Introduction: Digital methods for the exploration, analysis and mapping of e-diasporas

Dana Diminescu

TelecomParisTech and Fondation Maison des Sciences de l'Homme, Paris

One of the major changes affecting diasporas the world over since the 1980s has been the increasing number of communities scattered throughout physical space, along with new forms of presence, regrouping, interaction and mobilization within digital territories.

This change calls for a renewal in epistemological approaches. The topics under study, as well as the conceptual and methodological tools used to analyse them, need to be reconsidered in the face of this evolution of diasporas. The articles published in this issue of SSI[1] bear witness to such an effort: researchers and engineers involved in the *e-Diasporas Atlas*[2] project have sought to find the most appropriate concepts, tools and methods to explore the Web of diasporas, based on a number of case studies. This work represents a vast new area of investigation, which is still under way.

In this introduction, we examine the different conceptual tools used during the research, analyse their relevance for the different diasporic communities on the Web and present the methodological chain developed within the *e-Diasporas Atlas* project as well as the most important findings.[3]

The concepts

In several articles, and specifically in another issue of SSI, we have shown the emergence of a new migrant figure: the connected migrant.[4] S/he is no longer defined solely by life-experiences of disruptions and antagonisms – which have constantly been upheld as the organizing principles of any theoretical reflection on the uprooted migrant and his 'twofold absence' – but by different forms of 'presence at a distance'.

This change, which we had initially studied at the fundamental level of the migrant him/herself, can also be observed at the collective level of diasporas and transnational networks.

What kinds of diasporas are formed by connected migrants? Do the networks and interactions of migrants scattered throughout the world, which we have been able to

Corresponding author:
Dana Diminescu, FMSH, 190 avenue de France, Paris 75013, France
Email: dana.diminescu@telecom-paristech.fr

observe and visualize through an exploration of their traces on the Web, reveal traditional or novel functions of diasporas? Are these 'e-diasporas' an extension of physical diasporas, or merely their mirror image? Are they the source of new diaspora communities? Or are they, instead, an echo-chamber of globalization – of a society which is itself a diaspora in the making? All these questions show how difficult it is to find a generally accepted definition of what an e-diaspora is. Discussions around this concept are not settled, even after many years of work. In fact, this publication aims at stimulating further debate.

Historically, the emergence of e-diasporas occurred along with the diffusion of the Internet and the development of multiple online public services. At the end of the 1990s, a number of institutions joined forces with the new 'e'-technologies (e-administration, e-democracy, e-education, e-healthcare, e-culture, e-tourism), which gave rise to the first presence on the Web of associations run by migrant populations. If the earliest websites were those produced by IT professionals, we soon saw the diffusion of the Web into all the *diasporic* communities and at all levels within them. The past ten years have witnessed the use of both Webs 1.0 and 2.0 (blogs) in these communities as well as the widespread appropriation of the various social-networking platforms (Facebook, Twitter, LinkedIn, etc.).

What we call *e-diaspora* is a migrant collective that organizes itself and is active first and foremost on the Web: its practices are those of a community whose interactions are 'enhanced' by digital exchange. An e-diaspora is also a dispersed collective, a heterogeneous entity whose existence rests on the elaboration of a common direction, a direction not defined once and for all but which is constantly renegotiated as the collective evolves. An e-diaspora is an unstable collective because it is redrawn by every newcomer. It is self-defined, as it grows or diminishes not by inclusion or exclusion of members, but through a voluntary process of individuals joining or leaving the collective – simply by establishing hyperlinks or removing them from websites.

An e-diaspora is both 'online' and 'offline'. We are therefore interested in both the digital 'translations' of 'physical' actors/phenomena (the online activities of associations, for example) and the specifically ('natively') digital actors/phenomena (e.g. a forum and its internal interactions), which are sometimes called pure players. The question of 'rub-offs' – reciprocal influence between these two sorts of Web entities – is of capital importance in analysing an e-diaspora. It is thus clear that the research carried out in the context of the *e-Diasporas Atlas* presupposes knowledge of the diaspora in question and also implies knowledge of the Web and an appreciation of the singularity of the exchanges that take place there.

We prefer the term 'e-diaspora' to that of 'digital diaspora' because the latter may lead to confusion, given the increasingly frequent use of the notions of 'digital native' and 'digital immigrant', in a 'generational' sense (distinguishing those born before from those born during/after the digital era). The object of the *e-Diasporas Atlas* is not this 'digital migrant', however, but the *connected migrant* in his/her social and institutional context.

An *e-diaspora corpus* is a list of websites. The constitution of a corpus of websites is the method used to 'capture' an e-diaspora. To start exploring the Web of diasporas, we first need to identify diasporic websites.

A '*diasporic*' *website* or migrant website is a website created or managed by migrants and/or that deals with them (at any rate, a site for which migration or diasporas is a defining theme). This can be a personal site or blog, the site of an association, a portal/forum, an institutional site, or anything similar. Usage is not the criterion: a site often consulted by migrants (a media site, for example) is not necessarily a migrant site. What distinguishes 'activity' is first and foremost the production of content and the practice of citation (hyperlinks). On the other hand, a migrant site need not necessarily be located in a foreign country; it may just as easily be in the country of origin. Migrant sites testify to a given e-diaspora's occupation of the Web.

Some researchers have built their collection by starting with websites they already knew from the fieldwork they had done on migrations (Ingrid Therwath, Houda Asal, Tristan Bruslé, Yann Scioldo Zürcher). Others have used keywords and Google (Priya Kumar, Anat Ben-David, Francesco Mazzucchelli) or Twitter (e.g. through the hashtag #right2vote, for Marta Severo and Eleonora Zuolo) before crawling the Web with the tools of the *e-Diasporas Atlas*. The choice and classification of websites relies entirely on the researcher's expertise. Such choices have triggered debates which are still ongoing within the *e-Diasporas Atlas* community of researchers. This community is multidisciplinary, and the diversity of its members can be seen in the collection and categorization of websites. Houda Asal chose to work only on the websites of associations from the Lebanese diaspora. Ingrid Therwath chose to focus on websites related to the *hindutva* ideology. However, nearly all of them decided to collect official websites for a better understanding of the links between diasporas and their countries of origin. And all of them analysed the contents of each website, before deciding if they should keep it in their collection. The domain names, the languages used, the type of publication, the geolocation of the website, these are categories which have been used for every one of the case studies.

Frontier sites or neighbouring sites have also been collected.

A *neighbouring site* is a non-migrant site (or one belonging to an e-diaspora other than the one being studied) which distinguishes itself by its strong connection with the (migrant) sites of a given e-diaspora (governmental or media sites of the country of origin, for example). However, not every site strongly linked to an e-diaspora is necessarily a neighbouring site. To be one it needs to be 'specific' to the diaspora in question, which is why sites 'on the fringes of' the majority of Web communities, particularly those in the upper layers of the Web, Google, YouTube, Facebook and so on, are not counted as 'neighbours'.

In the *e-Diasporas Atlas*, a list of neighbouring sites may be drawn up alongside that of migrant ones. These neighbouring sites discovered during the prospecting phase are not crawled during subsequent prospection but only during the phase of validation so as to gather together all links with the migrant site.

An ensemble of the 'migrant sites' and 'neighbouring sites' (see below) of a given diaspora – whether such sites are 'living' or 'dead' – constitutes the *web of diasporas*. In a sense, this can be understood as the web ecosystem of a diaspora.

e-Diasporas methodology

The digital methodological chain and the tools we developed for building the *e-Diasporas Atlas* aim at mapping and analysing the occupation of the Web by diasporas. The chain is

composed of four interlocking steps: (1) equipped Web exploration and corpus building; (2) data enrichment (location, languages, text-mining); (3) network visualization-manipulation and graph interpretation; (4) collaborative sharing of (raw) data and findings.

Step 1: Web exploration

In order to compile a chapter of the Atlas, the first step is to build (and circumscribe) a *corpus* of websites. As we have already highlighted, the researcher plays a crucial role in this process inasmuch as his/her knowledge of the fieldwork allows him/her to select with discrimination the relevant resources for a given diaspora.

An e-diaspora is 'captured' by putting together a corpus of websites. This method entails breakdown and selection processes that allow a diaspora web to be extracted. But definition is also necessary because an e-diaspora presents itself to the researcher only as a product of this 'excision' performed upon the Web. Similarly, it is only because of such exploration/selection, the filtering/circumscription of a corpus, that what a migrant site actually is takes on meaning.

In order to complete this stage of *collection*, the researcher needs to be *equipped*. The identification of relevant websites is achieved semi-automatically thanks to a software called *Navicrawler*, which makes it possible to scan *web grounds* using a web-browser. Navicrawler is a Firefox add-on designed and developed by Mathieu Jacomy. The interface is located on the left of the currently browsed page.

Navicrawler works essentially by scraping the out-links of the visited websites (listed and stored as 'Next Sites'). The researcher can then incorporate each website into the corpus, where it becomes an 'in site', or can reject it, and then it becomes an 'out site'. The researcher can also describe the websites by adding tags.

The logic of exploration induced by Navicrawler combines browsing and crawling. Unlike automatic crawling, it allows the researcher to perceive the *context of links* and thus to avoid a blackbox effect. At the end of this exploration stage, s/he is able to export his/her corpus as a graph in which the nodes represent the websites and the edges stand for the links between them.

Step 2: Data enrichment (digital toolbox)

The social scientist plays a central role throughout the process of corpus building and description/enrichment. However, s/he can be assisted in the content analysis by automatic tools. Our research team developed a *digital toolbox* that makes possible various 'enrichment processes', which include:

–   Retrieving from a list of URLs the information provided by the registrar about the registrant (owner of the domain name), especially his/her geographical location, about the server hosting the website, etc.
–   Text-mining used on the index of the corpus in order to retrieve *named entities*: people, organizations, places, etc. (using Open Calais API).
–   Recognition of the *language*s used in each website (and the distribution of languages in order to study multilingualism, an important issue in migration studies).

Step 3: Network visualization

In order to visualize the exploration data, in other words to map the corpus previously built, we use a graph-visualization software called *Gephi*, a project initiated and hosted at first by our research team. The software was developed by Mathieu Bastian, Mathieu Jacomy and Sébastian Heymann. This tool allows the user to spatialize and manipulate the corpus network. Two types of visualization are available:

–   a spatialization based on the physical principle of attraction/repulsion (according to the presence or absence of a link between two nodes);
–   a geographical spatialization that uses geocoded data: location of website owner, website addresses, servers, etc. (especially information retrieved during the data-enrichment stage).

Note: It must be stressed that the graph is a tool for the researcher, and not a 'photograph' of a given diaspora, while recognizing that our cartography represents a fragment of the Web of diasporas, a snapshot at a particular point in time.[5]

Step 4: Collaborative platform http://maps.e-diasporas.fr

The *maps.e-diasporas.fr* platform is a *collaborative* platform initially developed and implemented by Mathieu Jacomy and the ICT Migrations team for hosting the *e-Diasporas Atlas*. It is a tool for publishing and sharing research findings among scientific communities. The platform comprises *chapters* (in our case, the various diasporas) and provides for each of them the following data:

–   *Maps*: browsable graphs of the corpus, with different views according to the fields of classification.
–   *Raw data*: the empirical data (texts, videos, images, interviews, etc.) produced/retrieved and used during the research. The *e-Diasporas Atlas* is part of the more general 'digital humanities' project to provide access; it diffuses not only the research results but also the research data.
–   *Statistics*: these are automatically generated from both the classification and the graph structure; they provide quantitative data about the relations between categories/actors. Statistics help strengthen the hypothesis formulated from the graph visualization.

Some significant findings

The *e-Diasporas Atlas* proposes at least two interpretation keys: (1) a topological key, centred on analysis of the connectivity between the actors on the Web; and (2) a quanti-qualitative key, which provides information derived from exploration of the contents of every site and from confrontation with the fieldwork and the expertise of all researchers involved in this project.

The comparison between geographical networks and networks observed on the Web seemed an obvious approach and gave rise to a few recurring observations, which included the following. The Web of diasporas does not match the geographical and demographic distribution of the dispersed populations. A large majority of e-diaspora sites are geolocated in North America and especially in the United States. This predominance is even more surprising when the presence on the Web strongly contrasts with the presence in the geographical areas, as was observed for the Palestinian, Nepalese or Egyptian diasporas. The online *hindutva* diaspora is thus very territorialized and symbolically linked to India, while operating from the United States. In this reallocation of geographical distribution due to presence or activity on the Web, the dominance of the English language and the majority practice of the Web favours English-speaking countries in the Web of diasporas. When English isn't dominant, the linguistic composition of the diaspora web reflects, above all, the diaspora's establishment in the host countries; very few websites operate exclusively in the language of origin.

Wishing to understand the collective life of the e-diasporas, we sought to identify clusters of expatriates on the Web and the exceptions to these clusters, their centrality, their hierarchies, their relations and their *assemblage*. The categorization of websites has enabled the identification of actors and clusters, and provides a glimpse, before going into further details, of: associations unaware of one another, bloggers creating their own world, activist groups or individuals seizing power on the Web and sometimes, as in the case of the Arab Spring, even managing to spark popular dissent and to impact on political events.

The absence of links between associations' websites cropped up repeatedly in the Web of diasporas. We also discovered, for instance, that relationships between different Lebanese organizations was problematic. The internal fragmentation of the Lebanese community sector appears to have been accentuated by the Web, and some alliances which exist on the ground are not visible in the graphs, as Houda Asal shows in this issue. Marta Severo observes that Egyptian associations based in different countries don't mention each other and have no common 'neighbour' websites. The websites of Palestinian associations are more frequently linked to frontier websites than amongst themselves.

When considering the categorization by publisher, the media are unquestionably the bridge nodes linking peripheral websites and institutional clusters. This may be explained by the fact that most associations' websites have a 'news' page related to the country of origin that links to articles published in major international or national media.

Mapping the e-diasporas enables us to analyse, among other things, the relationships undertaken and maintained by various diaspora actors with their homeland and with their institutions: clearly labelled links with a strong state, as in the case of France or India (an emergent state where we find a high visibility of government sites, which seek to attract the most privileged and influential migrants); one-way or even non-existent links, as in the case of Macedonia, Nepal and Lebanon for instance.

Two contributions to this special issue are particularly relevant for an understanding of the importance of neighbouring sites, and their relationship to diasporic websites.

In the corpus of the Palestinian diaspora put together by Anat Ben-David, 72% of the archived websites are neighbouring sites and 22% are diasporic websites. Together they form a densely knit network organized around two centres of gravity: the Palestinian cause and the Palestinian Territories. While analysing the dynamics of the Palestinian

diaspora as it emerges on the Web, Ben-David observes that it is no longer defined around Palestine as a *place of origin*, but rather that it is constructed around Palestine as a *place of reference*. Ben-David argues that its organization is built less on a network of family, social and transactional ties between *communities* of Palestinians who have been dispersed to many places in the world, and more on global *advocacy networks* that transcend their immediate social networks. Its members are no longer only *Palestinians abroad*, but also *natives of the host countries* who identify with the Palestinian cause.

Concerning the Hindu diaspora, Ingrid Therwath identifies a corpus of websites which 'mirrors' the mother-organization as well as three types of frontier websites: American associations located in the institutional neighbourhood of the Sangh Parivar in the United States; generalist conservative American websites like Fox News or neo-liberal think tanks, located in the neighbourhood of blogs and non-institutional *hindutva* websites in India; and, between these two locations, Therwarth discovers a cluster of particularly virulent Jewish diaspora groups opposed to Muslims. Beyond the common Islamophobic discourse, this neighbourhood, which juxtaposes pro-*hindutva* groups and extremist Jewish groups, is particularly interesting in that it puts into contact diasporic groups from different regions.

The self-organization typical of Web networks facilitates the emergence of decentralized communities and acts as an ideal platform for different forms of mobilization. This is the case, for instance, of the Egyptian political e-diaspora, the nationalist religious *hindutva* movement, the memorial activism of French colonial repatriates, the Tunisian cyber-dissidence, the transnational solidarity mobilization in support of Tamil rights, or the boycott movement against Israeli commercial and cultural products.

In all these cases, we have observed that only part of a diaspora is active on the Web, and that among them only a small minority are involved in political action – despite their visibility and their dominance on the Web. The expansion of this activism depends on the events at hand and is generally associated with a specific context such as historical commemorations, radical regime changes or highly contested elections. In such cases, the movements are often instrumentalized by the home country, and gain access to the global public sphere through alternative media and non-diasporic actors.

Khachig Tölölyan wrote, in a famous article, 'Rethinking diaspora(s): stateless power in the transnational moment' (1996), that: '*Where once were dispersions, there now is diaspora*'. To conclude, we can paraphrase him by saying that *where once were diasporas, there now is the Web*… Most certainly, populations from different diasporas are now disseminated both throughout the world and on the Web – a phenomenon which deserves to be thoroughly studied.

Funding

Notes

1   Special thanks go to Matthieu Renault and Anne Rocha-Perazzo for their contribution.
2   New communication and organization practices have produced a vast, moving e-corpus, whose exploration, analysis and archiving have never before been attempted. The *e-Diasporas*

Atlas is the first of its kind, and is the fruit of the efforts of more than 80 researchers worldwide, with some 8000 migrant websites archived and observed in their interactions.

> The *e-Diasporas Atlas* was incubated and developed at the Fondation Maison des Sciences de l'Homme ICT Migrations program. Initiated and coordinated by Dana Diminescu, the project introduced digital methods into research on diasporas. This was made possible by the R&D innovations of Mathieu Jacomy and the technical coordination and training provided by Matthieu Renault. Some eighty researchers from diverse disciplines, laboratories and countries took part in the project. Several partners also contributed to its success: the Institut National de l'Audiovisuel, the Centre National de la Recherche Scientifique through its Migrinter laboratory, the Institut Mines-Telecom, Linkfluence and the design studio Incandescence. The *e-Diasporas Atlas* will continue to grow in the years to come.

3   For general definitions of the technical terms used visit the subsection 'Learn more about our concepts, tools and methodology' at the website: http://www.e-diasporas.fr

4   Diminescu D (2008) The connected migrant: An epistemological manifesto. *Social Science Information* 47(4): 565–579.

5   The coloured graphs and subgraphs produced by and for the contributors to this Special Issue are not reproduced within the body of each article, but have been brought together in an appendix section located at the end of the issue and can be accessed at: http://dx.doi.org/10.1177/0539018412456918.

Author biography

Dana Diminescu, a practising sociologist, is an associate professor at TelecomParisTech. Her empirical work has enabled her to approach a variety of fields: e.g. uses of mobile telephone and voice IT, the Internet (tailing, archiving, mapping of the Web), and identifying digitalization technologies and m-transactions by migrants. Since 2003 she has been the scientific director of the research program ICT Migrations at the Fondation Maison des Sciences de l'Homme, Paris. This program, which she launched 10 years ago, has made major contributions to the theorization and analysis of the 'connected migrant'. She is also coordinator of the *e-Diasporas Atlas*, for which she developed a digital methodological chain and tools for mapping and analysing the occupation of the Web by diasporas.