sistema de pesquisa geralmente apresentado como "ajuda" (Abadal & Guallar, 2010). Optou-se no contexto deste trabalho por não elaborar um conjunto definido e definitivo de opções como as mencionadas em forma de indicadores fixados anteriormente ao encontro com o material disponível nos *websites*. Em vez disso, procurou partir-se de uma perspectiva mais abertas a possibilidades ainda não pensadas de preservação. Assim, poderiam considerar-se não só as opções mencionadas, como outras possibilidades ainda não estudadas e que possam fazer parte da realidade actual dos jornais.

Em terceiro lugar, investigações recentes sobre a preservação de notícias digitalizadas e de materiais jornalísticos nativos digitais também mostram a relevância do diálogo com pessoas das empresas de *media* e com os jornalistas para compreender a forma como essas pessoas no seu conjunto criam e preservam esses materiais e,

portanto, que atitudes estão por detrás das suas opções de (não) preservação. Esta perspectiva permite questionar as percepções individuais acerca do presente e do futuro do armazenamento e da preservação das empresas de *media* e dos jornalistas (Carner, McCain & Zarndt, 2014; Cowen, 2001) e descobrir quem são as pessoas responsáveis pelas operações digitais nas respectivas empresas (Carner, McCain & Zarndt, 2014). Para se conseguir isto, começa por se procurar nos *websites* das empresas de *media* e dos jornais possíveis organizações próprias de preservação, como arquivos e/ou centros de documentação. Só em seguida se contactam as pessoas dessas empresas, desses jornais e dessas organizações (caso existam) para se fazerem duas perguntas. Têm meios de preservar, incluindo através de backup, conteúdos publicados *online* no *website* do jornal? E é possível que alguma pessoa possa hoje abrir uma página web do *website* do respectivo jornal que date de 1996, 1997 ou 1998?

4.2 Análise de resultados

Aplicou-se o ArchiveReady a cada *website* actual dos jornais estudados para perceber o seu estado técnico actual de preservação, dando logo à partida uma ideia da postura e/ou do conhecimento de cada jornal face à preservação digital.

A classificação geral do ArchiveReady para cmjornal.pt foi de 79% à data de 22 de Julho de 2017. Os resultados para cada faceta foram os seguintes:

- Acessibilidade: 46%;
- Coesão: 75%;
- Meta-dados: 100%;
- Conformidade de padrões: 94%.

Os resultados menos bons no campo da acessibilidade decorrem da existência de algumas hiperligações inválidas e da presença de comandos de "não-permissão" usados para bloquear o acesso dos agentes virtuais de aquisição a determinadas áreas do *website* causando problemas de preservação da web.

A classificação geral do ArchiveReady para dn.pt foi de 62% à data de 22 de Julho de 2017. Os resultados para cada faceta foram os seguintes:

- Acessibilidade: 42%;
- Coesão: 63%;
- Meta-dados: 80%;
- Conformidade de padrões: 64%.

Uma vez mais, os resultados menos bons no campo da acessibilidade decorrem de problemas com hiperligações e comandos de "não-permissão".

A classificação geral do ArchiveReady para expresso.sapo.pt foi de 56% à data de 22 de Julho de 2017. Os resultados para cada faceta foram os seguintes:

- Acessibilidade: 48%;
- Coesão: 0%;
- Meta-dados: 80%;
- Conformidade de padrões: 97%.

Os maus resultados no campo da coesão decorrem da existência de imagens remotas, que são usadas no *website* mas estão hospedadas noutros locais, e da existência de CSS remotos, que são usados no *website* mas estão hospedados noutros locais. Os maus resultados devem-se sobretudo ao problema das imagens, visto que o número de imagens remotas é muito elevado. Os resultados menos bons no campo da acessibilidade decorrem de problemas como os já referidos de hiperligações inválidas e de comandos de "não-permissão".

A classificação geral do ArchiveReady para publico.pt foi de 64% à data de 22 de Julho de 2017. Os resultados para cada faceta foram os seguintes:

- Acessibilidade: 58%;
- Coesão: 25%;
- Meta-dados: 100%;
- Conformidade de padrões: 72%.

Os maus resultados no campo da coesão devem-se a problemas semelhantes aos referidos em termos de imagens e CSS remotos. Uma vez mais, os resultados menos bons no campo da acessibilidade decorrem de problemas com hiperligações e comandos de "não-permissão".

Em seguida, a comparação dos procedimentos de preservação dos *websites* implicou procurar recursos a isso destinados na própria estrutura dos *websites*. Descobriu-se o seguinte sobre cada um dos *websites*:

- em relação ao CM, o *website* permite fazer pesquisa por termos (podendo utilizar-se como filtros data, local, pessoas, organizações e acontecimentos), permite navegar pelos conteúdos guardados no *website* nomeadamente por meio *scroll* e de clique, na conta de utilizador da Confina guarda as notícias marcadas como favoritas, a assinatura digital dá acesso a um formato digitalizado do jornal impresso (cujo conteúdo também pode ser guardado) e suplementos e a todos os respectivos conteúdos do *website* (incluindo os conteúdos fechados/pagos);
- em relação ao DN, o *website* permite fazer pesquisa por termos (podendo utilizar-se como filtros relevância e data), permite navegar pelos conteúdos guardados no *website* nomeadamente por meio *scroll* e de clique, a assinatura dá acesso a um formato digitalizado do jornal impresso (cujas edições são automaticamente guardadas numa galeria digital a que se dá o nome de "biblioteca") e suplementos e a todos os respectivos conteúdos do *website* (incluído os conteúdos fechado/pagos);
- em relação ao *Expresso*, o *website* permite fazer pesquisa por termos (podendo utilizar-se como filtros autores e artigos), permite navegar pelos conteúdos guardados no *website* nomeadamente por meio *scroll* e de clique, a assinatura dá acesso a um formato digitalizado das várias partes do jornal impresso (cujas edições são automaticamente guardadas numa galeria digital a que se dá o nome de "biblioteca") e a todos os respectivos conteúdos do *website* (incluído os conteúdos fechado/pagos), a galeria digital dá ainda a possibilidade de se comprarem edições anteriores à edição vigente até 2014;
- em relação ao Público, o *website* permite fazer pesquisa por termos (podendo utilizar-se como filtros data, tipo, secção, autor e tópico), permite navegar pelos conteúdos guardados no *website* nomeadamente por meio *scroll* e de clique, na conta de utilizador do Público podem marcar-se conteúdos para guardar, a assinatura dá acesso a um formato digitalizado do jornal e dos seus suplementos

e a todos os respectivos conteúdos do *website* (incluído os conteúdos fechado/pagos) e ainda permite a consulta de edições digitalizadas do jornal impresso desde 2001 (pode fazer-se pesquisa por palavras e data) num formato a que dão o nome de "arquivo digital".

Estas evidências mostram que a ideia de "arquivo" dos jornais está ainda muito focada num modelo de shovelware e que, ainda assim, nem todos os materiais impressos dos jornais foram digitalizados e disponibilizados nos respectivos *websites*. Quanto aos conteúdos dos *websites* propriamente ditos, nada garante que estão todos nos *websites* e desde quando estão nos *websites*, visto que não há colecções dos conteúdos dos *websites* organizadas. Mais claro é o facto de que muito embora alguns desses conteúdos mais antigos (não parece haver nada anterior a 2000), até possam estar nos *websites*, os *websites* actuais não reproduzem as condições de leitura digital que os *websites* onde tais conteúdos foram primeiramente colocados tinham. Em resumo: não parece possível ter acesso a conteúdos anteriores a 2000, mas mesmo quando tal possa acontecer não é possível ter a mesma experiência de utilização que se teria à época porque os *websites* estão completamente diferentes. O "arquivo" dos *websites* destes jornais reduz-se, então, à acumulação de informação impressa digitalizada recente organizada em pequenas bibliotecas e à informação nascida digital a ser organizada segundo pedidos por parte do utilizador através de filtros de pesquisa e operações de guardar. O foco não é, claramente, a preservação digital, menos ainda a preservação de todo o material digital (não só conteúdos, mas as próprias estruturas e respectivos aspectos estético-funcionais dos *websites*).

O próximo passo foi compreender se as empresas de *media* possuíam recursos complementares de preservação digital. Para o fazer, procurou-se nos *websites* dos jornais e das empresas de *media* referências a tais recursos. A partir da informação disponível *online* descobriu-se o seguinte:

- a Cofina (empresa de nedia do CM) tem algo designado como "Arquivo", mas que apenas serve para preservar os relatórios societários (http://www.cofina.pt/investors/archive/2012.aspx?sc_lang=pt-PT);
- o Global Media Group (a empresa de *media* do DN) possui uma organização destinada a preservar e indexar jornais e revistas do grupo chamada "Arquivo e

Centro de Documentação", no caso DN preservam-se as versões impressas desde 1864 (http://www.globalmediagroup.pt/marcas/arquivo/);

- a Impresa (a empresa de *media* do Expresso) possui uma organização semelhante designada como Gesco, que funciona como arquivo electrónico para preservação e gestão de documentos, designadamente material jornalístico impresso (http://www.impresa.pt/marcas/marca/2013-11-07-GESCO);
- o Público possui um departamento a que é dado o nome de "Centro de Documentação", que não apresenta em nenhuma página web informação acerca das tarefas no seu âmbito realizadas (http://www.publico.pt/nos/ficha-tecnica).

Finalmente, contactou-se os jornais e as referidas organizações. Procurou-se contactar as pessoas responsáveis pelos *websites* e pela preservação por meio de ligação telefónica e através de email e também através de contacto pessoal. Procurou realizar-se contactos destes tipos de forma sucessiva durante dois meses e meio. Mas foi muito difícil obter qualquer espécie de respostas – sobretudo respostas suficientemente claras e definitivas. Em relação a cada jornal, em particular, percebeu-se o seguinte:

- o CM não possui um arquivo digital público, embora exista a possibilidade de que os seus conteúdos possam estar preservados em algum lugar no escopo da própria redação, já que o último contacto apontou o editor multimédia como a pessoa que poderia explicar como se processa tal preservação;
- o Arquivo e Centro de Documentação do Global Media Group tem uma equipa responsável pelos materiais digitais, mas não se não obteve resposta de ninguém nesta equipa;
- a Gesco da Impresa não preserva os materiais digitais do *Expresso*, mas o editor multimédia deste jornal afirmou que possuem uma base de dados digital interna onde o conteúdo do *website* é protegido;
- o Centro de Documentação do *Público* não preserva conteúdos digitais e as questões relativas à preservação de materiais digitais são alheias aos responsáveis.

O único aspecto transversal em todos os contactos foi o da percepção de que a referência a preservação do material jornalístico é imediatamente associada à

preservação de “conteúdos” e não de toda a estrutura das páginas web. Se alguém quisesse ver algum texto ou alguma imagem publicada nos *websites* desses jornais durante 1996, 1997 ou 1998 talvez fosse possível, mas seria impossível aceder ao todo do *website* como ele se apresentaria nesse tempo.

A súmula destes resultados sugere que o CM, o DN, o *Expresso* e o *Público* não têm uma cultura de preservação digital desenvolvida e informada, nem possuem políticas de preservação digital estruturadas. Graças a isso, o que esteve nos seus *websites* pode estar perdido para sempre e o que ainda neles se encontra pode estar prestes a perder-se. Parece assim justificar-se que se fale de problemas de memória no jornalismo português.

5 Estabelecimento de uma rota de investigação

Uma maneira de resolver os problemas levantados pela navegação num arquivo da web consiste no desenvolvimento de uma visão comparativa dos elementos das páginas estudadas. A partir dessa comparação, pode elaborar-se um “índice” das versões

selecionadas para estudo e, a partir desse índice, proceder a uma investigação do material (Brügger, 2016). Como elaborar este índice? Neste projecto, o índice foi pensado a partir da perspectiva de que a investigação está ligada à estrutura específica dos arquivos da web utilizados e aos seus próprios meios técnicos. Estes meios podem, em algum momento, ir para além das barreiras da estrutura de um arquivo. Por exemplo, o Arquivo.pt tem uma funcionalidade que permite completar as páginas web com os elementos em falta através do "Memento Time Travel Web Archive Replay", um *software* que cruza dados preservados em diferentes arquivos em relação ao mesmo momento de um sistema, procurando reconstituí-lo. Não obstante esta possibilidade, a investigação num arquivo implica um envolvimento em termos de comandos de utilizador na interface gráfica do arquivo, operações ocultas do arquivo e respostas perceptíveis pelo utilizador. Este envolvimento é uma espécie de imersão na própria estrutura do arquivo utilizado. É com os meios desse arquivo que se fazem as perguntas e através dele que se obtêm as respostas. A imersão traduz-se nesse condicionamento. Assim, investigar no arquivo acaba por ser uma actividade de navegação muito específica sobretudo numa investigação que se debruça sistematicamente sobre os mesmos assuntos, como no caso do presente projecto. Poderá pensar-se esta navegação reconhecendo na ideia de "navegar na internet" uma metáfora com utilidade perceptiva e, estendendo essa metáfora, referir-se o trajecto específico dentro do arquivo da web como uma *rota de investigação* materializadora, no acto de investigação, da forma do índice anteriormente referido.

A metáfora da navegação é coincidente com a ideia de "ciberespaço". Pode ter-se uma ideia do que tal significa se se atender ao sentido do termo "cibernética", que foi cunhado por Norbert Wiener a partir do grego para se referir à arte do homem do leme e da boa vela (Wiener, 1954). Neste sentido, o ciberespaço é um "espaço navegável" por pessoas diferentes, as quais podem estar simplesmente à deriva ou a procurar desenhar uma rota concreta para fazerem o seu caminho (Manovich, 2001). É neste quadro que se deve compreender o conceito de *rota de investigação*, que corresponde à delimitação dos dados disponíveis num trajecto muito específico. Indo para além dos dados já sinalizados pela ideia de índice, a *rota de investigação* dá conta de um nível de experiência de utilização dos arquivos. Como no caso de uma viagem

marítima a terras desconhecidas, é necessário projectar e desenhar uma rota navegável, cujas possíveis falhas são corrigidas durante a própria navegação através de afinamentos dos instrumentos de navegação, da observação, do cálculo e também de um certo grau de intuição. A *rota de investigação* é, assim, um desenho um curso na internet através de *websites* de alguma forma relacionados entre si. Este conceito não fecha alterações em investigações futuras, o que ele faz é colocar uma posição intermediária de cartografia para ajudar a encontrar o melhor caminho entre as formas possíveis de definir o corpus de análise. Na prática, a *rota de investigação* é desenhada através de conexões entre *websites* ao longo do tempo, dando prioridade aos *websites* que possuem a maioria dos dados disponíveis e vendo como complementar aqueles que possuem dados que os anteriores não têm. Esta síntese crescente acaba, portanto, por incluir certos dados e excluir outros. É mesmo possível que ao estabelecer a rota de investigação, se excluam páginas web importantes para a presente investigação. No entanto, esta parece ser a melhor solução, atendendo aos métodos disponíveis.

5.1 Metodologia

A realização do projecto dependia profundamente do isolamento do material para estudo entre todos os materiais disponíveis tendo em conta os parâmetros mencionados. O ponto de partida para se desenhar a *rota de investigação* foi descobrir os endereços dos *websites* disponíveis no Arquivo.pt para cada jornal. O foco foi colocado apenas nas "páginas principais", excluindo, portanto, outras páginas web. Utilizaram-se três formas para pesquisar os endereços dessas páginas:

- consultou-se uma lista interna do Arquivo.pt, fornecida pelo responsável pelo arquivo, com endereços de entidades web cujo material está preservado no arquivo, designadamente *websites* dos jornais, trata-se de uma lista não exaustiva que começa apenas em 2014, mas que sinaliza alguns endereços;
- utilizou-se o sistema de pesquisa por termo do Arquivo.pt para procurar por entradas com os termos "Correio da Manhã", "CM", "Diário de Notícias", "DN", "Expresso" e "Público" entre 01/01/1996 e 31/12/2016;

- utilizou-se o Sistema de pesquisa por endereço do Arquivo.pt para procurar por entradas com os actuais endereços dos *websites* dos jornais (já acima mencionados) entre 01/01/1996 e 31/12/2016;

Para elaborar uma *rota de investigação* rigorosa foi necessário efectuar uma pesquisa exaustiva do material disponível. Para realizar isto, elaborou-se uma base de dados na forma de tabelas para relacionar dados sobre os endereços encontrados. Estas tabelas documentam e ordenam o material disponível. Com a base de dados, seria possível desenhar uma rota mais orientada funcionando como uma espécie de mapa de dados preservados, realmente disponíveis e relacionados. As tabelas para formar a base de dados foram elaboradas com base nas seguintes categorias:

- "Endereço", que contém a indicação de um endereço descoberto;
- "Número de versões", onde se assinala o número de versões disponíveis no Arquivo.pt de cada um dos endereços;
- "Primeira versão", na qual se indica a data da primeira versão disponível no Arquivo.pt de cada um desses endereços;
- "Última versão", na qual se indica a data da última versão disponível no Arquivo.pt de cada um desses endereços;
- "Anos com maior número de versões", onde se sugere o ano ou anos com mais versões disponíveis de um endereço por comparação a outros endereços;
- "Notas posteriores", no âmbito da qual se expõem observações resultantes de uma análise mais avançada dos materiais de investigação, análise apenas tornada possível depois do terceiro momento de análise deste projecto (correspondente ao próximo capítulo do relatório).

Esta última categoria não foi planeada de raiz, ela surgiu como um auxiliar na realização da perspectiva filológica de Brügger. Num primeiro momento, apenas se percorreram as páginas correspondentes a certos endereços em certas datas. Logo aqui, foi possível descobrir e sinalizar novos endereços a acrescentar aos previamente descobertos. Neste percurso, foram ignoradas todas as páginas com anomalias graves ao ponto de inibirem a sua reprodução. Num momento posterior, o material foi ajustado pela análise de imagem que se explicará depois. Assim, a última categoria é um indicador que expressa uma certa dialéctica desta investigação mostrando que a delimitação de

uma *rota de investigação* não está finalizada enquanto não se tiver em conta a disposição efectiva dos *websites* e das páginas web com os seus elementos preservados e não preservados.

5.2 Análise de resultados

Considerando o CM, com a lista do Arquivo.pt descobriram-se os endereços cmjornal.xl.pt e correiomanha.pt. Com a pesquisa por termos, descobriu-se o endereço correiodamanha.xl.pt. Durante a navegação descobriu-se o endereço correiodamanha.pt/index.asp. O endereço actual é cmjornal.pt.

Endereço	Número de versões	Primeira versão	Última versão	Anos com maior número de versões	Notas posteriores
cmjornal.xl.pt	1.567 versões	23 de Março de 2010	21 de Maio de 2016	2010, 2011, 2012, 2013, 2014, 2015	De 29 de Dezembro de 2011 a 3 de Fevereiro de 2012, as versões deste endereço têm problemas de preservação graves.
correiomanha.pt	1.203 versões	1 de Novembro de 1996	29 de Setembro de 2015	2004, 2005, 2006, 2007, 2010	De 17 de Outubro de 2000 a 1 de Junho de 2002, as versões deste endereço têm problemas de preservação. No entanto, neste caso pode ter-se em conta os dados deste período porque o material disponível permite diferenciá-lo em relação a outros. De 27 de Maio de 2004 a 7 de Janeiro de 2005, este endereço apresenta versões com uma hiperligação inserida na frase "Saltar para a Homepage" ou, dependendo da data, na frase ""Entrar directamente em www.correiomanha.pt". Esta hiperligação redirecciona para o endereço correiodamanha.pt/index.asp. De 13 de Outubro de 2005 a 28 de Dezembro de 2005, este endereço apresenta versões com uma hiperligação inserida na frase "Entrar directamente em www.correiomanha.pt". Esta hiperligação redirecciona para

					o endereço correiodamanha.pt/index.asp. Este foi o único endereço onde se encontraram páginas diferentes em horas diferentes. Em algumas horas durante certos períodos havia hiperligações como essas a dirigir para outros endereços, noutras horas durante os mesmos período não havia hiperligações, mas a página para no outro caso se era redireccionado era logo aberta. De 23 de Março de 2010 a 29 de Setembro de 2015, as versões deste endereço redirecionam para versões com o endereço cmjornal.xl.pt. No dia 20 de Maio de 2009, este endereço redirecciona para versão com o endereço correiodamanha.pt/#.
correiodamanha.xl.pt	182 versões	12 de Julho de 2007	24 de Abril de 2015	2010	De 23 de Março de 2010 a 24 de Abril de 2015, as versões deste endereço redirecionam para versões com o endereço cmjornal.xl.pt
cmjornal.pt	20 versões	27 de Março de 2010	24 de Abril de 2015	-	Cada versão deste endereço redirecciona para versões com o endereço cmjornal.xl.pt
correiodamanha.pt/index.asp	156 versões	2 de Julho de 2002	24 de Abril de 2015	2005, 2006, 2007	-
TOTAL DE VERSÕES	3. 128 versões contadas				

A análise e comparação do material referido permitiu elaborar a seguinte rota de investigação:

- correiomanha.pt de 1996 a 2009;
- complementarmente, teriam de ser observadas as versões do endereço correiodamanha.pt/index.asp de 27 de Maio de 2004 a 7 de Janeiro de 2005 e de 13 de Outubro de 2005 a 28 de Dezembro de 2005;
- cmjornal.xl.pt de 2010 a 2016;
- adicionalmente, teriam de ser observadas as versões do endereço correiodamanha.xl.pt de 12 de Julho de 2007 a 23 de Março de 2010.

Considerando o DN, com a lista do Arquivo.pt foram descobertos os endereços dn.pt e dn.sapo.pt. Na navegação pelos endereços descobriram-se os endereços http://dn.sapo.pt/homepage/homepage.asp,dn.lusomundo.net,dn.sapo.pt/homepage/homepage.asp,dn.pt/inicio/default.aspx,dn.sapo.pt/inicio,dn.sapo.pt/inicio/default.aspx, dn.sapo.pt/index.html e dn2.sapo.pt/homepage/homepage.asp. Na pesquisa por termo, não se descobriram endereços para além de dn.pt e dn.sapo.pt. O endereço actual é dn.pt. Enquanto o processo de investigação estava em curso, os engenheiros do Arquivo.pt descobriram novas versões relacionadas com DN no endereço dn.pt/inicio.

Endereço	Número de versões	Primeira versão	Última versão	Anos com maior número de versões	Notas posteriores
dn.pt	780 versões	13 de Outubro de 1996	21 de Maio de 2016	2001, 2004, 2005, 2010, 2015	De 1 de Outubro de 2001 a 3 de Outubro de 2001, este endereço apresenta uma página em branco com uma hiperligação inserida na frase "Novo site do DN". O material da hiperligação não está arquivado. De 6 de Outubro de 2001 até ao final de 2001, este endereço redirecciona para versões do endereço dn.lusomundo.net. De 24 de Janeiro de 2002 a 19 de Outubro de 2004, este endereço redirecciona para versões do endereço dn.sapo.pt/homepage/homepage.asp. De 22 de Outubro de 2004 a 22 de Outubro de 2008, este endereço redirecciona para versões do endereço dn.sapo.pt. De 20 de Maio de 2009 a 19 de Janeiro de 2011, este endereço redirecciona para versões do endereço dn.sapo.pt/inicio. De 22 de Maio de 2011 a 1 de Julho de 2015, este endereço redirecciona para versões do endereço dn.pt/inicio/default.aspx.

					Não há praticamente nenhum material preservado no âmbito deste endereço em 2011, 2012, 2013 e 2014. Depois de 14 de Outubro de 2015, as versões disponíveis têm problemas de preservação graves.
dn.sapo.pt	1.274 versões	14 de Dezembro de 2001	9 de Março de 2012	2004, 2005, 2006, 2007, 2010	De 26 de Janeiro de 2002 a 21 de Setembro de 2004, este endereço redirecciona para versões no endereço dn.sapo.pt/homepage/homepage.asp. De 7 de Janeiro de 2006 a 29 de Agosto de 2006, as versões deste endereço têm problemas de preservação graves. De 20 de Maio de 2009 a 18 de Outubro de 2010, este endereço redirecciona para versões do endereço dn.sapo.pt/inicio/. A 8 de Janeiro de 2011, este endereço redirecciona para versões do endereço dn.sapo.pt/inicio/default. A 9 de Janeiro de 2011, este endereço redirecciona para versões do endereço dn.sapo.pt/inicio/. A 19 de Janeiro de 2011, este endereço redirecciona para versões do endereço dn.pt/inicio. Depois de 22 de Maio de 2011, este endereço redirecciona para versões do endereço dn.pt/inicio/default.aspx.
dn.sapo.pt/homepage/homepage.asp	190 versões	17 de Janeiro de 2002	25 de Março de 2012	2004	De 22 de Outubro de 2004 a 4 de Abril de 2005, este endereço redirecciona para versões no endereço dn.sapo.pt/index.html. A 9 de Dezembro de 2004 este endereço redirecciona para versões do endereço dn2.sapo.pt/homepage/homepage.asp.

					De 4 de Fevereiro de 2005 a 4 de Abril de 2005, este endereço redirecciona para versões do endereço dn.sapo.pt/index.html. A 13 de Dezembro de 2005, este endereço redirecciona para versões do endereço dn2.sapo.pt/homepage/homepage.asp. De 2 de Janeiro de 2006 a 21 de Outubro de 2008, este endereço redirecciona para versões do endereço dn.sapo.pt/index.html. Depois de 20 de Maio de 2009, este endereço apresenta páginas com anomalias.
dn.lusomundo.net	0 versões	-	-	Sem significados significativos	Este endereço não tem materiais arquivados no Arquivo.pt.
dn.pt/inicio	1.368 versões	19 de Janeiro de 2011	31 de Dezembro de 2015	2011, 2012, 2013, 2014, 2015	Depois de 21 de Janeiro de 2011, as versões arquivadas têm problemas de preservação tão graves que se torna muito difícil garantir que aquilo que se vê nas versões arquivadas esteja próximo daquilo que estaria na web primeiramente devido a problemas de preservação. Deste modo, as versões deste endereço não serão tidas em conta para a análise.
dn.pt/inicio/default.aspx	985 versões	8 de Janeiro de 2011	25 de Dezembro de 2015	2011, 2012, 2013, 2014, 2015	A 8 de Janeiro de 2011, este endereço redirecciona para versões do endereço dn.sapo.pt/inicio/default/aspx. Todas as versões arquivadas deste endereço têm problemas de preservação.
dn.sapo.pt/inicio/	633 versões	30 de Setembro de 2009	9 de Março de 2012	2010, 2011	Depois de 14 de Junho de 2011, este endereço redirecciona para versões do endereço dn.pt/inicio.
dn.sapo.pt/inicio/default.aspx	1 versão	8 de Janeiro de 2011	8 de Janeiro de 2011	Sem resultados significativos	-

dn.sapo.pt/index.html	327 versões	23 de Outubro de 2004	22 de Maio de 2011	2004, 2005	Depois de 1 de Outubro de 2009, este endereço apresenta páginas com anomalias.
dn2.sapo.pt/homepage/homepage.asp	10 versões	18 de Novembro de 2004	25 de Março de 2005	Sem resultados significativos	-
TOTAL DE VERSÕES	5.568 versões contadas				

A análise e comparação do material referido permitiu elaborar a seguinte rota de investigação:

- dn.pt de 1996 a 2002;
- dn.sapo.pt/homepage/homepage.asp de 2002 a 2009;
- dn.sapo.pt/inicio/ de 2009 a 2011;
- dn.pt/inicio em 2011.

Considerando o *Expresso*, com o apoio da lista de endereços do Arquivo.pt descobriram-se os endereços clix.expresso.pt, expresso.sapo.pt e aeiou.expresso.pt. Na pesquisa por termos do arquivo, encontraram-se os endereços www.expresso.pt e expresso.clix.pt. O endereço actual é expresso.sapo.pt. Na navegação encontraram ainda os seguintes endereços: actualizacaodiaria.expresso.pt/primeira,online.expresso.pt/ e semanal.expresso.clix.pt.

Endereço	Número de versões	Primeira versão	Última versão	Anos com maior número de versões	Notas posteriores
clix.expresso.pt	858 versões	8 de Dezembro de 2007	1 de Julho de 2015	2010, 2011, 2012	Depois de 2008, as versões deste endereço têm problemas de preservação.
expresso.sapo.pt	795 versões	12 de Março de 2012	21 de Maio de 2016	2012, 2013, 2014, 2015	Depois de 5 de Maio de 2015, as versões apresentadas têm problemas muito graves de preservação. Estas versões não serão por isso consideradas.
aeiou.expresso.pt	746 versões	18 de Outubro de 2007	15 de Maio de 2016	2010, 2011, 2012	Antes de 2008, as versões deste endereço têm problemas de preservação. Depois de 12 de Março de 2002, este endereço redirecciona para versões do endereço expresso.sapo.pt.
expresso.pt	511 versões	10 de Janeiro de 1998	17 de Maio de 2016	2000, 2001, 2004, 2010	De 16 de Agosto de 2000 a 19 de Outubro de 2000, este endereço redirecciona para versões do

					endereço actualizacaodiaria.expresso.pt/primeira. Nos dias 4 e 5 de Março de 2001, este endereço apresentar versões muito peculiares. De 18 de Abril de 2001 a 19 de Maio de 2001, este endereço redirecciona para versões do endereço online.expresso.pt/. De 29 de Setembro de 2003 a 18 de Dezembro de 2003, este endereço redirecciona para versões do endereço online.expresso.pt/.
expresso.clix.pt	561 versões	3 de Junho de 2004	24 de Abril de 2015	2004, 2005, 2006, 2007	De 29 de Junho de 2004 a 1 de Setembro de 2006, este endereço redirecciona para versões do endereço online.expresso.clix.pt/. A 1 de Novembro de 2004, o endereço expresso.clix.pt redirecciona para versões do endereço semanal.expresso.clix.pt. Depois de Julho de 2006, a data apresentada em muitas versões deste endereço não bate certo com a data apresentada na interface do Arquivo.pt.
actualizacaodiaria.expresso.pt	5 versões	24 de Agosto de 2000	15 de Maio de 2001	Sem resultados significativos	Este endereço conduz ao que parecem ser páginas complementares ou secundárias da actualização diária do *website*. Portanto, as versões deste endereço não serão tidas em conta para a análise.
online.expresso.pt/	220 versões	30 de Março de 2001	24 de Abril de 2015	2001	Quando a data indicada na interface no arquivo é a mesma, as versões deste endereço são semelhantes às versões do endereço expresso.pt. por isto, as versões deste endereço não serão tidas em conta para a análise, excepto no caso em que para elas se é redireccionado desde o endereço expresso.pt, como acontece no final de 2003.
semanal.expresso.clix.pt	240 versões	26 de Junho de 2004	2 de Junho de 2010	2004, 2005, 2006	Este endereço conduz ao que parecem ser páginas complementares ou secundárias da actualização semanal do *website*. Portanto, as versões deste endereço não serão tidas em conta para a análise.

online.expresso.clix.pt/	208 versões	26 de Junho de 2004	10 de Junho de 2010	2004, 2005, 2006	Depois de Julho de 2006, , a data apresentada em muitas versões deste endereço não bate certo com a data apresentada na interface do arquivo.
TOTAL DE VERSÕES	4.144 versões contadas				

A análise e comparação do material referido permitiu elaborar a seguinte rota de investigação:

- expresso.pt de 1998 a 2004;
- expresso.clix.pt de 2004 a 2008;
- complementarmente, clix.expresso.pt em 2008;
- aeiou.expresso.pt de 2008 a 2012;
- expresso.sapo.pt de 2012 a 2015.

Considerando o *Público*, com a lista interna do Arquivo.pt mostrou os seguintes endereços: digital.publico.pt, jornal.publico.pt, pdf.publico.pt e publico.pt. Na pesquisa de termos, não se encontrou nenhum resultado para além desses endereços. O endereço actual é publico.pt. Na navegação, descobriram-se os endereços publico.clix.pt e ww2.publico.clix.pt.

Endereço	Número de versões	Primeira versão	Última versão	Anos com maior número de versões	Notas posteriores
digital.publico.pt	618 versões	12 de Outubro de 2006	21 de Novembro de 2011	2010, 2011	De 29 de Outubro de 2010 a 16 de Novembro de 2011, as versões destes endereço têm problemas de preservação. Para além disto, as versões disponíveis neste endereço são semelhantes às versões do endereço publico.pt. Assim sendo, as versões deste endereço não serão tidas em conta para a análise.

jornal.publico.pt	1.910 versões	8 de Abril de 2000	15 de Março de 2016	2010, 2011, 2012, 2013, 2014 and 2015	As versões deste endereço são semelhantes às versões do endereço publico.pt. Como o ultimo será o endereço principal, as versões do endereço jornal.publico.pt não serão tidas em conta para a análise.
pdf.publico.pt	648 versões	26 de Novembro de 2002	22 de Novembro de 2011	2005	Este endereço conduz a uma página com os ficheiros PDF. das versões impressas do jornal. Assim sendo, as versões deste endereço não serão tidas em conta para a análise.
publico.pt	3.879 versões	13 de Outubro de 1996	20 de Maio de 2016	2010, 2011, 2012, 2013, 2014, 2015	De 19 de Maio de 2005 a 18 de Dezembro de 2009, este endereço redirecciona para versões do endereço publico.clix.pt. De 29 de Outubro de 2010 a 16 de Novembro de 2011, há problemas de preservação. Depois de 16 de Outubro de 2015, todas as versões deste endereço redirecionam para a versão actual e para páginas recentes do *website* do jornal e não para supostas versões arquivadas.
publico.clix.pt	1.145 versões	22 de Fevereiro de 2005	25 de Maio de 2016	2005, 2006	A partir 6 de Junho de 2007, este endereço redirecciona para versões do endereço ww2.publico.clix.pt. A partir de 2010, este endereço redirecciona para versões do endereço publico.pt.

ww2.publico.clix.pt	108 versões	19 de Maio de 2007	4 de Agosto de 2010	2007	-
TOTAL DE VERSÕES	8.308 versões contadas				

A análise e comparação do material referido permitiu elaborar a seguinte rota de investigação:

- publico.pt de 1996 a 2005;
- publico.clix.pt de 2005 a 2007;
- complementarmente, ww2.publico.clix.pt em 2007;
- publico.clix.pt de 2007 a 2010;
- publico.pt de 2010 a 2016.

A pesquisa para constituição de uma rota dentro do Arquivo.pt revelou a existência dos problemas habituais do material renascido digital dos arquivos da web no caso dos jornais estudados: há pouco material guardado, mas uma enorme quantidade de dados e de informações disponíveis sobre o que foi preservado, muitas páginas e respectivos endereços são difíceis de encontrar porque estão tapados pelas operações de redireccionamento, alguns materiais não têm data correspondente à data prometida na interface do Arquivo.pt, e muitos dos materiais arquivados são problemáticos ao ponto de não poderem ser estudados enquanto "representações" dos *websites* tal como eles teriam sido.

Ainda assim, há material suficiente para conduzir investigações desde que estas estejam suficientemente enquadradas teoricamente em relação aos objectos digitais disponíveis nos arquivos da web. Com esse enquadramento em mente, é possível delimitar os dados progressivamente e criar uma rota de investigação por entre o fluxo dos materiais crus e poucos ordenados que estão arquivados. É muito importante conhecer o arquivo da web como conceito e problematizá-lo e depois conhecer as estruturas específicas do arquivo com que se trabalha para se fazer uma investigação. O facto de que quando este projecto científico foi elaborado se tenham apenas encontrado o total de 3.166 versões dos *websites* desses jornais arquivados no Arquivo.pt e agora se tenham isolado 21.148 versões ilustra bem esta necessidade de conhecer os arquivos da web.

A ideia de *rota de investigação* pode, contudo, ser aproveitada para organização, sistematização e indexação manual dos materiais disponíveis de modo a criar colecções digitais sem os problemas originados pela automacidade da indexação do arquivo da web. Por exemplo, colecções que organizem os jornais por ordem dos seus *layouts*. Ou, outro exemplo, colecções de conteúdos relativos a temas jornalísticos específicos, como uma colecção sobre o caso Apito Dourado, onde se poderiam consultar todas as peças sobre o assunto. Estas colecções constituiriam o ponto de partida para análises de discurso, análises de conteúdo, análises de designs, etc.

O *website* do *Público* apresentou um problema singular neste estudo. Depois de 16 de Outubro de 2015, não há versões arquivadas desse *website*, porque todas as supostas versões redirecionam para páginas web actuais e não para páginas descontinuadas. Ou seja, a partir dos finais de 2015 o que se vê são páginas da data em que se consulta o arquivo. Assim que no decorrer da investigação se encontrou este problema, ele foi dado a conhecer aos engenheiros do Arquivo.pt, que rapidamente procuraram esclarecer a situação. Trata-se de um problema de base técnica a que se chama "fuga para a web viva" (*live-web leak*). O Internet Archive enfrenta o mesmo problema no caso do *Público*. Portanto, o problema é mesmo do *website* desse jornal. Alguém terá introduzido um código que inibiu completamente a recolha automática da web, ofuscando o material do *website*. Apesar desta investigação não cobrir o ano de 2017, neste ano observou-se que a anomalia já foi corrigida (o que significa que alguém se apercebeu de que algo estava mal!). Agora, se é verdade que o material de 2017 poderá estar salvo por ser recolhido pelos arquivos da web, o material de finais de 2015 e de todo o ano de 2016 pode ter-se perdido definitivamente.

Os endereços encontrados e as mudanças descobertas são muito valiosos do ponto de vista da análise social. Nessas mudanças vamos vendo como se monta o cenário político-económico das empresas de *media* em Portugal na sua relação com as empresas de telecomunicações. No CM, apenas há indicação do próprio grupo Cofina (xl.). No DN, aparecem a Lusomundo e a Sapo. No *Expresso*, surgem a Clix, o AEIOU e a Sapo. No *Público*, a Clix.

6 Análise diacrónica descritiva das páginas principais dos *websites* dos jornais de 1996 a 2016

Para realizar a análise diacrónica, é necessário que existam elementos comparativos que mudem ou, pelo menos, que tenham o potencial de mudar ao longo do tempo. A forma que existe de se chegar aos elementos em mudança dos *websites* é utilizando a interface gráfica do arquivo da web, neste caso do Arquivo.pt, num ecrã de computador alcançando o conjunto desses elementos como diferentes tipos de conjuntos. Quando falamos da mudança dos *websites*, são todos os elementos de todos os níveis que estão a mudar ao longo do tempo, porque mudam os *hardwares* e os *softwares* e muda o que se diz, o que se escreve e o que se dá a ver. Se nos focarmos naquilo a que Brügger chama textualidade, estaremos a caminho de fazer uma análise de conteúdo ou uma análise de discurso. É um desafio interessante para os arquivos da web. Trata-se de procurar perceber que tipos de textos e de imagens é que circulam em sociedade ou quais são pontos de vistas sobre um determinado assunto em debate no espaço público ou, ainda, como é a linguagem nos *media* veicula certas ideias e como contribui para a manutenção ou não do *status quo*, por exemplo, em relação à nacionalidade ou ao género (*vide* Richardson, 2007). Um exemplo interessante de estudo desta espécie seria a análise de discurso de grandes casos que passaram pelos tribunais portugueses e que tiveram um forte impacto na opinião pública, como os casos Apito Dourado ou Banco Espírito Santo. Com uma análise desse género perceber-se-ia como se construíram as tramas, quem é que teve que papel, como se estabeleceram as relações de poder.

Por muito interessantes que tais perspectivas sejam, não permitem responder à questão de como mudaram os *websites* dos jornais como um todo (quando muito respondem a questões relacionadas com o modo como mudou a sociedade espelhada, na medida em que o pode ser, nesses *websites*). Mais próximas do objectivo de partida deste projecto estão perspectivas do campo de estudos a que se tem dado o nome de Estudos Visuais, particularmente contributos da Semiótica e do Design das interfaces gráficas e da informação. Os Estudos Visuais são um campo interdisciplinar que pensa dispositivos visuais e a experiência que proporcionam a partir da sua especificidade material, conceptual e histórica (Smith, 2008). Estas perspectivas têm tomado o conceito de interface como instância dinâmica de relação entre o humano e a máquina

com disposições e mutações técnicas e estéticas específicas. Como se afirmou previamente, os ecrãs (e também as interfaces e as imagens) têm uma dimensão técnica, mas também uma dimensão social e semiótica. No âmbito do Design, tem sido dado um tratamento sobretudo descritivo dos elementos que emergem destas dimensões. Em Espanha vários estudos têm apresentado descrições de mudança nos *websites* dos jornais em termos de recursos audiovisuais e multimédia e de "elementos estruturais", como barras, colunas, topografias e cores (Esteruelas, 2014). A mudança dos *layouts* dos *websites* pode, nesse quadro, ser pensada como algo que decorre da adaptação das empresas de *media* às mudanças tecnológicas através de modificação de estruturas e de conteúdos nomeadamente através da integração de novos recursos e aspectos gráficos e do estabelecimento de novas relações com a publicidade (*ibid*.). Num estudo importante sobre os principais jornais americanos, John Nerone e Kevin G. Barnhurst investigaram as formas visuais das edições *online* desses jornais (Nerone & Barnhurst, 2001). Nesse estudo, mostraram que os *websites* dos jornais possuem estruturas distintivas em relação àquelas que existem na imprensa escrita. Nesses *websites* dominam as "lógicas" da hiperligação, dado que quase tudo é clicável e no acto de clique conducente a outros domínios da web, e da publicidade, visto que há colocação de anúncios em *banners* espalhados por toda a estrutura as colunas e das barras dos *websites*. Tomando uma posição crítica, os autores consideram que o mecanismo de clique entra em ruptura com a lógica da edição editorial previamente prevalecente através do web design, com interesses económicos específicos (referem-se, por exemplo, à contagem dos cliques). Para os autores, o *design* dos jornais encaminha-se por uma via que consideram ser "pós-moderna" e "neo-vitoriana" (*ibid*.). "Todos adoptam um vocabulário promocional para impulsionar os acontecimentos até aos consumidores", dizem os autores, "vendendo a sua carga moral em vez do discurso ponderado da sociedade civil" (*idem*: 480).

6.1 Metodologia

Com recurso a ferramentas da web disponíveis *online*, seria possível analisar o material dos *websites* dos jornais disponíveis no Arquivo.pt para se elaborar um estudo descritivo pormenorizado e exaustivo de todos os elementos que compõem o *layout* desses

websites. Isto seria o ideal para se responder às questões de partida. Contudo, a grande quantidade de dados condicionou a investigação na medida em que obrigou a que grande parte do trabalho se debruçasse na elaboração da *rota de investigação* como tal. A descrição pormenorizada e exaustiva das estruturas dos *websites* tomaria mais tempo do que aquele de que se dispôs neste projecto científico. Repara-se que, como já se afirmou, inicialmente se tinham apontado apenas 3.166 versões dos *websites* dos jornais sob estudo arquivados no Arquivo.pt e que, na verdade, se conseguiram isolar 21.148 versões (estando apenas a falar de páginas principais dos *websites*!). Com este número, bastaria que cada página web tivesse duas colunas para ser necessário medir e analisar mais de 42.000 colunas. Sendo que o número de colunas é bastante maior, sobretudo depois de 2000, tal descrição ao pormenor de todas as páginas não foi possível.

A alternativa encontrada para proceder à realização do projecto científico e resolver as questões iniciais é a da comparação geral dos *websites* na medida em que a interface dos arquivos os fornece como mensagens visuais. Em vez de uma análise pormenorizada, aposta-se então no poder ilustrativo dos *snapshots* convertidos em *browsershots*. Embora se mantenha a atenção em relação às cores e às formas, a análise estética estaria, deste modo, menos focada no carácter matemático da distribuição dos seus elementos do que numa perspectiva da sua na "composição", isto é, na "geografia interior da mensagem visual", fundamental na construção e hierarquização da mensagem e correspondente orientação da sua leitura (Joly, 1999). A partir desta perspectiva, procura assinalar-se os períodos de continuidade e de descontinuidade dos *layouts* dos jornais através de algumas descrições sumárias de uma página representativa de um certo período de continuidade. Posteriormente, procura-se atentar ao carácter social dessas mudanças.

6.2 Análise de resultados

A comparação entre as páginas web do(s) *website*(s) corresponde(s) ao CM encontradas na rota de investigação elaborada a partir do material desse jornal disponível no Arquivo.pt permite isolar dez desenhos diferentes de *layouts* desse jornal:

- o primeiro destes desenhos é visível no material arquivado com as datas que vão de 1 de Novembro de 1996 a 11 de Abril de 1997 no contexto do endereço correiomanha.pt, um exemplo deste desenho é o da versão apresentada na Figura 1, em que se apresenta um fundo vermelho que ocupa quase todo o campo visual, com uma coluna de pesquisa branca, o nome do jornal em letras gordas em branco e várias ilustrações de jornais animados com indicações de divisões de navegação (um quasi-menu);

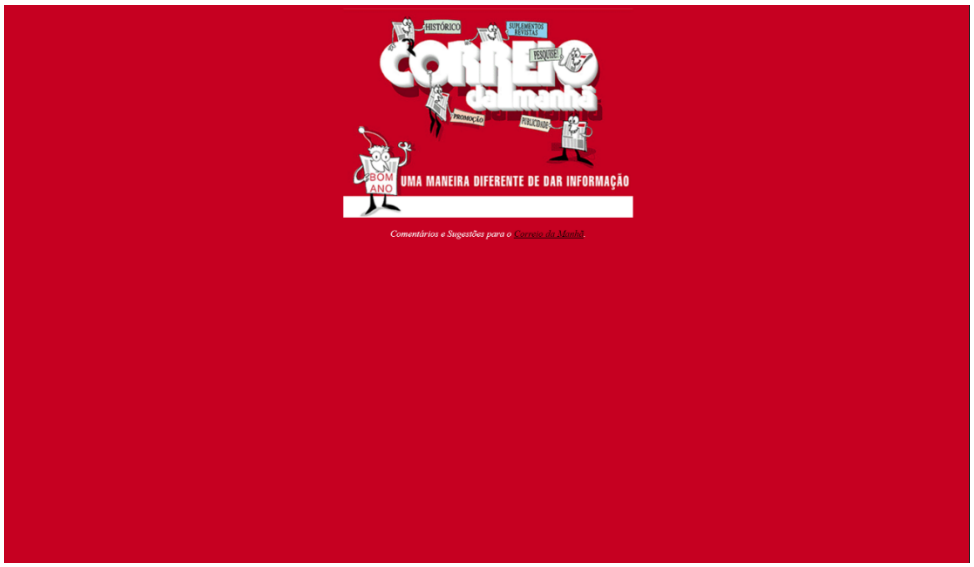

Figura 1 – Captura de ecrã de 1 de Novembro de 1996 (correiomanha.pt)

- um segundo desenho pode encontrar-se no âmbito do endereço correiomanha.pt, do período que vai de 11 de Dezembro de 1997 a 22 de Maio de 1998, um exemplo deste desenho encontra-se na Figura 2, o fundo é branco, destacam-se quatro grandes zonas horizontais, uma com um número em branco num fundo vermelho, outra que se divide entre um espaço para jornais animados idênticos aos anteriores e o nome do jornal em branco num fundo vermelho e outro com capas diárias, uma terceira zona com indicação dos suplementos que acompanham o jornal conforme o dia da semana, e uma quarta zona com os jornais animados a convidarem ao envio de "sugestões e comentários";

Figura 2 – Captura de ecrã de 22 de Maio de 1998 (correiomanha.pt)

- em correiomanha.pt, a partir de 3 de Dezembro de 1998 até 15 de Agosto de 2000, há um terceiro desenho, de que a versão apresentada na Figura 3 serve de exemplo, em que se encontra uma leitura vertical dividida em três colunas, uma vermelha com as categorias de navegação do *website*, uma branca com um *banner* horizontal sobre o próprio CM e com os textos informativos, e uma preta com alguns destaques do jornal e símbolos que indiciam a existência de algum tipo de publicidade comercial (referência à empresa de telecomunicações Optimus, por exemplo), nota-se que este *layout* tinha uma extensão vertical muito maior que os anteriores;

Figura 3 – Captura de ecrã de 3 de Março de 2000 (correiomanha.pt)

- o quarto destes desenhos encontra-se no material arquivado sob o endereço correiomanha.pt e respeitante ao período entre 17 de Outubro de 2000 e 1 de Junho de 2002, de que a captura de ecrã apresentada na Figura 4 é exemplar, na qual se observa ma organização claramente diferente embora seja difícil de caracterizar, dada a falta de dados apresentados, mas pode afirmar-se que, ainda que mantenha a divisão em três colunas verticais, a central tem mais textos acompanhados por aquilo que parecem ter sido imagens;

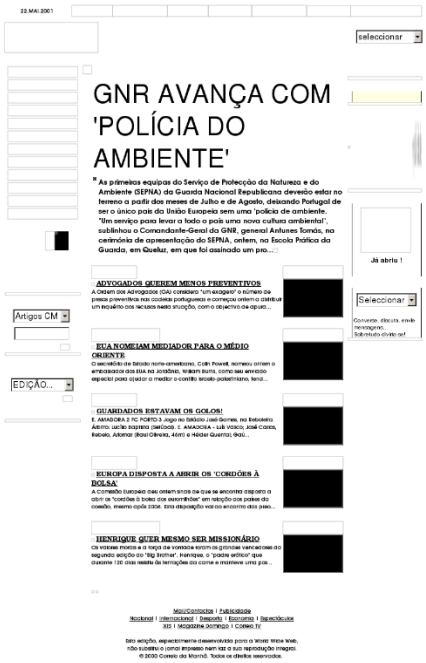

Figura 4 – Captura de ecrã de 22 de Maio de 2001 (correiomanha.pt)

- em correiodamanha.pt, de 2 de Agosto de 2002 a 23 de Junho de 2003, encontra-se um quinto desenho, de que a Figura 5 serve de exemplo, na qual se pode verificar que se mantém a divisão do espaço em três grandes zonas, mas que há mais distinções, dado que há uma espaço superior horizontalmente orientado onde está uma pequena barra com hiperligações para outras peças e que na parte central onde estão as notícias a apresentação é feita não como uma lista mas em quadrantes organizados tematicamente, para além de que o *website* adopta como cores apenas o vermelho e o branco;

Figura 5 – Captura de ecrã de 13 de Setembro de 2002 (correiomanha.pt)

- um sexto desenho pode ser encontrado nos materiais arquivados no âmbito dos endereços correiomanha.pt e correiodamanha.pt/index.asp, de 19 de Setembro de 2003 a 16 de Novembro de 2004, a versão apresentada na Figura 6 é um exemplo deste desenho, nela se observa um *layout* idêntico à versão anterior, no entanto a página é maior e debaixo das peças organizadas em quadrantes passa a haver mais peças em lista, parece também haver espaço publicitário mais especificado, mas a preservação da página não permite uma conclusão definitiva;

Figura 6 – Captura de ecrã de 16 de Junho de 2004 (correiodamanha.pt/index.asp)

- no contexto dos endereços correiomanha.pt, correiodamanha.pt/index.asp e correiodamanha.xl.pt, de 19 de Janeiro de 2005 a 12 de Março de 2008, encontra-se o sétimo desenho, como se pode observar na Figura 7, o *website* mantém uma certa estrutura geral, mas há muitas mudanças pequenas, destacando-se o aumento do número de peças na zona central do *website* através da diminuição dos quadrantes, tem também um *banner* da empresa de crédito Cofidis;

Figura 7 – Captura de ecrã de 1 de Abril de 2006 (correiomanha.pt)

- de 21 de Outubro de 2008 a 5 de Maio de 2010, em correiomanha.pt, correiodamanha.xl.pt e cmjornal.xl.pt, encontramos um oitavo desenho, que tem como caso exemplar a versão apresentada na Figura 8, toda a estrutura interactiva do *website* fica introduzida numa espécie de fundo cinzento, no topo destaca-se um grande *banner* para introduzir anúncios, a estrutura do *website* torna-se mais "arredondada" (barras com cantos arredondados, ícones arredondados), multiplica-se o número de colunas (há várias colunas verticais e vários cortes na horizontal, os quais isolam mais colunas verticais), multiplicam-se as possibilidades de clique (ícones com hiperligações, texto para outras páginas do *website*, texto com hiperligações), domina o tom e vários tons de cinzento, surgem galerias de imagens e de texto, destacam-se diferentes peças através da introdução de fundos coloridos;

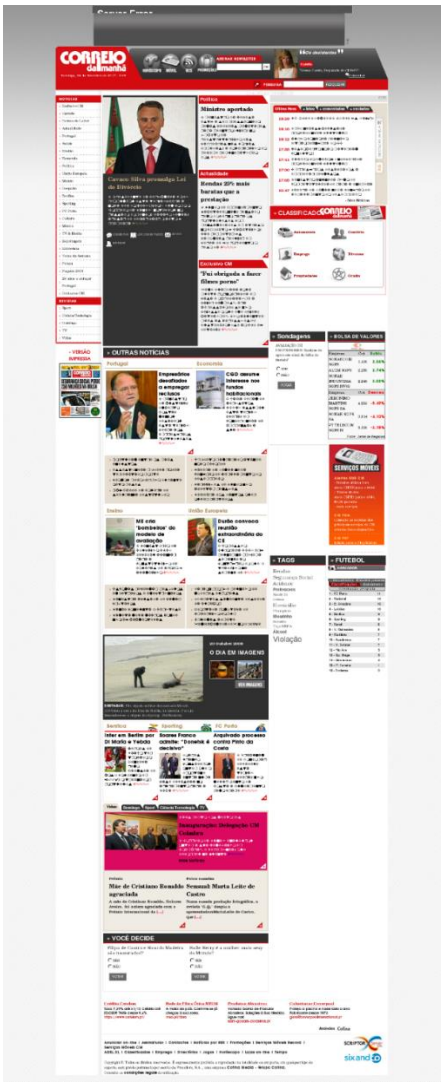

Figura 8 – Captura de ecrã de 21 de Outubro de 2008 (correiomanha.pt)

- servindo a versão da Figura 9 como exemplo, no âmbito do endereço cmjornal.xl.pt, de 7 de Maio de 2010 a 27 de Agosto de 2014, estamos na fase do nono desenho, a extensão vertical das páginas é muito maior, o campo visual passa a ser novamente ocupado por toda a página, há mais galerias (incluindo o

que parece ser vídeo), há muitas mudanças ao nível das pequenas coisas (mais colunas, mais cores, mais espaços em destaque);

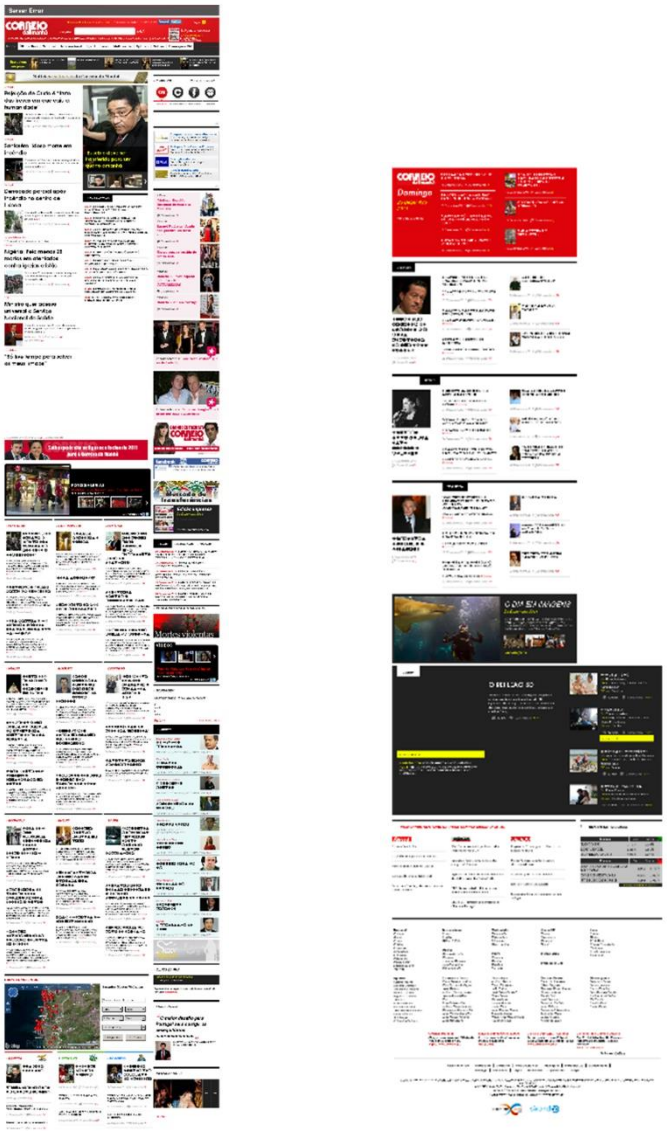

Figura 9 – Captura de ecrã de 25 de Dezembro de 2011 (cmjornal.xl.pt) (continuidade da página da esquerda para a direita)

- o décimo desenho encontra-se no material arquivado com as datas que vão de 1 de Setembro de 2014 a 21 de Maio de 2016 no contexto do endereço cmjornal.xl.pt, de que serve de exemplo a Figura 10, para além de novas mudanças em termos do desenhos das colunas e das letras, destaca-se o tamanho que as imagens passaram a ter, particularmente os retratos das pessoas que escrevem os textos de opinião.

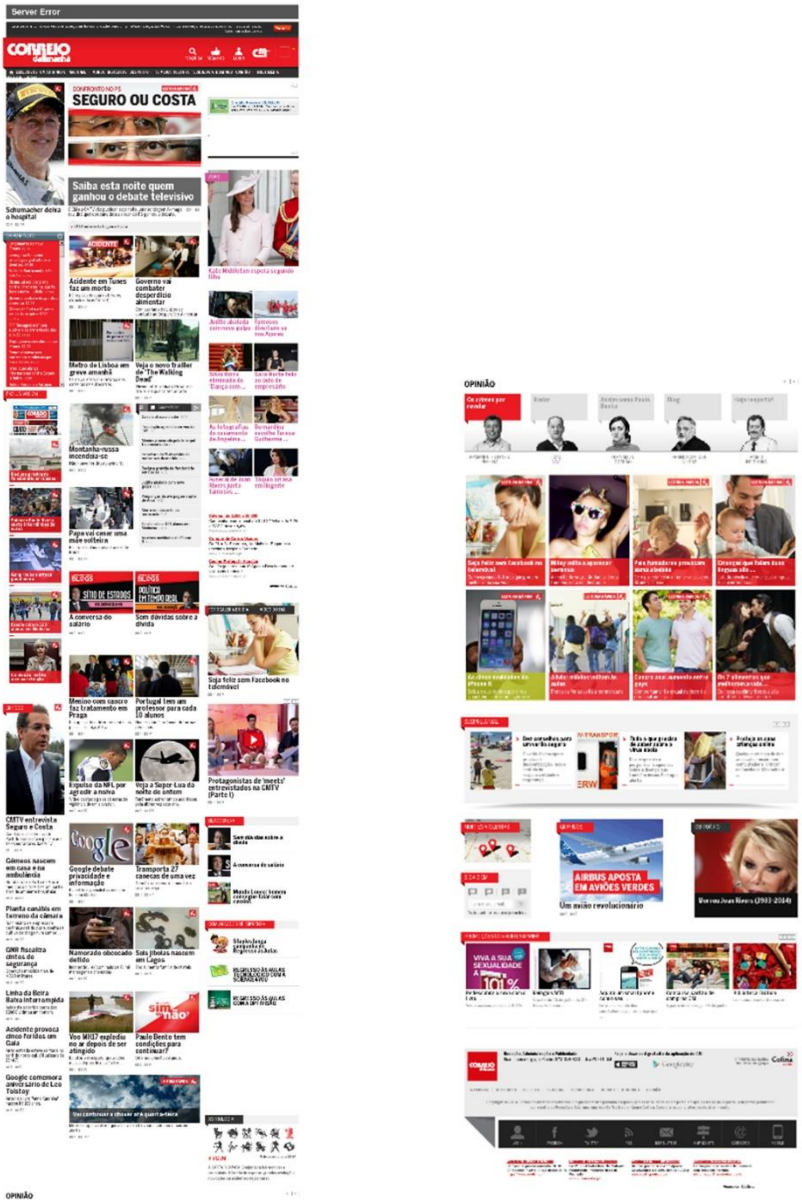

Figura 10 – Captura de ecrã de 9 de Setembro de 2014 (cmjornal.xl.pt) (continuidade da página da esquerda para a direita)

A comparação entre as páginas web do(s) *website*(s) corresponde(s) ao DN encontradas na rota de investigação elaborada a partir do material desse jornal disponível no Arquivo.pt permite isolar cinco desenhos diferentes de *layouts* desse jornal:

- no âmbito do endereço dn.pt, de 13 de Outubro de 1996 a 13 de Junho de 1997, com o exemplo da versão apresentada na Figura 11, encontra-se o primeiro desenho, a navegação do *website* está pensada na horizontal de tal modo que o título do jornal surge na vertical sobre um fundo branco, todo o espaço para as peças está num grande fundo azul;

Figura 11 – Captura de ecrã de 12 de Janeiro de 1997 (dn.pt)

- em dn.pt, de 10 de Dezembro de 1997 a 30 de Setembro de 2001, descobre-se o segundo desenho, de que a Figura 12 é exemplo, o fundo torna-se completamente preto, distingue-se uma barra de navegação verde e uma grande imagem para apresentar o jornal, há um espaço publicitário onde se vêem os logótipos da empresa de telecomunicações Portugal Telecomunicações e da empresa de audiovisual Lusomundo;

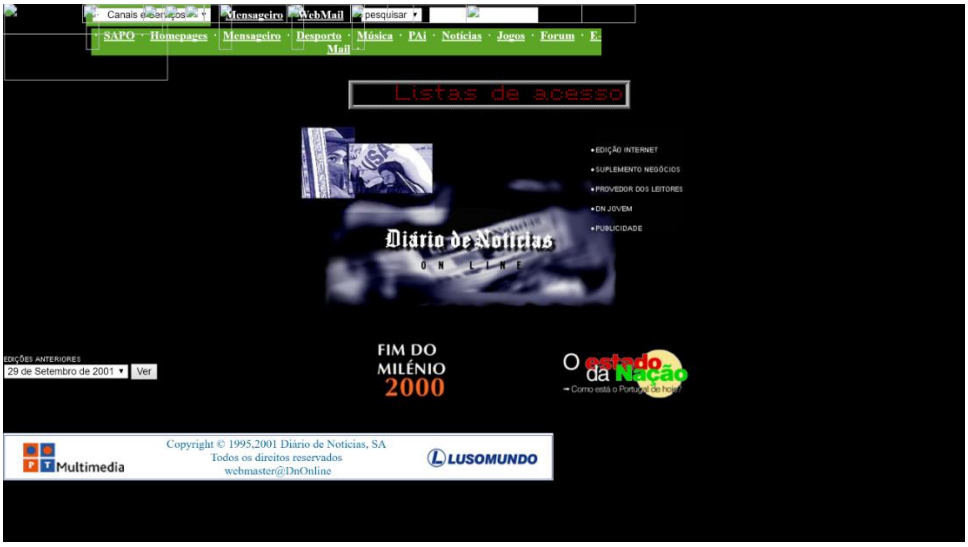

Figura 12 – Captura de ecrã de 30 de Setembro de 2001 (dn.pt)

- de 17 de Janeiro de 2002 a 21 de Outubro de 2004, no âmbito do endereço em dn.sapo.pt/homepage/homepage.asp, é possível encontrar um terceiro desenho, servindo de exemplo a versão da Figura 13, tudo está completamente diferente, há um espaço para navegação horizontal no topo da página, onde se destaca o azul como cor a acompanhar o logótipo do jornal e uma barra de navegação verde à direita do logótipo da empresa de telecomunicações Sapo, no final da página permanecem os logótipos da Portugal Telecom e da Lusomundo, ao longo da página na vertical há três grandes colunas com muitas cores diferentes, sendo que a do centro é mais larga e ocupada pelas ligações para as peças, a da esquerda destaca notícias de última hora e a da direita apresenta mais opções de navegação;

Figura 13 – Captura de ecrã de 18 de Novembro de 2002 (dn.sapo.pt/homepage/homepage.asp)

- em dn.sapo.pt/homepage/homepage.asp e dn.sapo.pt/index.html, de 23 de Outubro de 2004 a 29 de Outubro de 2008, encontra-se o quarto desenho, de que a versão apresentada na Figura 14 é exemplar, a parte superior da página

tem o logótipo da Sapo centrado, as peças ocupam a parte central da página, sendo que à direita se apresenta um grande *banner* vertical para classificados, as cores que antes enchiam a página desaparecem e dão lugar a uma versão quase totalmente branca:

Figura 14 – Captura de ecrã de 31 de Dezembro de 2005 (dn.sapo.pt/homepage/homepage.asp)

- no contexto dos endereços dn.sapo.pt/inicio/ e dn.pt/inicio, de 30 de Setembro de 2009 a 31 de Janeiro de 2011, pode descobrir-se um quinto desenho, exemplificado na Figura 15, a página tem uma extensão vertical muito maior do que anteriormente, mantém-se o azul como cor principal do jornal, o logótipo da empresa de *media* do jornal, na época ControlInveste, surge em cima e no fundo da página, demarcam-se claramente três colunas verticais onde se vão

diferenciando e destacando peças e há várias galerias para navegar em trabalhos destacados e galerias de imagens e de vídeo;

Figura 15 – Captura de ecrã de 12 de Janeiro de 2011 (dn.sapo.pt/inicio/) (continuidade da página da esquerda para a direita)

No caso do DN, podem fazer-se algumas observações adicionais. Primeiro, as comparações que permitiram isolar um *layout* no período entre 10 de Dezembro de 1997 e 30 de Setembro de 2001 mostram diferenças significativas em termos de ausência de elementos visuais, embora as semelhanças apontem para uma estrutura geral semelhante. Segundo, a partir de 2011 não se consideraram as versões arquivadas

porque apresentam problemas de preservação que não permitem fazer distinções visuais minimamente rigorosas, como se pode observar na Figura 16.

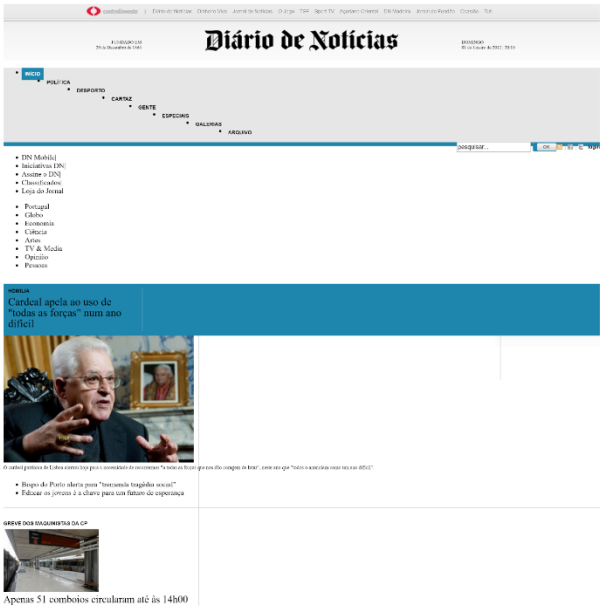

Figura 16 – Captura de ecrã de pormenor de 1 de Janeiro de 2012 (dn.pt/inicio/)

A comparação entre as páginas web do(s) *website*(s) corresponde(s) ao *Expresso* encontradas na rota de investigação elaborada a partir do material desse jornal disponível no Arquivo.pt permite isolar doze desenhos diferentes de *layouts* desse jornal:

- no âmbito dos endereços expresso.pt, de 10 de Janeiro de 1998 a 4 de Março de 2000, pode observar-se o primeiro desenho, de que a versão apresentada na Figura 17 é exemplo, encontramos desde logo páginas muito longas, um menu de navegação logo no topo do *website* e uma longa zona de peças colocadas sobre um fundo branco, com alguns destaques, há algumas imagens e há no fundo da página um *banner* de publicidade ocupado por um anúncio da empresa de informática Compaq;

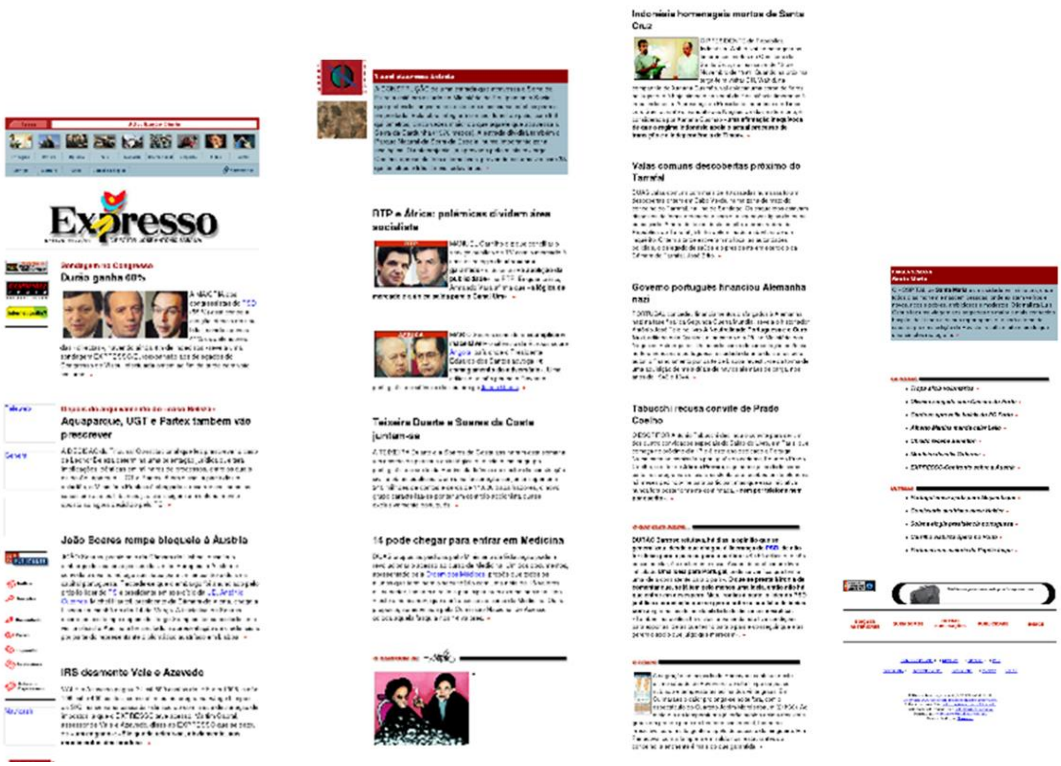

Figura 17 – Captura de ecrã de 3 de Março de 2000 (expresso.pt) (continuidade da página da esquerda para a direita)

- em expresso.pt e online.expresso.pt, de 18 de Abril de 2001 a 28 de Junho de 2001, encontra-se um segundo desenho, como se pode observar no caso exemplificativo da versão da Figura 18, o menu da parte superior é contraído

num barra e toda a navegação do *website* por materiais jornalísticos fica centralizada numa coluna vertical;

Figura 18 – Captura de ecrã de 25 de Junho de 2001 (expresso.pt) (continuidade da página da esquerda para a direita)

- no âmbito dos endereços expresso.pt, online.expresso.pt/ e online.expresso.clix.pt, de 22 de Agosto de 2001 a 1 de Julho de 2006, pode observar-se o terceiro desenho, de que serve de exemplo a versão apresentada na Figura 19, a página passa a ter três colunas verticais claramente diferenciáveis, a primeira destinada a assinalar a edição diária e a edição

semanal, a segunda concentrando as peças, a terceira com destaques, aparece o azul como cor ligada ao logótipo do jornal;

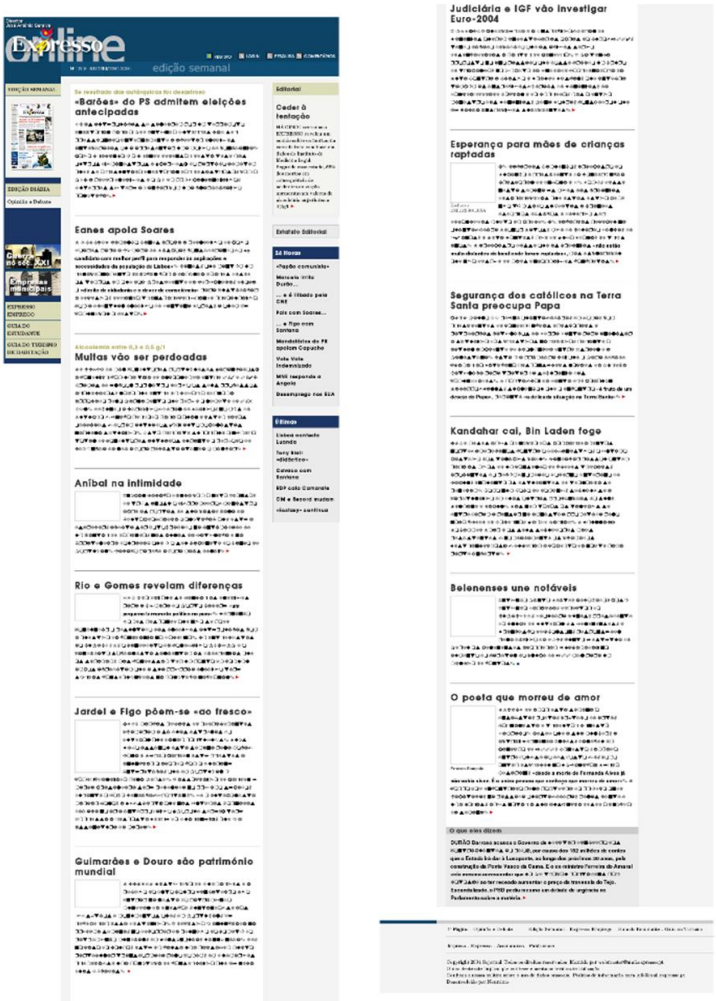

Figura 19 – Captura de ecrã de 9 de Dezembro de 2001 (expresso.pt) (continuidade da página da esquerda para a direita)

- de 4 de Julho de 2006 a 1 de Setembro de 2006, nos endereços expresso.clix.pt e online.expresso.clix.pt, está um quarto desenho, servindo de exemplo a captura de ecrã da Figura 20, o azul desaparece, o *website* fica praticamente branco e cinzento, com um botão de pesquisa cor de rosa na barra superior combinando com o logótipo da empresa de telecomunicações Clix, a zona de navegação do *website* continua dividida em três grandes áreas, mas há mais colunas verticais e vários cortes na horizontal isolando mais colunas verticais, à esquerda há uma coluna que serve de menu de navegação pelos conteúdos do *website*, no centro as principais peças, à direita estão alguns destaques, uma caixa com *scroll* para navegar por conteúdos do serviço de notícias do canal

televisivo SIC (também da Impresa, empresa proprietária do *Expresso*), há aí também um espaço para colocação de um anúncio;

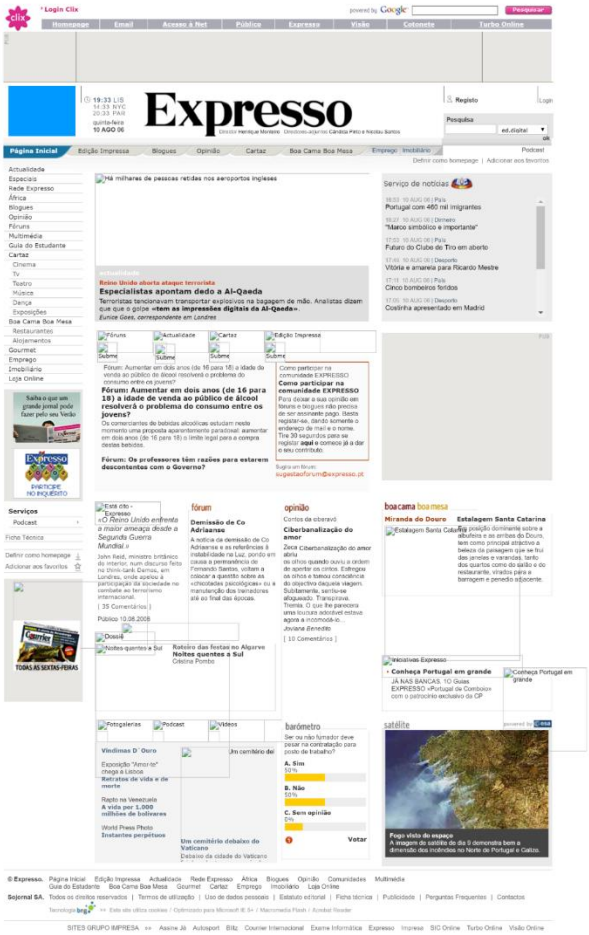

Figura 20 – Captura de ecrã de 10 de Agosto de 2006 (expresso.clix.pt)

- no contexto do endereço expresso.clix.pt, do período que vai de 6 de Outubro de 2006 a 19 de Agosto de 2007, usando como exemplo a versão apresentada na Figura 21, aparece o quinto desenho, a página parece ser mais extensa, há mais títulos clicáveis, mais destaques, poderia falar-se de várias colunas distinguíveis ao longo do *website*, o azul volta como cor ligada ao logótipo;

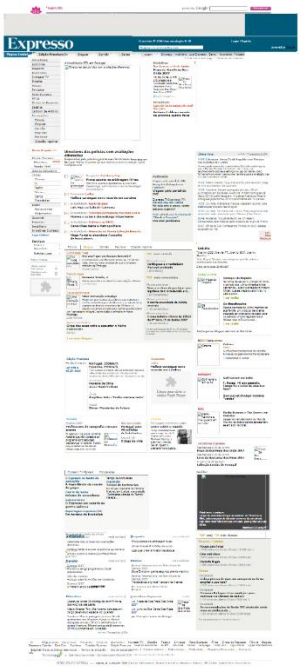

Figura 21 – Captura de ecrã de 27 de Junho de 2007 (expresso.clix.pt)

- em aeiou.expresso.pt, de 2 de Janeiro de 2008 a 20 de Maio de 2009, pode descobrir-se um sexto desenho, exemplificado na Figura 22, há mais imagens, o logótipo da Clix desaparece e é colocada uma pequena barra superior onde está o logótipo da empresa de telecomunicações AEIOU, há um *banner* longo na parte superior da página, todas as peças estão com um espaço maior de localização relativa entre si;

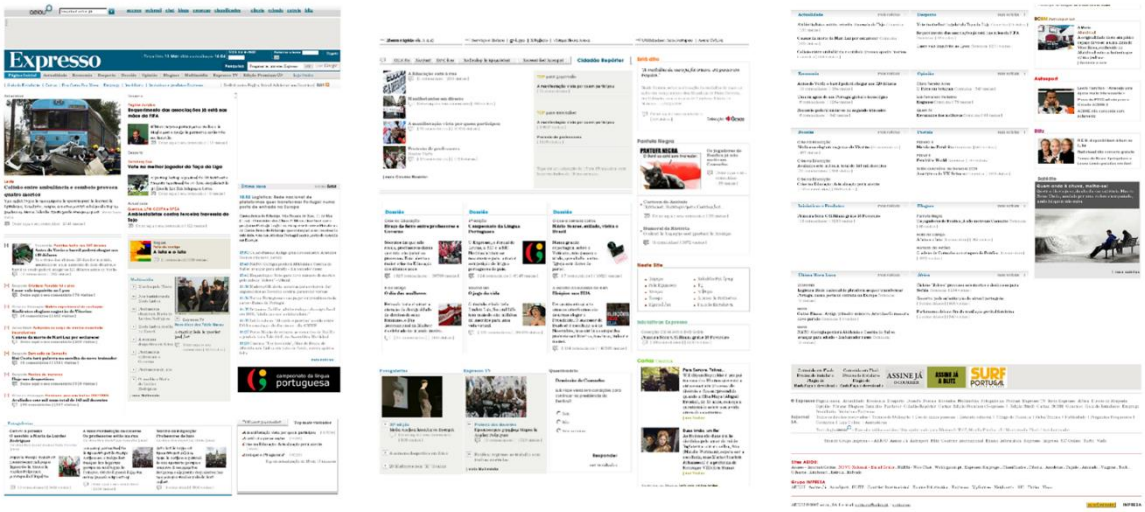

Figura 22 – Captura de ecrã de 11 de Março de 2008 (aeiou.expresso.pt) (continuidade da página da esquerda para a direita)

- do período que vai de 24 de Setembro de 2009 a 25 de Fevereiro de 2011, pode encontrar-se o sétimo desenho no âmbito do endereço aeiou.expresso.pt, utilizando-se como exemplo a versão da Figura 23 pode observar-se que há um ajustamento do tamanho das imagens, surgem galerias com imagens e vídeos, mantêm-se três grandes colunas verticais, o *layout* do *website* muda em muitos pormenores mas mantém-se o azul e o cinzento;

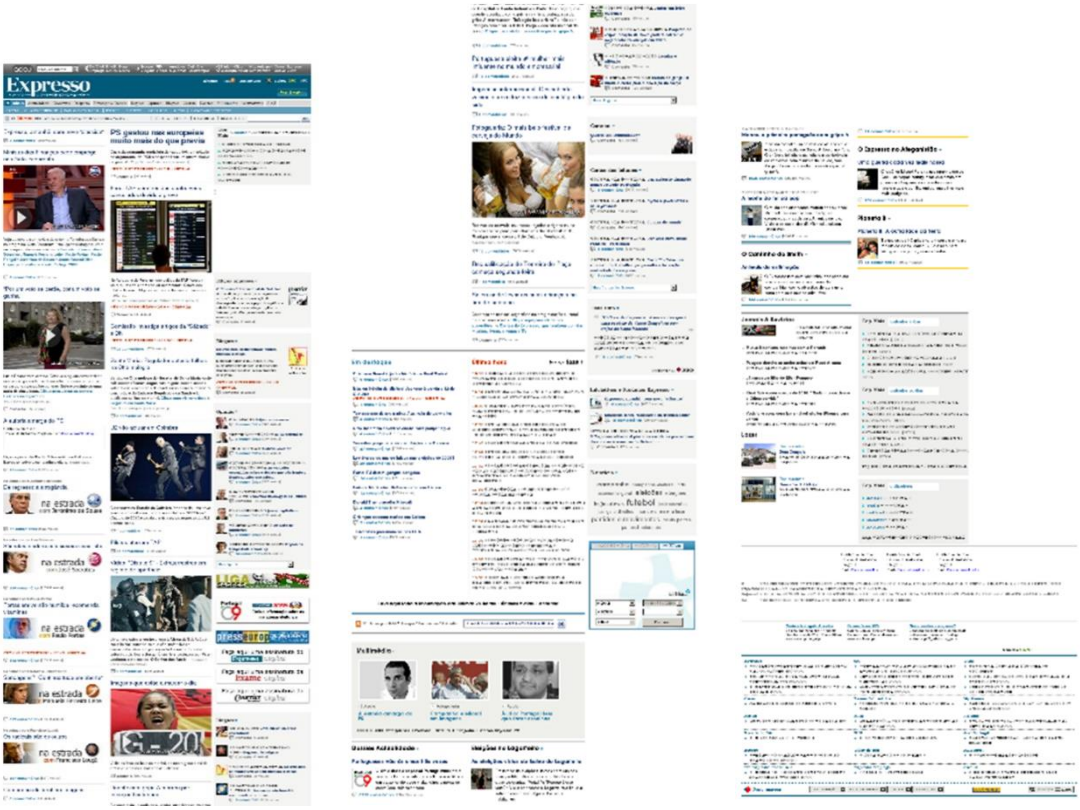

Figura 23 – Captura de ecrã de 10 de Fevereiro de 2011(aeiou.expresso.pt) (continuidade da página da esquerda para a direita)

- no âmbito do endereço aeiou.expresso.pt, do período que vai de 26 de Fevereiro de 2011 a 11 de Março de 2012, surge o oitavo desenho da página, tomando como caso exemplar a versão arquivada apresentada na Figura 24, pode verificar-se que surgem grandes destacas nas áreas de peças, com imagens

maiores, há um menu com *scroll* que concentra destaques à direita, o *layout* volta a mudar em pormenores pequenos;

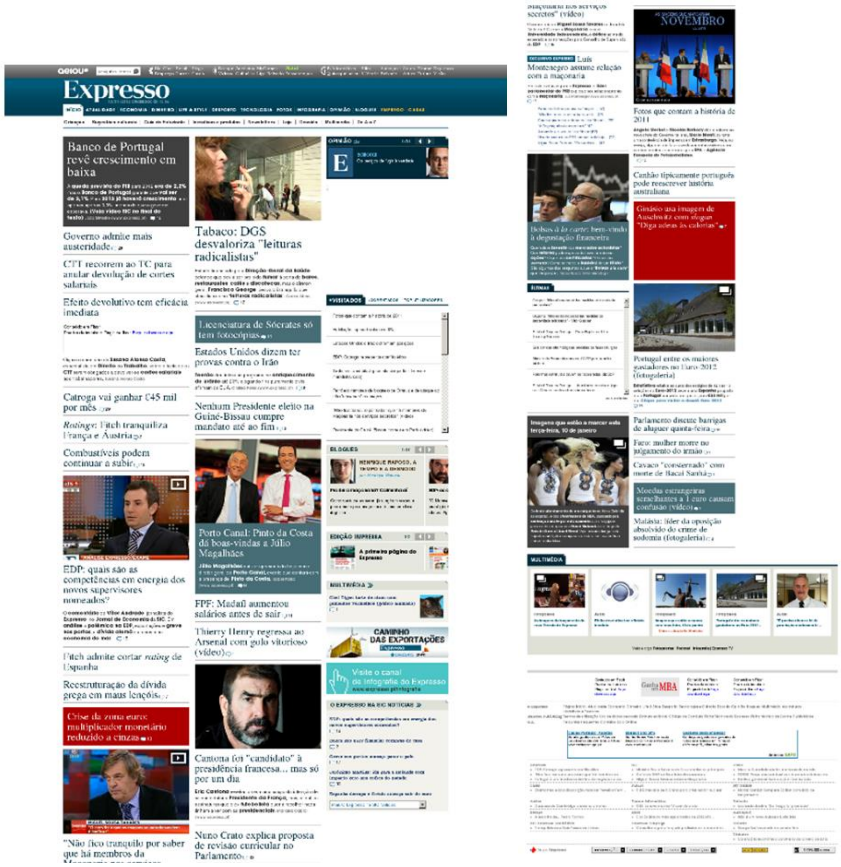

Figura 24 – Captura de ecrã de 10 de Janeiro de 2012 (aeiou.expresso.pt) (continuidade da página da esquerda para a direita)

- do período que vai de 12 de Março de 2012 a 31 de Dezembro de 2012, no contexto do endereço expresso.sapo.pt, pode ver-se, como na Figura 25 se evidencia, o nono desenho, a página apresenta um *layout* idêntico ao caso anterior, mas, para além de mudanças de pormenor, aparece uma barra com o logótipo da Sapo onde antes estava o da AEIOU;

Figura 25 – Captura de ecrã de 8 de Setembro de 2012 (expresso.sapo.pt) (continuidade da página da esquerda para a direita)

- no âmbito do endereço expresso.sapo.pt, de 10 de Janeiro de 2013 a 10 de Novembro de 2013, tomando como exemplo a imagem da Figura 26, observa-se que a zona de navegação da esquerda do *website* fica destinado à apresentação das peças, enquanto a zona central fica com grandes destaques e fotografias, à direita estão menus com navegação por *scroll*, surgem mais *banners* para publicidade à direita;

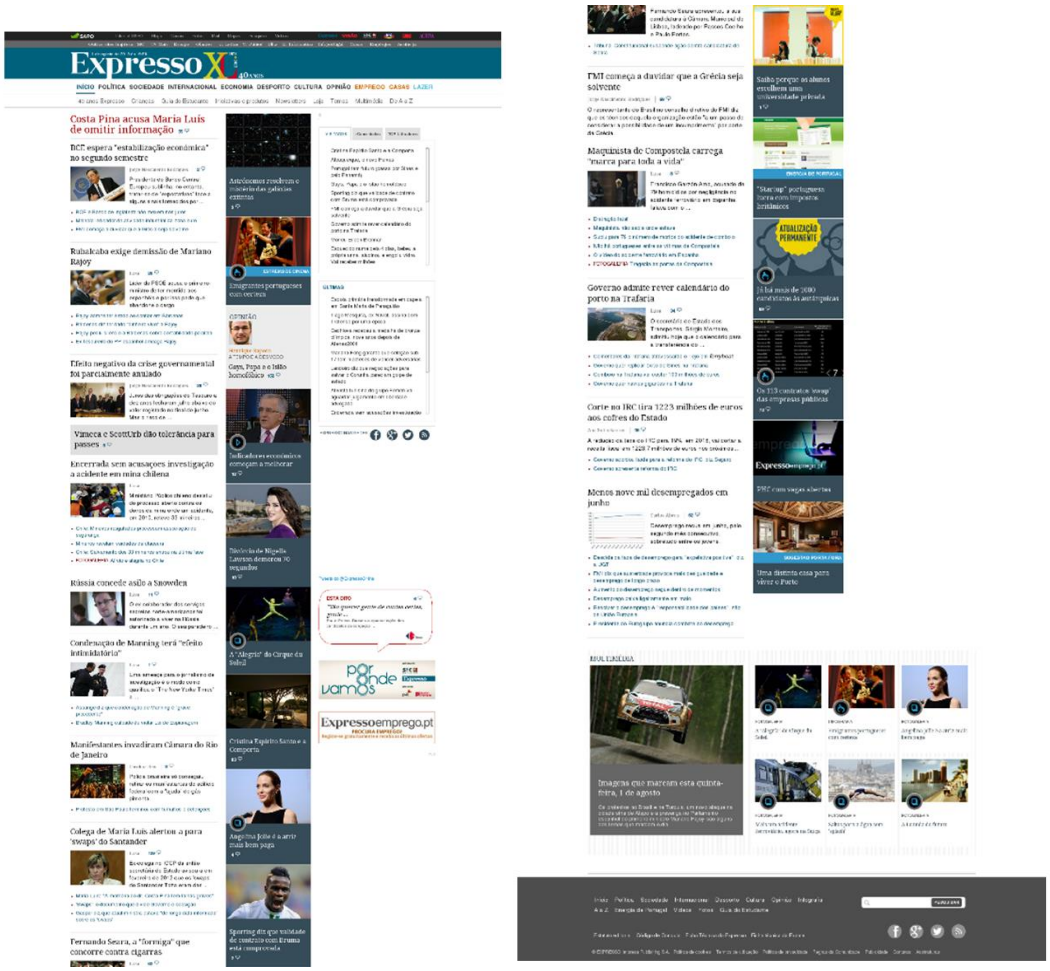

Figura 26 – Captura de ecrã de 1 de Agosto de 2013 (expresso.sapo.pt) (continuidade da página da esquerda para a direita)

- em expresso.sapo.pt, no intervalo entre 27 de Abril de 2014 e 6 de Maio de 2014, de que a Figura 27 serve de exemplo, o *website* apresenta a mesma linha, mas usam-se imagens maiores, há novamente mudanças em termos microscópicos (formas, feitios, tipos);

Figura 27 – Captura de ecrã de 27 de Abril de 2014 (expresso.sapo.pt) (continuidade da página da esquerda para a direita)

- no âmbito do endereço expresso.sapo.pt, do período que vai de 7 de Maio de 2014 a 3 de Maio de 2015, tomando como caso exemplar a versão apresentada na Figura 28, verifica-se que a página tem uma extensão vertical muito maior, surge um espaço na parte superior da página onde são imagens das pessoas que escrevem as colunas de opinião, na coluna da direita da página passa a existir destaque de peças do diário do jornal apresentando grandes fotografias, na parte de baixo do *website* há um série de tópicos separados por chavetas colocadas na horizontal, onde estão por exemplo as peças assinaladas como "multimédia";

Figura 28 – Captura de ecrã de 13 de Fevereiro de 2015 (expresso.sapo.pt) (continuidade da página da esquerda para a direita)

Na pesquisa sobre páginas do *Expresso*, foram encontradas algumas páginas peculiares que merecem atenção para reflexão futura, como aquelas que constam nas Figuras 29, 30 e 31.

[이동통신] 엘지 'ALL IP'망 4세대 직행론 '갸우뚱'

"벌써 4세대? 글쎄요…."

`올 아이피'(ALL IP)망을 조기에 깔아 2.5세대(CDMA2000-1X)에서 4세대 서비스로 직행하겠다는 박운서 데이콤 부회장(엘지 IMT-2000사업추진단장)의 발언에 대해 학계와 업계의 전문가들은 "뭔가 오해가 있는 것 같다"는 촌평을 내놓고 있다.

한국전자통신연구원(ETRI)의 한 박사는 "통신 표준화기구인 국제전기통신연합(ITU-R)이 4세대의 표준을 정하기 위한 회의를 겨우 4차례 열었을 뿐"이라며 "기술개발은 2007년이나 돼야 가능하고, 상용화 가능시기는 2010년께로 보고 있다"고 했다.

연구원의 또 다른 박사도 "올 아이피가 4세대에 적합한 시스템이기는 하지만 그것 자체가 4세대를 의미하는 것은 아니다"며 "3세대처럼 4세대도 정부가 새로 주파수 대역을 할당해야 서비스가 가능하다"고 했다.

올 아이피란 이동통신의 모든 네트워크를 인터넷 프로토콜 기반으로 바꾼다는 뜻이다. 올 아이피가 이뤄지면 이동통신 가입자도 유선 인터넷과 같은 수준의 인터넷 서비스를 받을 수 있고 교환효율이 9배 이상 높아진다.

올 아이피 1단계로 평가되는 IMT-2000의 경우 모바일스위칭센터(MSC) 등 핵심네트워크만을 아이피화하고, 교환기나 기지국 사이 등 나머지는 기존 회선망을 이용한다. 에스케이텔레콤 관계자는 "완전한 의미의 올 아이피화를 의미하는 3단계는 2007년에나 가능할 것"이라고 했다.

이재성 기자firib@hani.co.kr

Figura 29 – Captura de ecrã de 4 de Março de 2001 (expresso.pt)

[Innehåll | Nytt | Svara]

A!
Namn: Mathias
e-mail: meezziaz@hehe.com
Datum: 981029 17:16

Som svar på: Sv: Mez! postad av Anita, 981027 14:34:

Jag väntar med spänning på mailet. Har du filen mirc541t.exe kvar? annars finns den på www.mirc.co.uk
Hur kommer det sig att du inte har mirc kvar? har det strulat tidigare?

PoK Mathias

LineBreak

Efterföljande inlägg:

- Sv: Ah..... **Anita**, 981030 23:44
  - Sv: Ah..... **Mathias**, 981101 15:34
    - Sv: Ah..... **Anita**, 981103 14:13

LineBreak

Figura 30 – Captura de ecrã de 5 de Março de 2001 (expresso.pt)

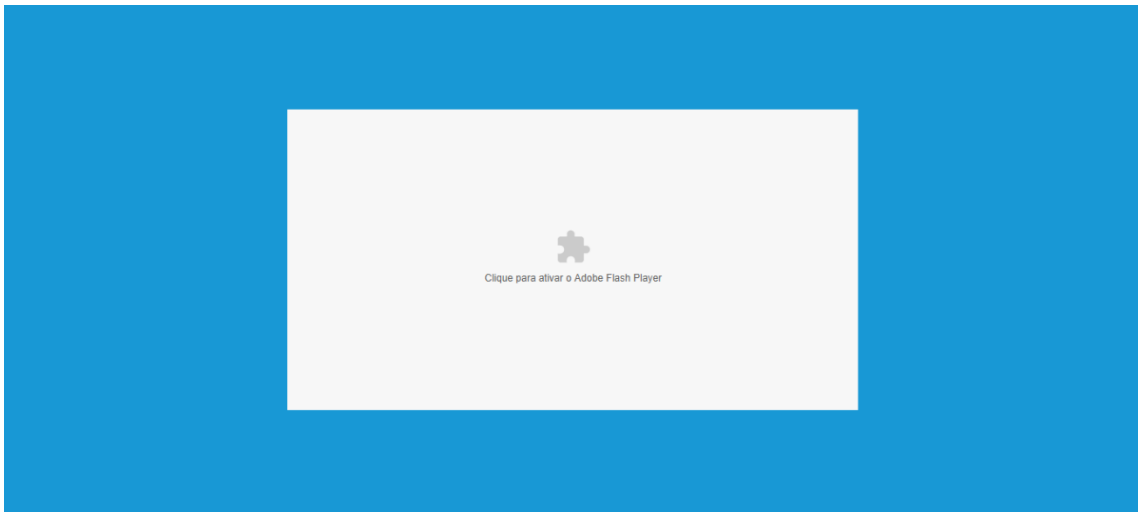

Figura 31 – Captura de ecrã de 29 de Setembro de 2003 (expresso.pt)

A comparação entre as páginas web do(s) *website*(s) corresponde(s) ao *Público* encontradas na rota de investigação elaborada a partir do material desse jornal disponível no Arquivo.pt permite isolar nove desenhos diferentes de *layouts* desse jornal:

- no âmbito do endereço publico.pt, à data de 13 de Outubro de 1996, encontra-se o primeiro desenho, como se observa na versão desse dia apresentada na Figura 32, vê-se uma página organizada na horizonta, com vários espaços sobre o próprio jornal, no topo está o nome do jornal dentro de um fundo cinzento;

Figura 32 – Captura de ecrã de 13 de Outubro de 1996 (publico.pt)

- em publico.pt, 12 de Dezembro de 1998, pode observar-se um segundo desenho, como mostra a Figura 33, a navegação é estabelecida na vertical, diferenciam-se duas grandes colunas com texto, com alguns destaques a amarelo, na zona superior, na zona da esquerda e em várias peças da zona central faltam imagens (que não foram preservadas) que não permitem a caracterização dessas zonas;

Figura 33 – Captura de ecrã de 12 de Dezembro de 1998 (publico.pt)

- no intervalo entre 25 de Janeiro de 1999 e 18 de Outubro de 2000, no contexto do endereço publico.pt, está o terceiro desenho, como se evidencia na versão apresentada na Figura 34, uma vez mais faltam muitos conteúdos não preservados, no entanto percebe-se que a cor preta e a cor vermelha são as dominante na página, aliás o logótipo do jornal é então a palavra “público” em vermelho separada da palavra “pt” por uma imagem do planeta Terra a fazer de ponto, percebe-se também que a navegação fica organizada na vertical numa pequena zona da esquerda mas muito horizontal de todo o lado direito;

Figura 34 – Captura de ecrã de 21 de Abril de 1999 (publico.pt)

- em publico.pt, de 18 de Abril de 2001 a 21 de Fevereiro de 2005, encontra-se o quarto desenho, como se vê na Figura 35, a página começa a aumentar a sua extensão na vertical, o logótipo do jornal é refeito, desaparece a cor preta e vários tons de azuis preenchem a página, surge uma barra para navegação na horizontal em cima, a zona principal do *website* divide-se em três colunas, uma coluna de navegação com menus à esquerda, uma zona central com as peças e com alguns *banners*, à direita uma coluna com destaque de notícias e com uma zona de formulário;

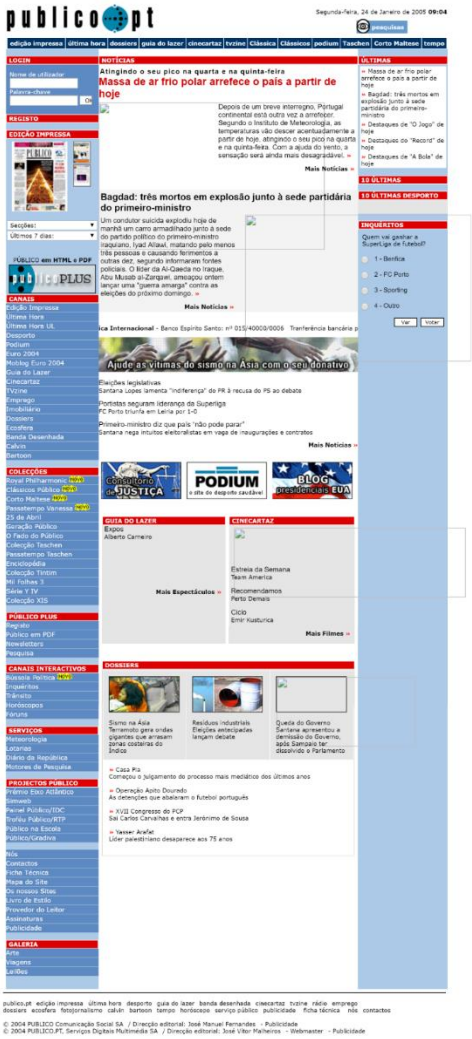

Figura 35 – Captura de ecrã de 24 de Janeira de 2005 (publico.pt)

- no âmbito do endereço publico.clix.pt, do período que vai de 22 de Fevereiro de 2005 a 10 de Fevereiro de 2007, como é exemplificado na versão da Figura 36, na parte superior da página surge um botão de pesquisa cor de rosa do lado direito e o logótipo também cor de rosa da Clix no lado esquerdo, embora se continuam a distinguir três grandes colunas verticais na página, no corpo da página multiplica-se o número de pequenas colunas (há várias colunas verticais e vários cortes na horizontal, os quais isolam mais colunas verticais) e respectivos conteúdos, na zona horizontal destinada aos trabalhos de opinião apresentam-se os textos pelos títulos mas também através das fotografias das pessoas que os escrevem;

Figura 36 – Captura de ecrã de 22 de Agosto de 2006 (publico.clix.pt) (continuidade da página da esquerda para a direita)

- no endereço ww2.publico.clix.pt, de 17 de Fevereiro de 2007 a 21 de Novembro de 2007, surge um quinto desenho, como mostra o caso exemplar da Figura 37, surge a imagem actual do logótipo do *Público*, desaparece o logótipo da Clix e as coisas cor de rosa, mas continua a existir uma caixa com indicações da Clix, na parte superior da página há um espaço para publicidade (cuja dimensão não se pode perceber devido a problemas de preservação), há várias mudanças em termos, uma vez mais, das coisas mais pequenas;

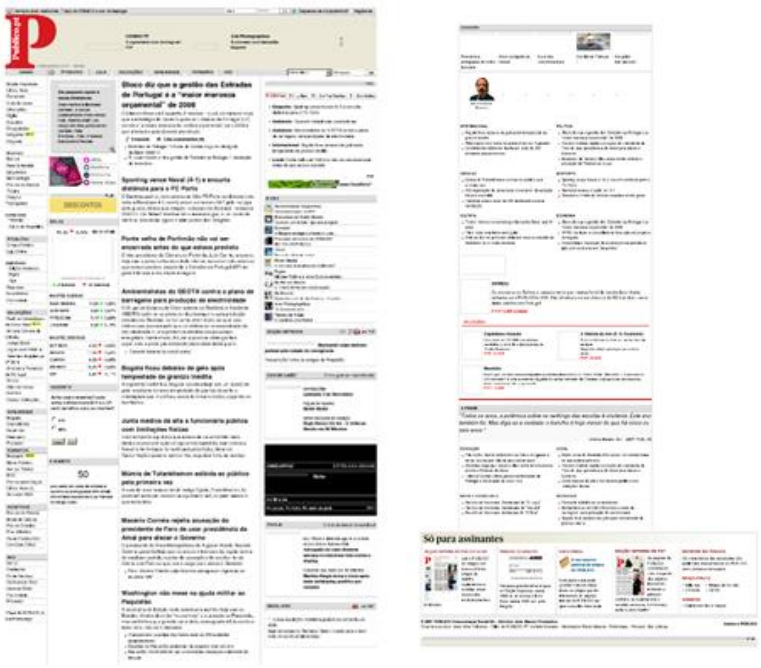

Figura 37 – Captura de ecrã de 5 de Novembro de 2007 (ww2.publico.clix.pt) (continuidade da página da esquerda para a direita)

- em publico.clix.pt, de 21 de Novembro de 2007 a 22 de Outubro de 2008, como mostra a versão da Figura 38, está-se perante um sexto desenho, a coluna de menus à esquerda desaparece e dá lugar a uma coluna de peças idêntica à coluna central e há uma série de mudanças mais pequenas, como a introdução de azul nos títulos das peças da coluna central;

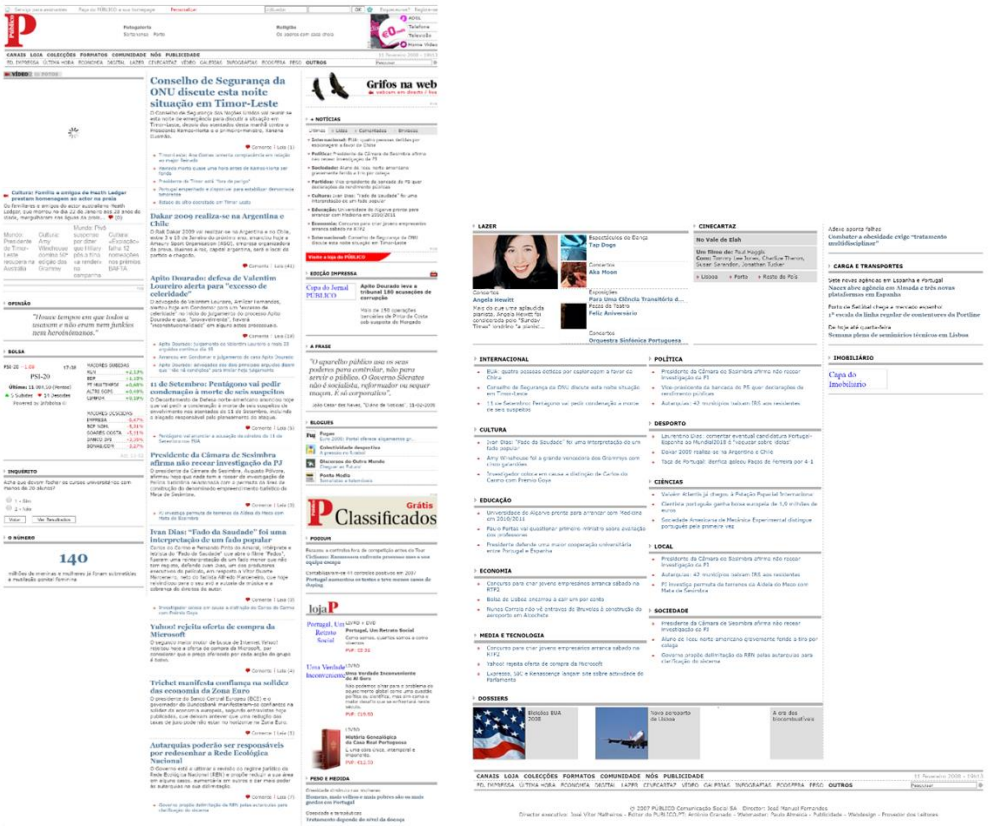

Figura 38 – Captura de ecrã de 11 de Fevereiro de 2008 (publico.clix.pt) (continuidade da página da esquerda para a direita)

- no âmbito dos endereços publico.clix.pt e publico.pt, do período que vai de 20 de Maio de 2009 a 16 de Novembro de 2012, encontra-se o sétimo desenho, de que a versão apresentada na Figura 39 serve de exemplo, a extensão vertical da página é muito maior, mantém-se a mesma estrutura de navegação, mas há muitas mudanças em termos microscópicos, designadamente com a introdução de pequenos separadores vermelhos ao longo das colunas e da generalização do azul na titulação;

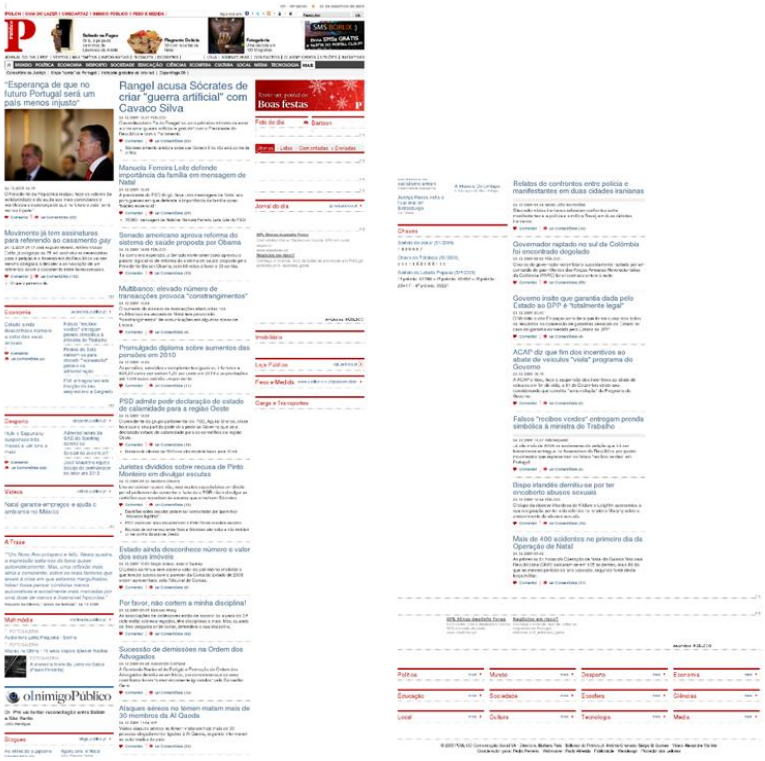

Figura 39 – Captura de ecrã de 25 de Dezembro de 2009 (publico.clix.pt) (continuidade da página da esquerda para a direita)

- no âmbito do endereço publico.pt, no intervalo entre 24 de Novembro de 2012 e 15 de Outubro de 2015, como se pode observar na versão apresentada na Figura 40, encontra-se um oitavo desenho, verifica-se que na parte superior da página estabelece-se uma barra vermelha com várias opções de navegação, o azul desaparece dos títulos dando lugar ao preto e surgem muitas mais imagens no corpo da página, há uma grande quantidade de pequenas mudanças estéticas;

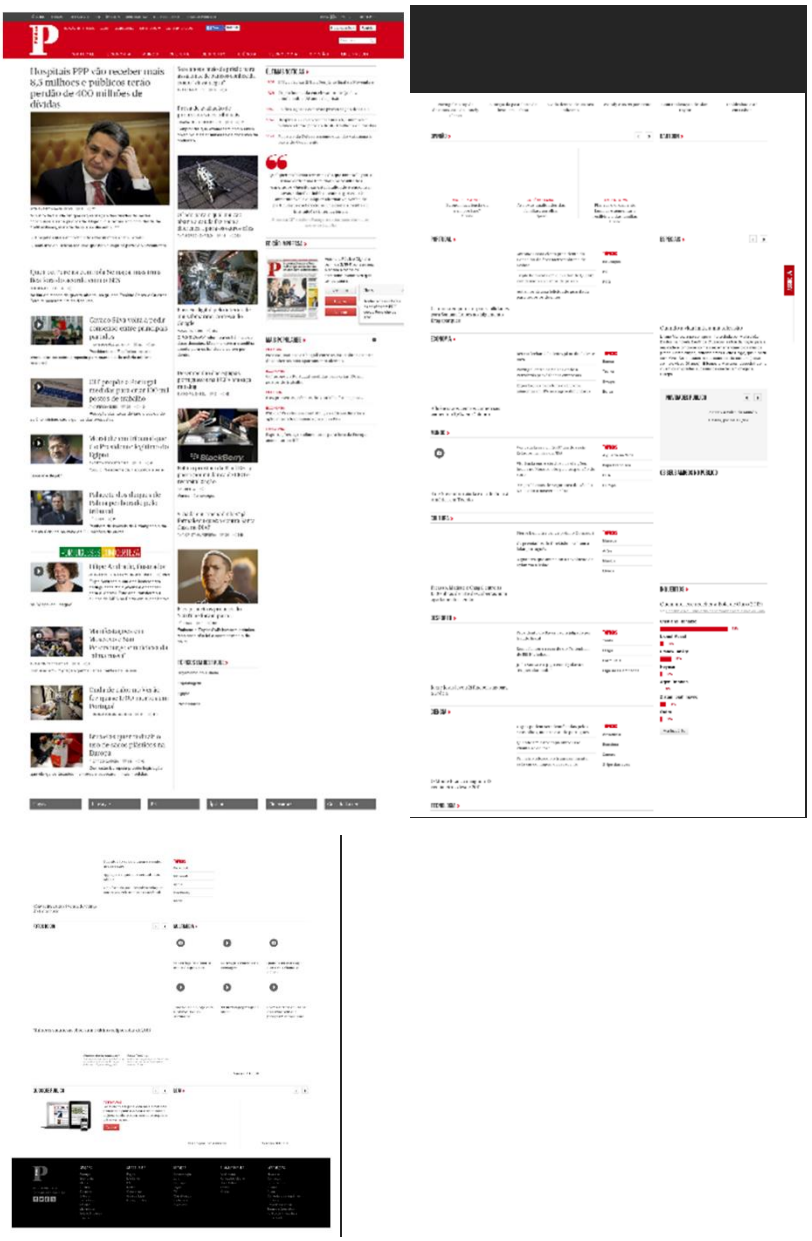

Figura 40 – Captura de ecrã de 4 de Novembro de 2013 (publico.pt)

A descrição, análise e comparação dos diversos materiais das páginas estudadas permitiu pensar períodos de continuidade de *layouts* dos *websites* e as respectivas rupturas que deram lugar a outro *layout*. A comparação entre todos estes *layouts* permite traçar algumas tendências gerais de desenvolvimentos dos *websites*:

- as páginas web são cada vez mais extensas, sobretudo na vertical;

- as páginas web vão tendo progressivamente uma orientação verticalizada com grandes barras de *scroll*, embora sejam inseridos pequenos espaços com galerias ou pequenas barras de *scroll*;
- as páginas web vão apresentando um acréscimo de estabilidade estética, consolidando a identidade visual do jornal;
- as mudanças são cada vez mais difíceis de assinalar na medida em que deixa são gradualmente menos "macro" e mais "micro", o que implica uma certa adaptação da perspectiva de análise;
- vão surgindo cada vez mais imagens e vídeos, muitas vezes enquadradas em galerias, aumentam também o número de hiperligações, botões, menus e barras de *scroll*.

Estas tendências são ainda permeadas pelo surgimento da imagem de várias empresas, sobretudo de telecomunicações, nas páginas dos *websites*. As mais comuns são a Clix e a Sapo.

Durante esta etapa da investigação, verificou-se que a conversão dos *snapshots* em *screenshots* pode provocar anomalias nos últimos, como problemas de resolução ou de correspondência com o que visualiza nos *snapshots*, ou pode haver erros informáticos que impedem a realização do próprio processo de conversão.

7 Conclusões

Neste relatório apresentaram-se os resultados e algumas reflexões decorrentes de um projecto que tinha como objectivo traçar uma linha de continuidades e descontinuidades sobre os desenvolvimentos do jornalismo em Portugal no contexto da web a partir de materiais disponíveis no Arquivo.pt. O foco da investigação dizia respeito às questões da preservação dos *websites* dos jornais, ao estudo da mudança efectiva desses *websites* ao longo do tempo e às implicações várias dessas mudanças em termos sociais e culturais. Dada a ancoragem temporal dos dados do Arquivo.pt, o período de estudo foi delimitado entre 1996 e 2016. E, devido à a quantidade e diversidade de jornais portugueses com *websites*, optou-se por estudar apenas as edições *online* do *Correio da Manhã*, do *Diário de Notícias*, do *Expresso* e do *Público*. O contexto do presente projecto de investigação e de redacção deste relatório contribuíram, contudo, para pensar o enquadramento histórico e teórico dos problemas em causa, tanto sobre a situação do jornalismo e da memória quanto acerca das metodologias que permitem fornecer contributos de investigação.

No âmbito da incursão dos temas em causa, foram, portanto, apresentadas duas reflexões distintas. Em primeiro lugar, foi dado um enquadramento teórico e histórico acerca da relação do jornalismo com a memória. Os jornalistas começaram, a uma escala global, passaram por uma forte alteração em termos do seu modo de produção, pois os seus processos produtivos ficaram focados *na* e *para* a internet. Num contexto de crescente complexidade do ecossistema de *media*, os trabalhos dos jornalistas foram digitalizados ou passaram a ser directamente pensados e produzidos para a web. No entanto, os jornalistas não vêem como parte da sua missão a preservação desses trabalhos. O próprio contexto de preservação da web enfrenta dificuldades *sui generis*, que contribuem para a perda de informação no médio e longo prazo. Assim, a função comunicacional-informativa de os jornalistas fornecerem às pessoas as informações necessárias para que se possam orientar e coordenar em sociedade não está estabelecida ao longo do tempo, porque dessas informações perde-se hoje o vestígio histórico. O que hoje é notícia amanhã poderá estar esquecido.

Em segundo lugar, considerou-se que, no quadro de um conjunto crescente de iniciativas de preservação de material digital, os arquivos da web podem funcionar como

fonte de dados para a investigação em ciências sociais, particularmente para investigação acerca do jornalismo. Para além de darem acesso a materiais concebidos para e por meios analógicos, estes arquivos permitem renascer materiais nativos digitais, como aqueles que se encontram nas edições *online* dos jornais estudados. A operação de renascimento levanta várias questões, designadamente em termos da codificação dos outputs dos arquivos da web em mensagens visuais. Todas essas questões parecem estar relacionadas com o modo de existência do arquivo da web como um objecto técnico. Arquivos da web, como o Arquivo.pt, dão acesso a capturas de um momento de páginas web de forma dinâmica e interactiva. Esta é a base do contributo desses arquivos para a investigação científica. Ainda que se levantem questões específicas sobre os condicionamentos culturais, sociais e técnicos que os usos dos arquivos da web representem para os cientistas sociais, o seu engajamento com esses arquivos torna-se relevante dado as possibilidades de acesso que permite num contexto de produção digital generalizado e a necessidade do contributo das ciências sociais para reflexão acerca do lugar das tecnologias e dos media digitais nas nossas vidas e na produção de conhecimento científico. Dito de outro modo, o uso de arquivos da web para investigar os fenómenos sociais no actual contexto tecnológico, particularmente sobre comunicação pública, pode ser prolífico. Os arquivos da web podem ser usados como dispositivos técnicos para fornecer acesso interactivo aos investigadores a materiais que sem esses arquivos não poderiam ser recuperados. Contudo, é fundamental que pessoas de todas as áreas científicas a quem este assunto possa interessar do ponto de vista da construção de teoria e de hipóteses de investigação e do ponto de vista das práticas de investigação, especialmente os cientistas sociais, se envolvam nas discussões críticas que nesse contexto emergem. O trabalho de Niels Brügger constituiu um ponto de partida fundamental para, pelo menos, apresentar algumas das questões metodológicas e práticas em causa, mas há muitos mais aspectos a discutir, entre os quais se destaca a relação entre as culturas epistémicas e de investigação da informática e das ciências sociais e os problemas de economia política em causa.

A partir destes enquadramentos, projectou-se a investigação em três passos de análise de aspectos concretos da preservação das edições *online* dos *websites*. Na

primeira etapa, procurou fazer-se uma avaliação técnica e sociológica da preservação e do perfil de preservação dos *websites* actuais dos jornais. Propôs-se o conceito de *camadas de preservabilidade* para pensar o processo de preservação nos seus vários aspectos técnicos diferenciáveis, várias pessoas e até mesmo várias culturas, mentalidades e políticas de preservação. No prolongamento deste conceito, encontrou-se aquilo que o *website* oferece directamente e aquilo que, como não é aí oferecido, se encontra apenas no diálogo com as pessoas que fazem parte das organizações responsáveis pela preservação, neste caso as empresas de *media*, ou consultando materiais adicionais que essas organizações possuem. A súmula dos resultados sugere que o CM, o DN, o *Expresso* e o *Público* não têm uma cultura de preservação digital desenvolvida e informada, nem possuem políticas de preservação digital estruturadas. O que esteve nos seus *websites* pode estar perdido para sempre e o que ainda neles se encontra pode estar prestes a perder-se, concretizando os problemas de memória no jornalismo. O único aspecto transversal em todos os contactos foi o da percepção de que a referência a preservação do material jornalístico é imediatamente associada à preservação de "conteúdos" e não de toda a estrutura das páginas web.

Na segunda fase da investigação empírica, construiu-se uma *rota de investigação* dentro do Arquivo.pt considerando os materiais aí disponíveis para o estudo dos *websites* dos jornais escolhidos. Uma maneira de resolver os problemas levantados pela navegação num arquivo da web consiste no desenvolvimento de uma visão comparativa dos elementos das páginas estudadas. A partir dessa comparação, pode elaborar-se um índice das versões selecionadas para estudo e, a partir desse índice, proceder a uma investigação do material. Este índice foi pensado a partir da perspectiva de que a investigação está ligada à estrutura específica dos arquivos da web utilizados e aos seus próprios meios técnicos. Ao desenho de um curso na internet através de *websites* interrelacionados entre si determinada por critérios para efeitos de investigação deu-se o nome de *rota de investigação.* Trata-se de um trajecto específico dentro do arquivo da web procurando concretizar o tal índice. Para elaborar esta rota, efectuou-se uma pesquisa exaustiva do material disponível, criando-se uma base de dados na forma de tabelas para relacionar dados sobre os endereços encontrados. A pesquisa para constituição de uma rota dentro do Arquivo.pt revelou a existência dos problemas

comuns do material renascido digital dos arquivos da web no caso dos jornais estudados: há pouco material guardado, mas uma enorme quantidade de dados e de informações disponíveis sobre o que foi preservado, muitas páginas e respectivos endereços são difíceis de encontrar porque estão tapados pelas operações de redireccionamento, alguns materiais não têm data correspondente à data prometida na interface do Arquivo.pt, e muitos dos materiais arquivados são problemáticos ao ponto de não poderem ser estudados enquanto "representações" dos *websites* tal como eles teriam sido. Ainda assim, encontrou-se material suficiente para prosseguir com a investigação. Em trabalhos futuros, a ideia de *rota de investigação* pode ser aproveitada para organização, sistematização e indexação manual dos materiais disponíveis de modo a criar colecções digitais. A análise dos endereços encontrados e as mudanças descobertas são muito valiosos do ponto de vista da análise social. Nessas mudanças, pôde observar-se que a gradual constituição de uma economia política da informação em Portugal, sugerida pela evidência de que a alteração de endereços sinaliza a existência de relação entre as empresas de media e empresas de telecomunicações, como a Clix e a Sapo.

Na terceira e última etapa de investigação, procede-se à análise das mudanças estético-funcionais dos *websites* dos jornais estudados. A forma que existe de se chegar aos elementos em mudança dos *websites* é utilizando a interface gráfica do arquivo da web, neste caso do Arquivo.pt, num ecrã de computador alcançando o conjunto desses elementos como diferentes tipos de conjuntos. Parte-se de perspectivs que considera a interface como instância dinâmica de relação entre o humano e a máquina com disposições e mutações técnicas e estéticas específicas para se dar um tratamento descritivo das estruturas em mudança. A grande quantidade de dados condicionou a investigação obrigando à realização não de uma análise pormenorizada e exaustiva de todas essas estruturas com recurso a ferramentas próprias para tal, mas de uma comparação geral dos *websites* na medida em que a interface dos arquivos é fornecedora de mensagens visuais. Em vez de uma análise pormenorizada, fez-se uma comparação atenta a cores, formas e composição dessas mensagens, procurando estabelecer continuidades e descontinuidades da sua disposição ao longo do tempo. A descrição, análise e comparação dos diversos materiais das páginas estudadas permitiu

pensar períodos de continuidade de *layout* dos *websites* e as respectivas rupturas que deram lugar a outro *layout*. A comparação entre todos estes *layouts* conduziu ao esclarecimento das seguintes tendências gerais de desenvolvimentos dos *websites*:

- as páginas web são cada vez mais extensas, sobretudo na vertical;
- as páginas web vão tendo progressivamente uma orientação verticalizada com grandes barras de *scroll*, embora sejam inseridos pequenos espaços com galerias ou pequenas barras de *scroll*;
- as páginas web vão apresentando um acréscimo de estabilidade estética, consolidando a identidade visual do jornal;
- as mudanças são cada vez mais difíceis de assinalar na medida em que deixa são gradualmente menos "macro" e mais "micro", o que implica uma certa adaptação da perspectiva de análise;
- vão surgindo cada vez mais imagens e vídeos, muitas vezes enquadradas em galerias, aumentam também o número de hiperligações, botões, menus e barras de *scroll*.

Em correspondência com os aspectos político-económicos observados na mudança de endereços, também na análise dos *layouts* se observou que sobretudo as empresas de telecomunicações têm uma grande expressão nos *websites* dos jornais, inclusivamente em aspectos do seu *layout*.

A experiência desta investigação permite sugerir dois aspectos a serem pensados e melhorados em investigações futuras. O primeiro desses aspectos diz respeito à discussão conceptural. Em traços largos, pode afirmar-se que as transformações globais em termos tecnológicos trazem a necessidade de que é fundamental discutir o carácter dessas transformações. Parte dessa discussão tem um contexto próprio no âmbito dos arquivos da web, dado a sua situação histórica enquanto arquivos digitais de material digital. Isto significa que é importante ir além das especificidades técnicas dos arquivos da web mergulhando nas dimensões económicas, culturas, políticos e sociais em que estão envolvidos e de que fazem parte. Mas é também importante perceber as especificidades técnicas e modo como essas mesmas especificidades se relacionam com essas dimensões. O segundo aspecto a ser pensado e melhorado diz respeito à necessidade de criar projectos idênticos ao "Investiga XXI", mas com maior duração e

número de pessoas, para que a discussão possa ser mais diversa e plural e para que a investigação empírica possa ser mais exaustiva e pormenorizada.

8 Referências bibliográficas

Abadal, E. & Guallar, J. (2010). The digital press archives of the leading Spanish online newspapers. *Information Research*, 15 (1), disponível em: http://www.informationr.net/ir/15-1/paper424.html.

Alexander, J. C. (2015). The Crisis of Journalism Reconsidered: Cultural Power. *Fudan Journal of the Humanities and Social Sciences*, 8 (1), 9-31.

Appadurai, A. (2003). Archive and Aspiration. In Brouwer, J. & Mulder, A. (Eds.), *Information is alive* (pp. 14-25). Rotterdam: V2_Publishing/NAI Publishers.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 de Junho. Disponível em: www.wired.com/2008/06/pb-theory/.

ArchiveReady (s.d.). ArchiveReady.com: Website Archivability evaluation tool. Disponível em: http://www.archiveready.com/.

Arquivo.pt (2016). Bolsas. Disponível em: http://sobre.arquivo.pt/pt/colabore/id/bolsas/.

Arquivo.pt (s.d.). O que é o Arquivo.pt? Disponível em: http://sobre.arquivo.pt/pt/ajuda/o-que-e-o-arquivo-pt/.

Assmann, J. (2008). Communicative and cultural memory. In Erll, A. & Nünning, A. (Eds.), *Cultural memory studies: An international and interdisciplinary handbook* (pp. 109-118). New York: Walter de Gruyter.

Bastos, H. (2010). *Origens e evolução do ciberjornalismo em Portugal: os primeiros quinze anos (1995-2010).* Porto: Afrontamento.

Bastos, H. (2011). *Ciberjornalistas em Portugal: práticas, papéis e ética*. Lisboa: Livros Horizonte.

Benjamin, W. (2000). *The arcades project*. Cambridge, MA: Harvard Unviersity Press.

Berger, P.L. & Luckmann, T. (2010). *A Construção Social da Realidade: um tratado da sociologia do conhecimento*. Lisboa: Dinalivro.

Boczkowski, P. J. (1999). Understanding the development of online newspapers: Using computer-mediated communication theorizing to study Internet publishing. *New Media & Society*, 1(1), 101–26.

Boczkowski, P. J. (2004). *Digitizing the News: Innovation in Online Newspapers*, Cambridge, MA: MIT Press.

Brügger, N. (2005) *Archiving Websites: General Considerations and Strategies*. Aarhus: Centre for Internet Research.

Brügger, N. (2008). The archived website and Website Philology. *Nordicom Review*, 29 (2), 155-175.

Brügger, N. (2009). Website History and the website as an object of study. *New Media & Society*, 11(1-2), 115-132.

Brügger, N. (2010). *Website Analysis: Elements of a conceptual architecture*. Aarhus: The Centre for Internet Research.

Brügger, N. (2012). When the present web is later the past: Web Historiography, Digital History, and Internet Studies. *Historical Social Research*, 37 (4/142), 102-117.

Brügger, N. (2013). Web Historiography and Internet Studies: Challenges and perspectives. *New Media & Society*, 15 (5), 752-764.

Brügger, N. (2015). *Humanities, Digital Humanities, Media Studies, Internet Studies: An inaugural lecture*. Aarhus: The Centre for Internet Research.

Brügger, N. (2016). Digital Humanities in the 21st century: Digital material as a driving force. *Digital Humanities Quarterly*, 10 (2).

Brügger, N. & Finnemann, N. O. (2013). The web and Digital Humanities: Theoretical and methodological concerns. *Journal of Broadcasting & Electronic Media*, 57 (1), 66-80.

BUDDAH (s.d.). Bursaries. Big UK Domain Data for Arts and Humanities. Disponível em: https://buddah.projects.history.ac.uk/news/bursaries/.

Camponez, C. (2004). A crise do jornalismo face aos novos desafios da comunicação pública. In *Atas dos Ateliers do V Congresso Português de Sociologia Sociedades*

Contemporâneas: Reflexividades e Ação (pp. 9-12), Lisboa: Associação Portuguesa de Sociologia.

Camponez, C. (2012). Jornalismo: o fracasso de um contrapoder entre poderes. In Sá, A., Peixinho, A.T. & Camponez, C. (Orgs.), *Aprofundar a crise: olhares multidisciplinares* (pp. 159-174). Coimbra: Imprensa da Universidade de Coimbra.

Carner, D., McCain, E. & Zarndt, F. (2014). Missing links: The digital news preservation discontinuity. Comunicação apresentada na International Federation of Library Association Satellite Conference, Geneva, Switzerland. Disponível em: https://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-carner-en.pdf.

Carey, J. W. (1977). The problem of journalism history. In Munson, E.S. & Warren, C.A. (Eds.), *James Carey. A critical reader* (pp. 86-94). Minneapolis/London: University of Minnesota Press.

Chalaby, J. (2003). O jornalismo como invenção anglo-americana: comparação entre o desenvolvimento do jornalismo francês e anglo-americano (1830-1920). *Media & Jornalismo*, 3, 29-50.

Cohen, E. L. (2002). Online Journalism as Market-Driven Journalism. *Journal of Broadcasting and Electronic Media*, 46 (4): 532–548.

Costa, M. (2014). *Information Search in Web Archives*. Tese de Doutoramento, Universidade de Lisboa.

Costa, M., Gomes, D. & Silva, M. J. (2016). The evolution of web archiving. *International Journal of Digital Library Systems*, 1-15.

Cowen, N. (2001). The future of the British broadsheet newspaper on the World Wide Web. *Aslib Proceedings*, 53 (5), 189-200.

Cowls, J. (2017). Cultures of the UK web. In Brügger, N. & Schroeder, R. (Eds.), *The Web as History. Using Web Archives to Understand the Past and the Present* (pp. 200-237). London: Press University College London.

Crompton, R. (2008). Forty Years of Sociology. Some Comments. *Sociology*, 42 (6), 1218-1227.

Cunha, D.S. (no prelo). Expresso: Transformações tecnoeconómicas na digitalização de um jornal impresso. In Garcia, J. L. & Martinho, T.D. (Coords.), *Os media em mudança em Portugal: implicações da digitalização no jornalismo*. Lisboa: Entidade Reguladora para a Comunicação Social.

Derrida, J. (1995). Archive fever: A Freudian impression. *Diacritics*, 25 (2), 9-63.

Downie, L., Jr. & Schudson, M. (2009). The Reconstruction of American Journalism. *Columbia Journalism Review*. Disponível em: http://www.cjr.org/reconstruction/the_reconstruction_of_american.php?page=all.

Elias, N. (1997). *Envolvimento e Distanciamento: Estudos sobre sociologia do conhecimento*. Lisboa: Dom Quixote.

Elias, N. (2008). *Introdução à Sociologia*. Lisboa: Edições 70.

Esteruelas, N. C. (2014). Estudio evolutivo del diseño periodístico en Internet: la edición digital de El País (1996-2013). *Zer: Revista de estudios de comunicación*, 37, 137-155.

Featherstone, M. (2000). Archiving cultures. *The British Journal of Sociology*, 51 (1), 161-184.

Featherstone, M. (2006). Archive. *Theory, Culture & Society*, 23 (2-3), 591-596.

Fenton, N. (2010). Drowning or Waving? New Media, Journalism and Democracy. In Fenton, N. (Ed.), *New media, old news. Journalism & democracy in the digital age* (pp. 3-16). London/Thousand Oaks/New Delhi/Singapore: Sage.

Floridi, L. (2012). Big Data and Their Epistemological Challenge. *Philosophy and Technology*, 25 (4), 435-437.

Fidalgo, J. (2006). O lugar da ética e da auto-regulação na identidade professional dos jornalistas. Tese de Doutoramento, Universidade do Minho.

Fidalgo, J. (2016). Disputas nas fronteiras do jornalismo. In Gomes, R. (Ed.) (2016). *Digital Media Portugal – ERC 2015* (pp. 35-47). Lisboa: Entidade Reguladora para a Comunicação Social.

Finnemann, N. O. (2002). Perspectives on the Internet and Modernity Late Modernity, Postmodernity or Modernity Modernized? In Brügger, N. & Bødker, H. (Eds.), *The Internet and society?: questioning answers and answering questions* (pp. 29-38). Århus: The Centre for Internet Research.

Foucault, M. (2014). *A arqueologia do saber*. Lisboa: Edições 70.

Frade, C. (2016). Social Theory and the Politics of Big Data and Method. *Sociology*, 50 (5), 863-877.

Garcia, J.L. (2007). O Novo Ethos Científico e a Biotecnologia. *Revista de Comunicação e Linguagens*, 38, 207-219.

Garcia, J. L. (2009a). Introdução ao estudo dos jornalistas portugueses: os jornalistas e as contradições do capitalismo jornalístico no limiar do século XXI. In Garcia, J. L. (Ed.), *Estudos sobre os Jornalistas Portugueses - Metamorfoses e Encruzilhadas no Limiar do Século XXI* (pp. 23-46). Lisboa: Imprensa de Ciências Sociais.

Garcia, J. L. (2009b). Principais Tendências de Profissionalização dos Jornalistas no Período Pós-Transição Democrática. In José Luís Garcia (Eds.), *Estudos sobre os Jornalistas Portugueses - Metamorfoses e Encruzilhadas no Limiar do Século XXI* (pp. 63-91). Lisboa: Imprensa de Ciências Sociais

Garcia, J.L. (2016). A Web, a expansão do digital e a nova economia política da informação. In Gomes, R. (Ed.) (2016). *Digital Media Portugal – ERC 2015* (pp. 9-13). Lisboa: Entidade Reguladora para a Comunicação Social.

Garcia, J.L. & Martins, H. (2009). O ethos da ciência e suas transformações contemporâneas, com especial atenção à biotecnologia. *Scientiae Studia: Revista Latino-Americana de Filosofia e História da Ciência*, 7 (1), 83-104.

Garcia, J., Marmeleira, J. & Matos, J.N. (2014). Incertezas, vulnerabilidades e desdobramento de atividades. In Rebelo, J. (Ed.), *As novas gerações de jornalistas portugueses* (pp. 9-19). Lisboa: Mundos Sociais.

Garcia, J. L. & Martinho, T.D. (Coords.) (no prelo). *Os media em mudança em Portugal: implicações da digitalização no jornalismo*. Lisboa: Entidade Reguladora para a Comunicação Social.

Ghosh, P. (2015). Google's Vint Cerf warns of 'digital Dark Age'. *BBC News*. Disponível em: http://www.bbc.com/news/science-environment-31450389.

Gomes, A. (2012). Crise, novo paradigma, o próximo jornalismo e arbitragem da ética. *Atas do VII Congresso Português de Sociologia: Sociedade, Crise e Reconfigurações.* Porto: Associação Portuguesa de Sociologia.

Gomes, D. & Silva, M. J. (2006). Modelling Information Persistence on the Web. In the Sixth International Conference on Web Engeneering, Palo Alto, USA. Disponível em: http://visibilidade.net/daniel/docs/papers/gomes06urlPersistence.pdf.

González, M.A.C. (2000). Convivencia de la prensa escrita y la prensa on line en su transición hacia el modelo de comunicación multimédia. Disponível em: http://www.ucm.es/info/perioI/Peri od_I/EMP/Numer_07/7-4-Comu/7-4-0 1.htm.

Grabowicz, P. (2014). Tutorial: The Transition To Digital Journalism. Disponível em: https://multimedia.journalism.berkeley.edu/tutorials/digital-transform/.

Halford, S. & Savage, M. (2017). Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research. *Sociology* (Online First). Disponível em: http://journals.sagepub.com/doi/abs/10.1177/0038038517698639.

Hartley, J. (1996). *Popular reality: journalism, modernity, popular culture*. London: Arnold;

Hemeroteca (s.d.). Hemeroteca Digital. Um projecto em construção. Disponível em: http://hemerotecadigital.cm-lisboa.pt/

Hui, Y. (2012). What Is a Digital Object?. *Metaphilosophy*, 43 (4), 380–395.

IIPC (s.d.). About the IIPC. Disponível em: http://netpreserve.org/about-us/.

Jenkins, H. (2006). *Convergence culture: where old and new media collide*. New York: New York University Press.

Joly, M. (1999). *Introdução à análise da imagem*. Lisboa: Edições 70.

Kawamoto, K. (2003). *Digital Journalism: Emerging media and the changing horizons of journalism.* London: Rowman & Littlefield Publishers.

Karlsson, M. & Sjøvaag, H. (2016a). Introduction. Research methods in an age of digital journalism. *Digital Journalism*, 4 (1), 1-7.

Karlsson, M. & Sjøvaag, H. (2016b). Content Analysis and Online News. Epistemologies of analysing the ephemeral Web. *Digital Journalism*, 4 (1), 117-192.

Kitch, C. (2008). Placing journalism inside memory – and memory studies. *Memory Studies*, 1 (3), 311-320.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: Sage.

The Data Revolution Big Data, Open Data, Data Infrastructures and Their Consequences Rob Kitchin

Kovach, B. & Rosenstiel, T. (2004). *Os elementos do jornalismo: o que os profissionais do jornalismo devem saber e o público deve exigir*. Porto: Porto Editora.

Kranzberg, M. (1986). Technology and History: "Kranzberg's Laws". *Technology and Culture,* 27 (3), 544-560.

Lévy, P. (1994). *As tecnologias da inteligência: o futuro do pensamento na era informática*. Lisboa: Instituto Piaget.

Lewis, S.C. & Westlund, O. (2015). Big Data and Journalism. Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3 (3), 447-466.

Library of Congress (s.d.). About the Program. Disponível em: https://www.loc.gov/ndnp/about.html.

Lopes, F. (2015). *Jornalismo: profissão ameaçada*. Lisboa: Alêtheia Editores.

Mahrt, M. & Scarkow, M. (2013). The Value of Big Data in Digital Media Research. *Journal of Broadcasting & Electronic Media*, 57 (1), 20-33.

Mancini, P. (2013). Globalization and communication. Research what scholars can learn from the crisis of journalism. *International Journal of Communication*, 7, 127–136.

Manovich, L. (2001). *The Language of New Media*. Cambridge: MIT Press.

Martins, H. (1996). *Hegel, Texas e outros Ensaios de Teoria Social*. Lisboa: Edições Século XXI.

Mateo, R., Bergés, L. & Garnatxe, A. (2010). Crisis, what crisis? The media: business and journalism in times of crisis. *Triple C: Cognition, Communication, Co-operation*, 8(2), 251–274.

Matos, J.N., Baptista, C., & Subtil, F. (Eds.) (2017). *A crise do jornalismo em Portugal*. Lisboa: Le Monde Diplomatique e Deriva.

Mauss, M. (2005). *Sociologia e Antropologia*. São Paulo: Cosac & Naify.

Mazzocchi, F. (2015). could big data be the end of theory in science? A few remarks on the epistemology of data-driven science. *EMBO Rep*., 16 (10), 1250-1255.

McChesney, R. (2013). *Digital Disconnect. How Capitalism is Turning the Internet Against Democracy*. New York: New Press.

McFarland, D.A., Lewis, K. & Goldberg, A. (2016). Sociology in the Era of Big Data: The Ascent of Forensic Social Science. *The American Sociologist*, 47 (1), 12-35.

McManus, J. (1994). *Market-Driven Journalism*. *Let the Citizen Beware*. Thousand Oaks: Sage.

Mendonça, P.F.X. (2010). Objectos, poder e oculto – Sobre a experiência do ecrã. *Comunicação e Sociedade,* 17, 51-66.

Millward, G. (s.d.). Digital barriers and the accessible web: disabled people, information and the internet. In BUDDAH, Bursaries. Big UK Domain Data for Arts and Humanities. Disponível em: http://sas-space.sas.ac.uk/6104/1/Millward%20case%20study.pdf.

Miranda, J. B. (1996). A virtualização do arquivo. *Revista da Faculdade de Ciências Sociais e Humanas*, 9, 95-117.

Mützel, S. (2015). Facing Big Data: Making sociology relevant. Big Data & Society, 2 (2).

Nanni, F. (2013). L'archiviazione delle pagine dei quotidiani online. *Diacronie. Studi di Storia Contemporanea*, 15 (3). Disponível em: http://diacronie.revues.org/418.

Nanni, F. (2017). Reconstructing a website's lost past – Methodological issues concerning the history of www.unibo.it. *Digital Humanities Quarterly*, 11 (2). Disponível em: http://www.digitalhumanities.org/dhq/vol/11/2/000292/000292.html.

National Research Council. (2014). *Convergence: Facilitating transdisciplinary integration of life sciences, physical sciences, engineering and beyond*. National Research Council.

Nerone, J. & Barnhurst, K.G. (2001). Beyond modernism. Digital design, Americanization and the future of newspaper form. *New Media & Society*, 3 (4), 467-482.

Nora, P. (1989). Between memory and history: *Les Lieux de Mémoire*. *Representations*, 26, 7-24.

Palácios, M. (2003). Jornalismo online, informação e memória: apontamentos para debate. In Fidalgo, A. & Serra, P. (Orgs.), *Informação e comunicação online, Volume 1, Jornalismo online* (pp. 75-89). Covilhã: Universidade da Beira Interior.

Pavlik, J. (2000). The Impact of Technology on Journalism. *Journalism Studies*, 1 (2), 229-237.

Pavlik, J. V. (2001). *Journalism and New Media.* New York: Columbia University Press.

Pinto-Coelho, Z. (2010). Da instabilidade do ecrã. In Pinto-Coelho, Z. & Neves, J. (Eds.), *Ecrã, Paisagem e Corpo* (pp. 17-33). Coimbra: Grácio Editor.

Pizarroso Quintero, A. (Coord.) (1996). *História da Imprensa*. Lisboa: Planeta Editora.

Presnell, J. L. & Morris, S. E. (2017). The Historical Newspaper Crisis: Discoverability, Access, Preservation, and the Future of the News Record. Comunicação apresentada na International Federation of Library Association Satellite Conference, Reyjavík, Iceland. Disponível em: https://www.ifla.org/files/assets/newspapers/2017_Iceland/2017-presnell-en.pdf.

Proulx, S. (2001). Les recherches nord-américaines sur la communication: l'institutionnalisation d'un champ d'etudes. *L'Année sociologique*, 50 (2), 467-485.

Rebelo, J. (2011). *Ser Jornalista em Portugal: Perfis Sociológicos*. Lisboa: Gradiva.

Richardson, J. E. (2007). *Analysing newspapers: an approach from critical discourse analysis*. New York: Palgrave Macmillan.

RJI (2017). Dodging the Memory Hole 2017. Disponível em: https://www.rjionline.org/events/dodging-the-memory-hole-2017.

Rokeby, D. (2003). Construire l'expérience. L'interface comme contenu. In Poissant, L. (Dir.), *Interfaces et sensorialité* (pp. 91-114). Québec: Presses de l'Université du Québec/Publications de l'Université de Saint-Étienne.

Ruellan, D. (1993). *Le professionalisme du flou. Identité et savoir-faire des journalistes français*. Grenoble: Presses universitaires de Grenoble.

Sá, A., Peixinho, A.T. & Camponez, C. (Orgs.) (2012). *Aprofundar a crise: olhares multidisciplinares.* Coimbra: Imprensa da Universidade de Coimbra.

Santos, B. S. (2010). *Um Discurso sobre as Ciências*. Porto: Afrontamento.

Savage, M. & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, 41 (5), 885-899.

Schudson, M. (2003). *The sociology of news*. New York: W.W. Norton & Company.

Schudson, M. (2007). O modelo americano de jornalismo: excepção ou exemplo. *Comunicação e Cultura*, 3, 115-130.

Scott, B. (2005). A Contemporary History of Digital Journalism. *Television & New Media*, 1 (6), 89-126.

Silva, A. S. (1986). A ruptura com o senso comum nas ciências sociais. In Silva, A.S. & Pinto, J. M. (Eds.), *Metodologia das Ciências Sociais* (pp. 29-53). Porto: Afrontamento.

Skinner, K. and Schultz, M. (2014). *Guidelines for Digital Newspaper Preservation Readiness*. Atlanta, GA: Educopia Institute.

Smith, M. (2008). Introduction. Visual Culture Studies: History, Theory, Practice. In Smith, M. (Ed.), *Visual Culture Studies* (pp. 1-16). London: Sage.

Sousa, H. (2006). Information *Technologies*, Social Change and the Future. The Case of Online Journalism in Portugal. *European Journal of Communication*, 21 (3), 373-387.

Tirohl, B. (2000). The photo-journalist and the changing news image. *New Media & Society*, 2 (3), 335-352;

UNESCO (2004). Charter on the Preservation of Digital Heritage [Resolução adoptada no relatório da Comissão V no 18.º encontro plenário, a 15 de Outubro de 2003]. In UNESCO, *Records of the General Conference* [32nd Session], Volume 1 *Resolutions* (pp. 74-77). Paris: United Nations Educational, Scientific and Cultural Organization.

Urbano, J. & Guarda, D. (2001) O mundo dos compossíveis [Entrevista a Hermínio Martins]. *Número Magazine*, 11, 112-121.

Vattimo, G. (1992). *A Sociedade Transparente*. Lisboa: Relógio D'Água.

WARD, Mike (2002) *Journalism Online*. Oxford: Focal Press.

Weber, M. (1997). *Conceitos Sociológicos Fundamentais*. Lisboa: Edições 70.

Weber, M. S. (2017). The tumultuous history of news on the web. In Brügger, N. & Schroeder, R. (Eds.), *The Web as History. Using Web Archives to Understand the Past and the Present* (pp. 83-100). London: Press University College London.

Webber, R. (2009). Response to 'The Coming Crisis of Empirical Sociology': An Outline of the Research Potential of Administrative and Transactional Data. Sociology, 43 (1), 169-178.

Wiener, N. (1954). *The human use of human beings: cybernetics and society*. New York: Doubleday Anchor.

Winner, L. (1980). Do Artifacts Have Politics?, *Daedalus*, 109 (1), 121-136.

Williams, K. (2006). Competing models of journalism? Anglo-American and European reporting in the information age. *Journalistica*, 2, 43-65.

Wikipedia (s.d.a). List of online newspaper archives. *Wikipedia*. Disponível em: https://en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives.

Wikipedia (s.d.b). URL redirection. Disponível em: https://en.wikipedia.org/wiki/URL_redirection.

Winters, J. (2017). Coda: Web archives for humanities research – some reflections. In Brügger, N. & Schroeder, R. (Eds.), *The Web as History. Using Web Archives to Understand the Past and the Present* (pp. 238-248). London: Press University College London.

Zamith, F. (2008). Uma Proposta Metodológica para analisar o Aproveitamento das Potencialidades Ciberjornalísticas da Internet. *Observatorio (OBS*) Journal*, 2 (2), 165-191.

Zamith, F. (2013). *A contextualização no ciberjornalismo*. Porto: Afrontamento.

Zelizer, B. (2008). Why memory's work on journalism does not reflect journalism's work on memory. *Memory Studies*, 1 (1), 79-87.

Zelizer, B. (2012). A "Voz" do Visual na Memória. *Media & Jornalismo*, 20*,* 19-42.