

# Stage ARVIVA – Visualisation de données, graphes et sémantique

Niveau : L3, M1 ou M2 (les objectifs seront adaptés selon le cas)

## Contexte

Le projet ARVIVA, initié en 2014, a pour objectif d'étudier l'activité artistique en Val de Loire entre Moyen Age et Renaissance (de 1470 à 1600<sup>1</sup>) et de créer les outils informatiques pour collecter les données, assister leur analyse ainsi que permettre leur valorisation. Il réunit pour un travail collaboratif des spécialistes de différents domaines : historiens, historiens de l'art, informaticiens. L'équipe ARVIVA est ainsi constituée de chercheurs du Centre d'Etudes Supérieur de la Renaissance (**CESR**), du Laboratoire de recherche en Informatique de L'université François-Rabelais Tours (**LI**), de l'Institut de Recherche en Histoire des Textes (**IRHT**) et d'ingénieurs d'études, spécialistes en histoire, histoire de l'art et informatique, certains embauchés pour le projet et certains relevant de la Maison des Sciences de l'Homme Val-de-Loire (**MSH VdL**). Les étudiants en Master du CESR, stagiaires du projet, aident à la recherche et à la saisie dans la base.

La collaboration entre historiens et historiens de l'art permet de croiser les recherches sur les œuvres avec les dépouillements d'archives pour trouver des informations sur les commandes artistiques et sur les œuvres non-conservées. Les témoignages d'archives permettent aussi d'étudier les acteurs de la production artistique et leurs réseaux, de mettre en évidence leurs liens professionnels et familiaux dans une perspective d'histoire sociale.

Cette collaboration s'est d'abord exercée dans la conception d'un outil de collecte des données, conçu en étroite concertation entre historiens, historiens de l'art et informaticiens. La **base de données relationnelle** qui est au coeur de cet outil accueille tous les liens précédemment évoqués. Elle est aussi le support d'un travail collaboratif au sein de l'équipe ARVIVA. Elle est également utilisée pour la **recherche d'information et l'analyse des données**. L'ensemble de ces services est destiné à être regroupés dans un **portail web**, qui permettra de rendre les œuvres et les archives plus accessibles aux chercheurs, de pousser leur réflexion sur le foyer artistique en Val de Loire et de valoriser leurs résultats de recherche à travers différents modes de présentation adaptés à tous types de publics : spécialistes-chercheurs, étudiants en histoire et histoire de l'art et grand public.

Actuellement la base est implémentée et alimentée via l'outil de collecte. Elle contient déjà plusieurs centaines d'œuvres, de personnes et d'extraits d'archives, plus les liens entre ces différents éléments. Le coeur du portail web existe, il comprend un outil pour rechercher des informations et naviguer dans les données. La MSH VdL assure l'hébergement et la maintenance de l'ensemble.

## Objectifs

L'objectif de ce stage est de construire d'autres outils de visualisation à intégrer dans le portail pour mieux valoriser les contenus de la base et mieux soutenir les analyses interactives menées par les chercheurs. Les aspects IHM sont évidemment primordiaux.

En cas de stagiaire de Master 2, un objectif supplémentaire sera la conception de l'ontologie correspondante à cette base (en testant des outils de bootstrapping), avec un alignement sur CIDOC-CRM (Doerr, 2007).

## Méthodologie

Une première contribution de ce stage concerne la possibilité de visualiser les données géolocalisées sous la forme d'une carte, qui puisse être un point d'entrée à la navigation dans les informations de la base. Pour cela on pourra utiliser un visualiseur de web cartographie comme par exemple Leaflet (Leaflet, 2017). On étudiera par ailleurs la possibilité d'avoir des cartes historiques avec le géoportail de l'IGN (IGN, 2017).

Une deuxième contribution de ce stage consiste à développer un mode de visualisation du réseaux de relations stocké dans la base. Cela peut passer par un export vers une base de données graphes (Angles, Gutierrez, 2008), comme par exemple Neo4J (Neo4J, 2017). Un outil a été d'ailleurs développé à la MSH VdL pour faire de la visualisation de graphes, et la base de données du projet ARVIVA a été choisie pour les tests de cet outil du fait de la forte connectivité de son schéma conceptuel. Peut-être que cet outil est directement intégrable dans le

---

<sup>1</sup> Les limites chronologiques ont été repoussés jusqu'à la fin du XVIe siècle pour pouvoir suivre la mutation du style artistique dans la région à la Renaissance et contester l'idée de rupture de la tradition artistique dans le Val de Loire dans des années 1530. Même si la cour royale quitte la région, l'activité artistique continue et les liens ne se rompent pas.

portail ARVIVA mais encore faut-il s'assurer de son utilisabilité par l'équipe ARVIVA, question qui n'a pas été considérée pour son développement (puisque ce n'est pas pour ARVIVA qu'il a été conçu) : c'est donc un cas d'étude IHM intéressant.

D'autres apports seront attendus en fonction de la durée du stage : un tableau de bord des statistiques de la base, calculé à la demande, serait également utile à l'équipe ARVIVA, de même qu'une "documentation" d'utilisation du portail (y compris en mode administration), qui peut prendre la forme de courts tutoriels placés dans le portail lui-même (mode flash), serait utile aux visiteurs du portail.

Par ailleurs, si le stagiaire est en M2 alors une représentation de la sémantique des données de la base sera également demandée. L'ontologie représentant la base et son contenu sera générée au départ à partir de la base elle-même en utilisant des outils de bootstrapping (Jiménez-Ruiz, 2015), puis validée ensuite par interview d'experts. Dans le but de préparer la suite du projet ARVIVA, qui consistera à intégrer cette ressource dans le web sémantique, un alignement sur CIDOC-CRM sera effectué. Là encore, on utilisera d'abord des outils d'alignement puis on complètera et validera le résultat manuellement.

### **Compétences attendues**

Le/La candidat(e) est étudiant(e) en 3ème année de Licence d'Informatique, ou en 1ère ou 2ème année de Master en Informatique ou en dernière année d'Ecole d'Ingénieur.

Le/La candidat(e) devra avoir une bonne connaissance des bases de données relationnelles, de la programmation Java, et des principes de l'IHM. Techniquement, les systèmes utilisés pour la base et le portail vont de MySql au framework Symfony (design pattern MVC), en passant par PHP et Javascript.

Le reste des compétences utiles dépend de l'année d'étude et de la durée du stage.

### **Encadrement**

- Béatrice Markhoff, Maître de conférences, Université de Tours (beatrice.markhoff@univ-tours.fr)
- Jean-Yves Antoine, Professeur, Université de Tours (jean-yves.antoine@univ-tours.fr)
- Thomas Devogele, Professeur, Université de Tours (thomas.devogele@univ-tours.fr)

### **Contacts**

Les candidatures sont à envoyer par mail avec comme sujet [Stage ARVIVA] à beatrice.markhoff@univ-tours.fr

Le mail doit comporter :

- Une lettre de motivation
- Un CV détaillé
- Les notes obtenues dans l'année en cours et l'année précédente
- Des références ou lettres de recommandations

### **Déroulement du stage**

Le stage se déroulera dans le laboratoire LI, antenne de Blois, et dans l'équipe BdTIn. Le/la candidate s'intégrera ainsi dans une équipe de recherche qui collabore régulièrement avec des chercheurs en sciences humaines en appliquant ses travaux de recherche en science des données aux problématiques rencontrées en SHS. Il/Elle pourra ainsi échanger avec des spécialistes et participer à des séminaires. Des réunions très régulières auront lieu avec l'équipe ARVIVA et avec l'équipe informatique de la MSH VdL.

Financements : 554€/mois

### **Bibliographie**

- Angles, Renzo; Gutierrez, Claudio (2008). "Survey of graph database models". *ACM Computing Surveys*. Association for Computing Machinery. 40 (1). doi:10.1145/1322432.1322433
- Neo4j (2017) : [https://neo4j.com/developer/graph-db-vs-rdbms/#\\_from\\_relational\\_to\\_graph\\_databases](https://neo4j.com/developer/graph-db-vs-rdbms/#_from_relational_to_graph_databases)
- Jiménez-Ruiz, Ernesto; Kharlamov, Evgeny; Zheleznyakov, Dmitriy; Horrocks, Ian; Pinkel, Christoph; Skjæveland, Martin G; Thorstensen, Evgenij & Mora, Jose (2015). "BOOTOX: Bootstrapping OWL 2 ontologies and R2RML mappings from relational databases". *CEUR Workshop Proceedings*. ISSN 1613-0073.
- Martin Doerr, Christian-Emil Ore, and Stephen Stead (2007). "The CIDOC Conceptual Reference Model: a new standard for knowledge sharing". In *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling*-Volume 83. Australian Computer Society, Inc., 51–56.
- Leaflet (2017) : <http://leafletjs.com/>
- IGN (2017) : [http://remonterletemps.ign.fr/telecharger?x=0.863380&y=47.457909&z=9&layer=GEOGRAPHICALGRIDSYSTEMS.MAPS&demat=DEMAT.CASSINI\\$GEOPORTAIL.DEMAT;MAPS](http://remonterletemps.ign.fr/telecharger?x=0.863380&y=47.457909&z=9&layer=GEOGRAPHICALGRIDSYSTEMS.MAPS&demat=DEMAT.CASSINI$GEOPORTAIL.DEMAT;MAPS)