

Agnès Vatican – Archives départementales, directrice

... and is now called Aquitaine Cultures Connectées, right? It's about heritage. You are going to do a two or three voice presentation, I don't know. So Nathalie Leuret...

Nathalie Leuret :

It's going to be a three voice presentation because I wanted to do a quick summary in the program's introduction, because the website is linked to the program. But I'll be brief. And then I want to present the website in two parts, and the two specialists will each present a part.

Agnès Vatican :

In any case thank you so much for coming, because I'm sure you're not often available. We thought it would be interesting to show the relevance of a new portal that integrates collections and resources gathered through the years by the partners to this project.

Nathalie Leuret :



This is the Aquitaine Cultures Connectées program that you may have known as BNSA, which has existed for 16 years. We started this program in 2000 with the DRAC Aquitaine (Regional Directorate of

Cultural Affairs in Aquitaine). The joint goal was to support the development of web in the cultural sector. It is a partnership-based program, because beyond our mutual commitment, this project became part of the different regional program contracts in place since 2000. This is the third program contract and even though it isn't a cultural project, there is a cultural component. In the digital section we identified the Aquitaine Cultures Connectées program with specific ringfenced budget line (CPER). Lasting till 2020, we have a budget of €285,000 per year and per partner.

It is also supported by our local partners, the program's members. A dozen supporters signed a membership charter that I'll show you later. We put all our partners' logos on it, as well as on other Aquitaine operatives who support the operations that will be available on our website. So, essentially digitalisation.

Today we are thinking about Aquitaine Cultures Connectées' scope and that's why we changed the name. It is already larger than the BNSA, which was for mediation and the digital enhancement of heritage collections. Today, Aquitaine Cultures Connectées deals with every aspect of culture: books, art, cinema, live theatre. (...) In any case we support any digital innovation in the cultural sector, even beyond the project's main aspect: heritage. Here are our partners. A dozen of them are local organisations. We aim to expand. Behind these local organisations there are their archive services, museums, libraries.

They all sign a charter, which was recently modified, so they will have to re-sign it. Moreover, since it only lasted till the end of the program contract, it has now expired, so we'll have to renew it. The Aquitaine Culture Connectées program has two different sections. The Collection section, which is probably more interesting for you this morning, is the constitution of collections online. We only support member partners, who can ask for support from the digitalisation phase to the metadata's online release onto our website.

For the last dozen or so years, we have been looking at all of the requests. Today we are looking into having an editorial line, with maybe a committee of experts to help us establish topics and to digitalise those that are more important to the region. This is an Aquitaine program and as part of an interregional contract it is limited Aquitaine's borders. One of the challenges with the merger is to expand it to the other two regions, which are one, to the other region's territories which will allow us to open up our network of member partners.

The second section is the Factory, an Aquitaine initiative. From 2010-2011, we wanted to diversify our documentary and editorial offering. We also wanted to communicate with a wider audience. So we put in place a program, a measure, which turned into 3 calls for projects, in 2011, 2013 and 2015. These calls were devised to support the most innovative creative contents, notably by supporting the development of services with digital applications.

All of the cultural sector was included, not just heritage. The idea was to expand. Of course the member partners were a priority, and can apply in the same way as the others. But we wanted to open up

to other audiences. Private operators and companies applied. From the last calls for projects, 16 digital productions were supported. They were web documentaries for example, I don't know if you were able to view the one we made about the river. Our member partners can obviously provide files.

We are thinking about how to open the program (...), from a management viewpoint. We want to make it into a long-lasting program on our website that will allow all those with a project to submit their files, either in the Collections or the Factory section. And I forgot to say, I can see that Jean-François is in the room, as well as the interregional contract, we signed a memorandum of understanding with the state, which includes these two segments.

Now, let's present the website.

Nadège Galibert :



The Aquitaine Cultures Connectées website echoes the program Nathalie presented, for us and for the region it is an echo of its intervention in cultural digital innovation, and the support we give to the documentary databases, the editorial section, and to the most innovative productions. I'm going to show you the 4 parts of the website, but first I will tell you who it is for. It's addressed to an Aquitaine and a non-Aquitaine audience. First of all, the audience that plays an active part in the digital cultural world: culture, digital and communication professionals, tourists and those interested in the region, as we deal with Aquitaine heritage, and also schools and young people. We have resources specifically for younger

people. The website is managed by the Direction de la Culture de la Région (the Regional Culture Directorate) and we work alongside Bertrand Hubert. He works a lot more on the digital aspect, so he will present the documentary databases and the OAI system later on. I deal with the editorial aspect. So I am going to present the 4 sections of the website, which we can see on the home page. I'll scroll down to show you the whole page. You'll be able to visit the website from home and browse more easily. The website is divided into 4 sections. First there is "Digital Exploration", which is like a map. It presents the creative digital productions dealing with Aquitaine culture. You can see them here on the map. There are more than 150 of them, dealing with various cultural topics: architecture, fine arts, nature, landscapes, all developed in original digital formats. They can be found online. They may be web documentaries, interactive games, or mobile apps. Let's go onto one of the pages. It's a resource indication. This is a project we helped with in the Landes, since we also present assisted projects and Aquitaine projects that had no help from the region. This is an assisted project which is part of the Factory, it's a transmedia project concerning music. This is what it looks like. Every time there is a visual, a geolocation, information on the project and a link to the resource. Now, we're going onto the Aquitaine Cultures Connectées website and the digital creation. I'm going back to show you. We can see the geolocation and a list of resources that allow us to search amongst the digital creations. Next, the second section, Digitalised Collections and the documentary databases, Bertrand will talk about this later. The next part is called Innovate Initiatives, where we promote places, creations and the contributors to cultural innovation in Aquitaine.

It was a choice made by the Culture Directorate, a fairly arbitrary choice of remarkable regional experiences or projects in terms of digital innovation. There are 4 categories: places of experience, creative actors, innovative approaches and online digital creations. Let me show you one. The format's a lot longer than that of Digital Exploration. There's a presentation of the project, often with feedback, that can be used by other potential partners who create digital cultural projects. There is also geolocation, which is taking time to upload, but that normally works. There are images, videos, and any other type of media on the topic, with links to learn more. The last section, Creative Box, is a bit different from the others, in that they aren't cultural projects. This is where cultural and digital resources are available, in two sections. There are links towards monitoring sites, or resource tools used to retrieve digital cultural resources and information on the support and financing of a digital cultural project, offered by Aquitaine and France, and even by Europe, with various calls for projects or other measures.

To finish with the website, it also offers a newsletter that I urge you to subscribe to. It comes out monthly and recaps all the news that we can publish, and monitoring tools we retrieved elsewhere. We are also present on social networks, mainly Facebook and Twitter, with a targeted community that follows us on these subjects. Now I'll let Bertrand speak.

Bertrand Hubert :



Hello everybody. We are going to look back at the Digital Collections section. Let's start with BNSA's history with a service provider, using OAI-PMH metadata. There have been two main stages and we are perhaps currently in a third. We started with an Aquitaine Patrimoine (Aquitaine Heritage) format, which some of you will know, based on Dublin Core and enriched. Why? Because we were on an internet portal where we had to be able to promote data that didn't exist in Dublin Core – images, geolocation – and more importantly be able to gather various types of data, from archives, books, museums, so a fair bit of metadata, and the Aquitaine Patrimoine model created at that time, was supposed to combine all of this.

We are not comparing ourselves to Europeana or to the BNF, but we draw inspiration from them. We were challenged by some technical difficulties, mainly the partners – I heard the feedback and questions you had earlier, with human and technical problems, the ability to internalise technical skills, even though you rely on service providers to create your websites or repositories. There were however things that weren't necessarily accessible or simple. There were always some manual adjustments to carry out to allow the data to be promoted on the portal. We are not going to go over all the problems we had, but putting the images up was very difficult to do with the AP format. Why? Because there were always images in formats that didn't allow us to retrieve them. There were some very heavy formats. You can imagine how long it took us to process everything when we receive 300,000 images weighing 3 mega per image in a repository that hosts 400,000 data items. It was quite unmanageable.

We also had problems with geolocation, of which I heard some of the feedback. Principally concerning the norm. We had all types of nuances of the norm: Lambert III, Lambert 93, and other Google formats, for the slightly limited existing browser to be able to put these notices up effectively. This new portal raised a few questions, mainly concerning the reworking of an aggregator that could adapt to all types of data. The first thing we did was open to Dublin Core as well as Aquitaine Patrimoine. This allowed our partners who couldn't provide Aquitaine Patrimoine, to retrieve data labelled "Aquitaine" in a Dublin Core format. We had to work on an aggregator capable of retrieving these two models, and at the same time have a display on the portal capable of displaying these two types of data with the same results. So that was one of the first tasks.

The second task was to work on a search engine capable of promoting all this data – at the time we had between 500,000 and 700,000 data items depending on the repositories we had at any given time – and to find a technical solution. We started with Solr. Today it raises questions because with Big Data and suchlike, we could maybe find other solutions that allow structured search with a system of facets, as presented by the previous speaker. So we worked on these sections. Even though they now work there is still a problem with the quality of the original data, to know whether the metadata was adequately entered. In the aggregator, which I can't show you as it is an internal tool but will be accessible to the partners in a future update later in the quarter, we worked on a metadata quality evaluation algorithm. What does this mean? It means that the partner will be able to evaluate whether their repository provides all the data. It gives a score. And better, the partners score themselves. The algorithm makes this easier. If the fields are not filled in for the entire repository, it's easier to see whether or not we can get an effective dissemination in France. As is the case with Europeana and other websites, after this we can no longer ask for a good dissemination of this metadata if the fields aren't filled in. The aggregator will be open to the partners. We will be able to estimate the quality of the metadata input. This section will be very interesting.

After seeing the difficulties the partners had providing a structured and enhanced AP model and the resulting infrastructure network, this year we accepted the Dublin Core format. This doesn't mean we will keep the standard Dublin Core, we may switch back if the models aren't homogenised, because we also take the BNF's usage into account. Everyone has a model. In general terms, a partner will have to be capable of providing as many repositories and models as they want to disseminate on these portals. We are small compared to the others. We went back to the simple Dublin Core format that will allow us to gain as many partners as possible. We placed the emphasis on referencing, which Europeana and the BNF haven't focused on, because we are addressing the general public as well as researchers.

For this we need to be indexed on the users' search engines, so we must work on an XML feed, and the metadata sent to Google's robots and others', and to then get more visits from the general public, as getting them on our website is difficult. To attract them through a search engine that is nothing more than a browser, with Google or suchlike, may be easier.

So we are going to work on the indexation, which has worked quite well so far. Besides the statistics show that a lot of the web users arrive mostly via Google or other search engines rather than via the website.

We have to update all of our technical documentation because it is constantly evolving. Our job is to listen to the partners. Even though we test their repositories, we can't always intervene as the technical structures are all different. We can identify their problems and support them in their retrieval. Updating the aggregator could also allow us to be an intermediary, and to be re-harvested by bigger organisations, something that hasn't been done yet.

Europeana and the BNF must also settle because this sort of thing regularly evolves. When I said earlier that we draw inspiration from the BNF, I meant that our regulator will allow our partners direct access to a partnering back-office, where they themselves will be able to declare their data records, their repositories, evaluate them, and harvest and integrate the ACC portal, without our validation. We are going to do this in two steps. First, we have to see how difficult the repositories are before we can automate this. And if the partners are mature and experienced enough to inject the data into the portal themselves, then...

Bertrand Hubert :

We will make the insertion even more automatic. We can also discuss the other problems with geolocation that you seem to be interested in.

Any questions?

Agnès Vatican :

Thank you.

It's interesting to see how a long term project can mature enough to raise questions that weren't among those at the beginning. Concerning the formats, with our usage in mind, there is software that allows us to produce search devices in numerous formats. I believe there are new formats underway, including one on Wikipedia? But Georges isn't here, so we'll have to come back to that.

So for us this is an issue. When we have an editor that allows multi-publications, it's fine, but it's not necessary viable long term, so we use a Dublin Core format. In the LoCloud experiment the enrichment tools were interesting. Compared to a Dublin Core format and leaving aside the referencing strategies, did you also raise questions about enriching Dublin Core?

Bertrand Hubert :

Yes, we did. The problem is that if you take a step back from our projects, you will realise that Europeana proposes its own enrichment service, another one its own too, and so on. Whereas the Archives Départementale is a big partner, there are also smaller projects, with small repositories and lacking the human resources, who also want to know where to get enrichment. Someone said earlier "it's the source that I'm interested in". For me the enrichment should really be done at the data's source and not necessarily with Europeana or the BNF etc. just because it's good. The problem is that to be able to open up and be operable with everyone, the enrichment needs to be done at the source. And that is hard to manage because the partners don't always have this internal technical ability.

Agnès Vatican :

Any questions?

Louis Colombani :

I have a further question relating to the formats. In a short term utopia, can we hope to see universal formats? You told us a lot about the RDF models etc. And the second question, you mentioned abandoning Solr so...

Bertrand Hubert :

At the moment we are using the Solr search engine and we are satisfied with it given the amount of data we have. If at some point we had to increase, if in a perfect world we had perfect data and a lot of data, then yes, we could go much further with others, such as Elasticsearch. This is a tried and tested solution. We are going to test it on other datasets, although not on this program, but we will keep it in mind for the system's development. The RDF is already present. In the Aquitaine Cultures Connectées solution, again the problem is to know how to reference the datasets to get a good RDF feed. We put the RDF in place straight away, but not necessarily for the partners even though they can provide them. But if we had to use rebounds, at that point the RDF will be exploited and exploitable by those who want us to retrieve it.

Myriam Pauillac :

So your pages are in RDF?

Bertrand Hubert :

Yes, the data structure behind them is RDF.

Myriam Pauillac :

OK.

Agnès Vatican :

Any more questions?

Nathalie Gascoin :

I don't really understand your connection to Europeana?

Bertrand Hubert :

None.

We draw technical inspiration from what Europeana does, we retrieve data. I have spoken to them about their old models that were extremely rich. We compare solutions but we don't have a program links with them

Nathalie Gascoin :

So you don't envision putting Aquitaine Cultures Connectées' data onto Europeana?

Bertrand Hubert :

Sending the data, yes. But the problem will be managing all the duplicates. What you have to know is that Europeana and the BNF keep only 10% of the data, apart from when they deposit their own datasets. It's an interesting number to get from them. This means that if we send them our repository and if there aren't any duplicates, they will still select a certain amount. For whatever reason – editorial or due to the quality of the data - they don't take 100% of the data. There is a lot of noise about data that wasn't properly qualified. And they identified this problem straight away, so they have the human resources to filter all this, so we can't compare ourselves to them.

James Lemaire :

I didn't understand whether Europeana retrieves all the data that we give to them?

Bertrand Hubert :

Yes, they retrieve it all, but then they filter it, and the BNF does too. They retrieve everything. If the data isn't qualified enough, they filter it beforehand.

[Speakers 1 and 6 talk at the same time]

Agnès Vatican :

I understand data and quality of data, but the increase in the number of portals and the possibilities of displaying the resources, the archives, also raise the question of a strategy we could use. There are global strategies that could allow us to contribute or not to the portal. At the moment we haven't thought about it like this. In LoCloud's case, we put everything we could onto Europeana. Everything might not be relevant, given the way we use Europeana. I don't know if others have considered these questions, but it really makes me wonder. And I tell myself that the leaders and the aggregators do have strategies relating to the type of data and the way they are displayed: referencing and structuring. Europeana has just redesigned its interface with Europeana Collections. Do we, as archivists, also have strategies relating to the ways we want to display our resources that depend on the portal's own strategies?

Louis Colombani :

I have a provocative question. With all these portals, do the archives really need their own website?

Agnès Vatican :

Good question. In any case, I think that the archives' new websites can't be designed in the same way they were 10 years ago, when we believed we were alone in the world and thought that we had to put everything forward even if it meant nothing was visible, because no one else would. We have to consider the redesign of our website in the context of this proliferation. And this is why I use the term strategy. But this doesn't mean that the archives won't have a portal anymore. For example, concerning the Archives Départementales library, we are in contact with our colleagues from the University and the Agence de l'ABES - Agence Bibliographique de l'Enseignement Supérieur (the Bibliographic Agency for Higher Education) to see whether or not this library should be integrated into the SUDOC (University documentation system) and bought back to us through specific tools that will allow us to consult from the archives website, or in any case where we won't be the main reservoir of these resources. That is one project. We are conducting a feasibility study with Catherine Wilmet and Michel Constantin at the moment, but this is a project we have for part of the library, given the way the big reservoirs and library catalogues are structured these days and how they interconnect, being part of the SUDOC means you can be directly consulted on the SUDOC, or in Bordeaux on Babord, the portal for the university libraries in Bordeaux, or on the portal for the libraries of Bordeaux which has a federated query with Babord. It allows us to be indirectly visible, to make our modest archive library visible in numerous ways.

Concerning the archives, we believe we still have full legitimacy.

In any case, thank you very much.

