Towards Internet Domain Names in Tifinagh

Abderrahman Ait Ali
PhD Candidate – KTH Stockholm – Member of ISOC Morocco & Sweden
abde@kth.se
Anass Sedrati
PhD Candidate – INPT Rabat – Member of ISOC Sweden
anass@kth.se

Abstract
As a global network of networks, the Internet is nowadays one of the most important tools for cultural and information exchange. With its increasing popularity, the Internet is used by more and more people from different cultural and particularly linguistic backgrounds. Language scripts such as Tifinagh are thus increasingly used. In order to integrate such language scripts, The International Corporation for Assigned Names and Numbers (ICANN) has created Internationalized Domain Names (IDN) which allow URLs to be used in different language scripts. The Amazigh language, with its script (Tifinagh), is currently an official language in Morocco. The Royal Institute of the Amazigh Culture (IRCAM) has as central objective to promote the Amazigh culture, and particularly Tifinagh both nationally, regionally and locally (IRCAM, 2016). In this sense, this paper aims at promoting the use of the language and its script in the Internet through integrating Tifinagh in URLs.

Keywords: Internet domain names, URL, DNS, language scripts, Amazigh language, Tifinagh, IRCAM, ICANN.

1. Introduction

The Internet is defined as a global interconnected system of computer networks. It can be seen as a network of sub-networks connected using a different wired or wireless communication technologies. The Internet started in the 1980s with ARPANET in the United States, the first network of interconnected regional academic and military networks. It became globally accessible when CERN launched the World Wide Web (WWW) in 1989 (Stewart, 2000). At this stage and as the result of the origin of the Internet, most of the internet content was in English and Latin script. Thus, early computer systems were limited to the characters in the American Standard Code for Information Interchange (ASCII) which allows to write in a subset of Latin script (IANA, 2007). In particular, domain names that are used in the internet protocols were always in Latin script. However, little by little the number of internet users in the world increased to reach a larger and global extent (Figure 1). Due to this rise in the number of internet

users, many different languages started to be used, particularly languages based on scripts other than the Latin one.

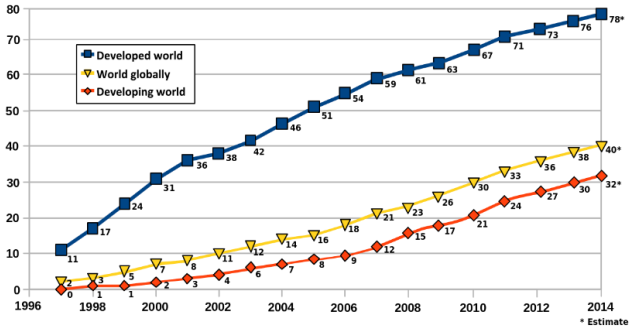

Figure 1- Internet users per 100 inhabitants (ITU, 2015)

An internalized domain name (IDN) was needed to allow internet domain names to contain language-specific scripts such as Arabic, Hebrew and Chinese. In this paper, we will discuss the case of Tifinagh, the writing script used for Amazigh languages. The case of Tifinagh is interesting and is increasingly attracting attentions especially with the recent evolutions. As of 2016, the Amazigh language is an official language in Morocco and a proposed one in the draft of the new Algerian constitution. Regardless of the common use of the Latin script and lack of knowledge of Tifinagh by native speakers, the standard Tifinagh IRCAM alphabet is used as the official script to write and teach the language in Morocco. This means that there will be a need to integrate Tifinagh alphabet in the current information systems in general and Internet domain names in particular. The integration of such a language script into the Internet domain name system faces several challenges. An administrative process, managed by ICANN, is needed in order to allow the use of Tifinagh script. Besides, some applications such as e-mail software or web browser restrict the characters which can be used as domain names. This paper will discuss the administrative process of enabling the use of Tifinagh alphabet in Internet domain names. We first present ICANN, the governing body of the Internet. Secondly, we discuss some useful technical details related to Tifinagh script. Thirdly, a description of the administrative process will be given. Finally, a set of challenges will be presented and discussed before giving some conclusions and recommendations for future work.

2. ICANN

The Internet Corporation for Assigned Names and Numbers, commonly known as ICANN, is a not-for-profit public-benefit corporation with participants from all over the world dedicated to keeping the Internet secure, stable and interoperable. It promotes competition and develops policy on the Internet's unique identifiers. Through its coordination role of the Internet's naming system, it does have an important impact on the expansion and evolution of the Internet.

2.1. Structure

ICANN has a multi-stakeholder structural model. It treats the public sector, private sector, and technical experts as peers. The ICANN community consists of various members such as registries, registrars, internet service providers (ISPs), commercial and business interest, non-commercial and non-profit interests, governments, and individual users (ICANN, 2016). ICANN has a very structured organization which slightly changes from time to time. As of 2016, it is managed by 16 members voting members of the Board of Directors in addition to four non-voting liaisons; six representatives of different Supporting Organizations, sub-groups that deal with specific sections of the policies under ICANN's purview; an At-Large seat filled by an At-Large Organization; and the President / CEO, appointed by the Board.
Currently, ICANN has the following three Supporting Organizations:

- **GNSO:** The Generic Names Supporting Organization which deals with policy making on generic top-level domains (gTLDs).
- **ccNSO:** The Country Code Names Supporting Organization which deals with policy making on country-code top-level domains (ccTLDs).
- **ASO**: The Address Supporting Organization which deals with policy making on IP addresses.

ICANN has also different advisory committees and mechanisms to receive recommendation on the needs of other parts that are directly linked to the three Supporting Organizations. These committees include (ICANN, 2016):

- **GAC:** Governmental Advisory Committee which is composed of representatives of a large number of national governments from all over the world.
- **ALAC:** At-Large Advisory Committee which is composed of individual Internet users from around the world.
- **RSSAC**: Root Server System Advisory Committee which provides advice on the operation of the Domain Name System (DNS) root server system.
- **SSAC:** Security and Stability Advisory Committee which is composed of Internet experts who study security issues pertaining to ICANN's mandate;
- **TLG:** Technical Liaison Group which is composed of representatives of other international technical organizations that focus, at least in part, on the Internet.

The following chart (Figure 2) summarizes the organizational structure of ICANN.

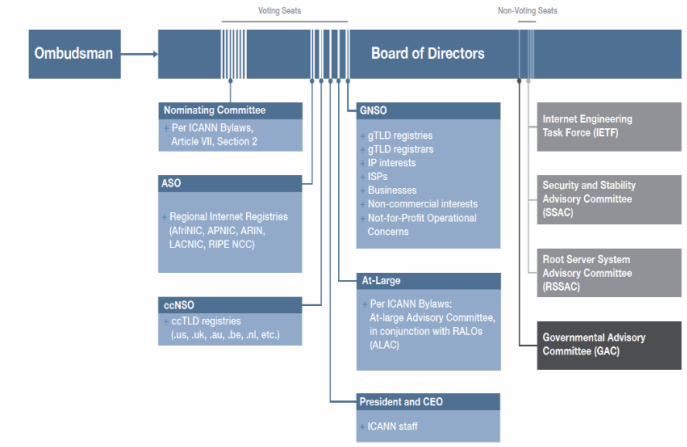

Figure 2- ICANN Organizational Structure (ICANN, 2016)

2.2. Activities

ICANN has several activities with regard to the regulation and governance of internet domains. One of its important activities is to run TLDs (Top-Level Domains) and deal with the assignment of IP addresses and ranges, ports, and other related attributes. This activity is done through an institution called IANA, Internet Assigned Numbers Authority. For instance, it is responsible for the transition from IPv4 to the new IPv6. Another important activity of ICANN is to preserve the security, stability and resiliency of the Domain Name System (DSN) and domain name registration services (ICANN, 2016). Another interesting activity of ICANN, which is important in the context of this paper, is the Internationalized Domain Names (IDNs) management. IDNs permit the global community to use a domain name in their native language or script. In this sense, ICANN focuses on the planning and implementation of the IDN Top-level Domains (TLDs) that include both ccTLDs and gTLDs. This is of a paramount importance to integrate language scripts such as Tifinagh. ICANN also organizes a one week-long conference three times per year, one of them is the organization's annual General Meeting, where new board directors take their appointed seats. These meetings are held in a different location each time.

3. Tifinagh Script

The Amazigh language is the oldest attested language in the Maghreb. Its area extends east to west, from the Egyptian-Libyan border to the Canary Islands, and north to south from the southern shore of the Mediterranean to Niger, Mali and Burkina Faso.

3.1. Tifinagh alphabet

Tifinagh is a traditional script system used to write the Amazigh language, with its various dialects. The modern Tifinagh, also known as Neo-Tifinagh is based on the traditional script of the Touareg people. It was introduced in the 20th century. A slightly modified version, called Tifinagh IRCAM is used in a number of publications in Morocco and particularly in elementary schools to teach the Amazigh language to children. Tifinagh is written from left to right. Figure 3 illustrates the Neo-Tifinagh alphabet that is used in Morocco.

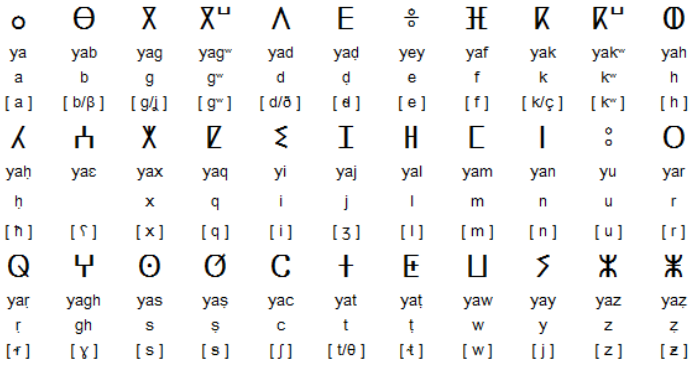

Figure 3- Neo-Tifinagh alphabet used in Morocco

3.2. Unicode

Tifinagh has currently a Unicode block in the computing industry standard for encoding and handling of texts in the world's writing systems. As of Unicode version 8.0, Tifinagh Unicode block ranges from 2D30 to 2D7F with 59 code points assigned and 21 unused reserved code points. IRCAM had proposed to the International Standardization Organization (21/06/2004) the Tifinagh alphabet that was adopted. This proposal includes four subsets of Tifinagh characters:

- IRCAM basic set;
- IRCAM extended set;
- Other neo-Tifinagh used letters;
- The modern Touareg letters whose use is attested.

The following table shows the official Unicode code chart of Tifinagh block (The Unicode Consortium, 2015):

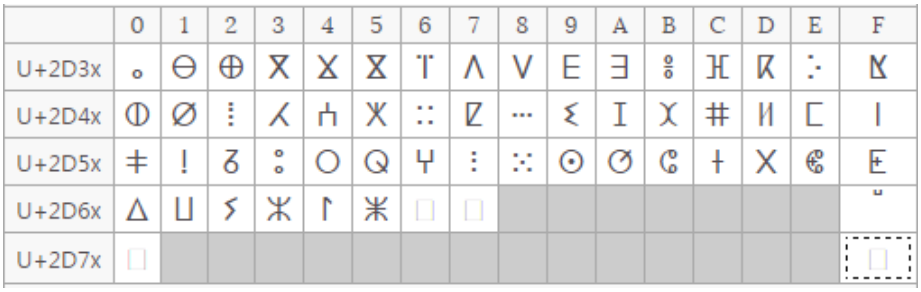

Figure 4- Tifinagh Unicode block (Unicode code version 8.0)

It is worth mentioning that Tifinagh Unicode covers the alphabets present in all the four subsets making it very flexible and easily extendable. This Unicode standardization was particularly important for the integration of Tifinagh in the new information systems such as the Internet.

3.3. Tifinagh presence in Internet

Until recently, the use of Tifinagh in the Internet was very limited. Nowadays, it is increasingly used and supported by different platforms. For instance, Microsoft started to support Tifinagh in their Operating Systems starting from Windows 7 when Microsoft web browser Internet Explorer started to support the characters. Another example is Google browser Chrome which also supports the alphabet as of 2016. Apple in its turn started to support Tifinagh starting from its mobile operating system (iOS 9).

Figure 5- Difference between a web browser supporting (right) and not supporting (left) Tifinagh Unicode

This support trend of Tifinagh in the Internet means that more and more people will be able to read and write using the alphabet on the web. Similarly, to Arabic and languages with special writing scripts, Tifinagh needs to be supported in domain name systems so that internet users can write their web addresses in Tifinagh. This has already been done with languages as shown in the following map:

Figure 6- Some examples of internationalized domain names, adapted from (Blackbird, 2016)

Given the fact that Tifinagh is integrated in the International Unicode Standard as well as the increasing support of this Unicode by different internet services, the next step towards the integration of Tifinagh in the Internet systems will be to include it in the Internationalized Domain Names. The following sections will discuss the different steps of the administrative process and related challenges.

4. Domain Names in Tifinagh

After having familiarized us with ICANN and seen Tifinagh technical specifications, in this part we will describe the process for the creation of Domain Names in Tifinagh script, as it is still a hanging issue. But, first of all, what is a Domain Name?

A Domain Name is a succession of characters (typically letters and numbers) separated by dots marking hierarchy. The domain name is the "address" of any page in the Internet. An example of a domain name is "event.ircam.ma". The "ma" is a Domain Name Label of the highest or top level, called Top Level Domain (TLD), "ircam" is a Domain Name Label of second level, and finally "event" is the Domain Name Label of third and lowest level. The "event.ircam.ma" can be translated into "the event that is hosted by IRCAM institute in Morocco".

As devices communicate and identify each other in a Network with IP addresses, a system was created to translate each Domain Name to its corresponding IP address. This system is called Domain Name System (DNS).

Initially, and as Internet was developed in countries adopting the Latin script, only this script in ASCII (American Standard Code for Information Interchange) format was allowed in Domain Names, and this was due to the origin of DNS which only supported Latin script in ASCII format. At that time, the rules for creating new top-level domain labels were simple:

- Labels must only be composed of letters (a-z) in the English alphabet
- Labels must have a length from 2 to 63 characters

Those rules were simple, but very restrictive. Even languages other than English using the Latin script, but not ASCII standard, were restricted. A domain name could not

contain the French “é”, Portuguese “ç”, Swedish “å”, German “ö” nor the Turkish “ğ”. This issue was of course bigger in countries not adopting the Latin script at all, and created various confusing situations. In Morocco for instance, the exclusivity of Latin script forced many to write Amazigh or Arabic domain names in the Latin script, instead of Tifinagh or Arabic alphabets, as it was not possible.
As countries not adopting Latin script are many in the world, and as the aim of Internet initially is to spread and democratize knowledge to everybody, the issue of internationalizing domain names was rising more and more. It was finally solved in 2010, when a technical solution was implemented to enable the introduction of domain names in multiple scripts and languages at the top level of the DNS without destabilizing the Internet (ICANN, 2015).

After the introduction of the Internationalized Domain Names (IDNs), Top Level Domains (TLDs) were divided into 4 categories:

- **Country code Top Level Domains (ccTLDs):** This category represents the official names of countries and territories. It includes domains in Latin script following the ASCII standard, such as .ma (Morocco) or .cat (Catalonia). The full list, set by ISO 3166, can be found here: https://www.iso.org/obp/ui/#search
- **Internationalized Domain Names of Country code Top Level Domains (IDNs ccTLDs):** This category is similar to ccTLDs, but gathers names of countries and territories in scripts other than Latin. For instance **.المغرب** in Arabic or **.中国** in Chinese.
- **Generic Top Level Domains (gTLDs):** Regards any other Top Level Domain that is not a country or territory. Example: .com .org .ngo.
- **Internationalized Domain Names of generic Top Level Domains (IDNs gTLDs):** Gathers gTLDs that are in other scripts, such as **.삼성** (.Samsung) or **.संगठन** (.Organization).

In Morocco, according to the 2011 constitution, the two official languages are Tamazight and Arabic (Moroccan Government, 2011). A solution for domain names has already been implemented for Arabic by ICANN. ICANN has indeed since 2010 (and until mid-2015) approved 47 IDNs ccTLDs that cover 15 scripts and 24 languages (ICANN, 2015). Deployment is still ongoing, until covering all scripts that ask for TLDs.

The same solution ought now to be implemented for the Amazigh language and the Tifinagh script. Our aim through this paper is to initiate the process so that in the very near future one can type the URL “□□□□□.ⴰⴰⵜⵜ.□□□”[1][2] as an alternative to “event.ircam.ma”, and that it works. In order to achieve this, the following steps need to be achieved in collaboration with ICANN:

---

1 ⴰⴰⵜⵜ is the abbreviation if IRCAM in Tamazight, and stands for ⴰⵙⵉⵏⴰⴳ ⴰⴳⵍⴷⴰⵏ ⵏ.

2 ⵜⵓⵙⵏⴰ ⵜⴰⵎⴰⵣⵉⵖⵜ. ⵎⵖⵔ is our suggested ccTLD for Morocco (in Tamazight) that will be discussed further in the next section.

- **Create a Tifinagh Script Generation Panel**

A Script Generation Panel is a panel supported by ICANN that is responsible for developing proposals that determine script-specific Label Generation Rules (LGR) for the root zone. The Amazigh speaking community, interested in developing Tifinagh script in TLDs, can engage through the Tifinagh Generation Panel. Registering in a panel is done by sending an email to idntlds@icann.org, and specifying the interest in the Amazigh language and Tifinagh script (APNIC, 2014).
This panel should contain experts in DNS, Unicode, IDNs, linguistics (Amazigh language), and domain name operations and policy. All challenges and milestones are discussed in the panel, and receive support from ICANN experts if needed. Issues can be various, and are related to the script, either regarding the technological aspect, or specific dialects, or confusion between characters, etc.
In the case of Tifinagh, and as Morocco is not the only Tamazight-speaking country in the world, the Tifinagh Script Generation Panel might also include participants from Algeria, Tunisia, Libya, Egypt, Mauritania, Canary Islands, Mali, Niger and Burkina Faso. Moreover, any person interested in Tifinagh script and Amazigh language is welcome to join the panel and participate with their time and knowledge in the project.

- **Establish the Tifinagh Code Point Table**

A Code Point Table is a subset of Unicode code points that is designed by members of the Script Generation Panel, and covers all characters that represent the new script that should be in the TLDs. The table gathering all code point tables is called the Maximal Starting Repertoire (MSR). ICANN has already released a first version of MSR called MSR-1. This repertoire covers nowadays 22 scripts (among which Arabic, Cyrillic, Greek, Thai…). Once the Tifinagh Panel establishes its Code Point Table and agrees on a definitive version, it can suggest the Label Generation Rules for Tifinagh. The proposal will then be raised to an ICANN experts Panel called the Integration Panel. Once accepted, the Integration Panel will integrate the LGR for Tifinagh Script into the main LGR (APNIC, 2014).

5. Challenges

In this section, we will cover some of the technical challenges that have to be solved, before, during and after sending the Tifinagh LGR proposal.

- **Linguistic issues**

As in any language, Tamazight or the Amazigh language contains a variety of dialects and differences, not only in Morocco, but also in the whole North Africa. If IRCAM has done an appreciated work of standardization of Tamazight in Morocco, this work has still to be done at a larger extend within the other Tamazight-speaking countries. The Tamhasheq dialect (spoken by Touareg) has for example characters that are specific to it in Tifinagh, and are not used in Morocco. The Tifinagh Script Generation Panel must be very cautious regarding this issue. Moreover, differences in dialects

create a challenge. All confusions between dialects must be cleared in order to create a working Unicode Code Points Table. One important aspect is that all characters in the table must be part of an active and contemporary variant of the Tamazight language.
Tifinagh is the script gathering all Tamazight speakers, and enabling it should give the opportunity to all Tamazight speaking communities to use it in their TLDs. This will need a tight cooperation between communities, and even sometimes compromises in order to reach a satisfying consensus for all parts.

- **Choice of names for TLDs**

When applying for Top Level Domain names in Tifinagh script, the Generation Panel should at least have standards for the most commonly used TLDs. We suggest in the table below an example of ccTLDs in Tifinagh for some Tamazight speaking countries and territories.

Country/Territory Name	Name in Tamazight	ccTLDs in Latin Script (ASCII)	Suggested ccTLDs in Tifinagh
Morocco	ⵍⵎⵖⵔⵉⴱ	.ma	.□□□
Algeria	ⴰⵣⴰⵢⵔ	.dz	.□□
Tunisia	ⵜⵓⵏⵉ	.tn	□□
Libya	ⵍⵉⴱⵢⴰ	.ly	.□□
Egypt	ⵎⵉⵚⵔ	.eg	.□□□
Mauritania	ⵎⵓⵔⵉⵜⴰⵏⵢⴰ	.mr	.□□□
Canary Islands	ⵜⵉⴳⴰⵣⵉⵔⵏ ⵜⵉⴽⴰⵏⴰⵔⵢⵉⵏ	.ic (proposed, not accepted yet)	. □□
Mali	ⵎⴰⵍⵉ	.ml	.□□
Niger	ⵏⵉⵊⵔ	.ne	.□□□
Burkina Faso	ⴱⵓⵔⴽⵉⵏⴰ ⴼⴰⵙⵓ	.bf	.□□□

This suggestion is obviously subject to modifications in the future, and is given just in order to show potential ccTLDs in Tifinagh. Other suggestions can be considered such as □□□ (from □□□□□□) for Morocco, or. □□□ for Canary Islands.

- **Register a whole Domain Name in Tifinagh**

As we have seen in the previous section, a domain name is a succession of characters separated by dots. Until now, we have only seen how to implement Tifinagh script in Top Level Domains (TLDs). However, a Domain Name is more than a TLD. It contains in fact second, third (etc.) level labels that should also be in the same script. How is this done?
If we suppose that the ccTLD. □□□ will be the official Tifinagh ccTLD for Morocco, then registering a domain name under it has to follow the same rules as when registering domain names in .ma and .المغرب domains, that are both official domain

names managed by the National Agency for Regulation of Telecommunication (ANRT). In order to register the domain dame □□□□□□. □□□, for example, it has to be available. This can be done through checking in www.registre.ma/whois . If the domain name is available, then a registrar should be contacted in order to apply for this domain name. This implies that Moroccan registrars must also be aware of the introduction of the Tifinagh Script, and make sure that they can support hosting domain names in Tifinagh.

- **Technical challenges**

Unfortunately, one of the challenges still facing Tifinagh Script is technical. Not all devices and Operative systems support yet this script, but there is a significant amelioration in the last years.

Difficulties observed in reading content in Tifinagh are still present in many interfaces, but are also improved gradually by the different updates each manufacturer does.

As our paper is discussing domain names in Tifinagh, our technical concern is not only about output (the webpage), but also about the input means. In order to write a domain name, a keyboard or an input system is needed. Very few keyboards sold in Morocco have a native Tifinagh script, and most of users interested in writing in Tifinagh have either to use virtual keyboards or install different programs and software tools. We hope that in the future more keyboards in Tifinagh will be released (in combination with Arabic/Latin letters), because it is the input system that often motivates more people to write in a given script. This will encourage the creation of more domain names in Tifinagh, which is hopefully our common aim.

6. Conclusions

Internet is imposing itself more and more as one of the most important and powerful tools of communication in our modern era. As a universal network, it should be accessible to all human beings regardless of languages they speak. In this matter, ICANN has been deploying efforts since 2010 in order to make domain names available in international scripts, and not only in the Latin script that was initially the only access tool to the Internet. Nowadays, people can write URLs in Arabic, Cyrillic, Chinese and many other scripts. The Amazigh language deserves clearly its place in the list. Our aim through this paper was to present guidelines and steps to follow in order to allow domain names available also in Tifinagh script. In order to achieve this symbolic milestone, it is important for us that a team of linguistic experts works together with us in order to overcome the different challenges that might face the project. Experience from other languages and support from ICANN are provided helping tools, completing them by moving forward to panel work is the next step towards having internationalized domain names in Tifinagh.

References

APNIC. (2014, 09 30). *Speak up for your language*. Retrieved from APNIC - Asia Pacific Network Information Centre: http://blog.apnic.net/2014/09/30/speak-up-for-your-language

Blackbird. (2016, 05 24). *Internationalized domain names (IDN) and DNS / Bind9 problem*. Retrieved from Brainstorming of a sysadmin: http://blackbird.si/internationalized-domain-names-idn-and-dns-bind-9

IANA. (2007, 05 14). *Character Sets.* Retrieved from Internet Assigned Numbers Authority - IANA: http://www.iana.org/assignments/character-sets/character-sets.xhtml

ICANN. (2015, 09 08). *Linguistic Diversity in the Internet Root: The Case of the Arabic Script and Jawi*. Retrieved from ICANN - The Internet Corporation for Assigned Names and Numbers : https://www.icann.org/news/blog/linguistic-diversity-in-the-internet-root-the-case-of-the-arabic-script-and-jawi

ICANN. (2016, 05 20). *Resources - ICANN*. Retrieved from The Internet Corporation for Assigned Names and Numbers - ICANN: https://www.icann.org/resources

IRCAM. (2016, 05 25). *Présentation*. Retrieved from The Royal Institute of the Amazigh Culture - IRCAM: http://www.ircam.ma/?q=fr/node/620

ITU. (2015). *World Telecommunication/ICT Indicators database 2015.* International Telecommunication Union - ITU.

Moroccan Government. (2011, 06 29). *Constitution.* Retrieved from Maroc.ma: http://www.maroc.ma/ar/system/files/documents_page/BO_5964Bis_Ar.pdf

Stewart, W. (2000). *Living Internet.* Creative Commons.

The Unicode Consortium. (2015). *The Unicode Standard, Version 8.0.* Retrieved from Unicode: http://www.unicode.org/charts/PDF/U2D30.pdf