

Les Archives nationales font leur hackathon

PRÉSENTATION AUX ÉTUDIANTS - 9 OCTOBRE 2018

DMOASI des Archives nationales

Frédéric ZAMARRENO

Chef du département de la maîtrise d'ouvrage

Gaetano PIRAINO

Chef de projet métier et innovation du SI

Les Archives nationales

Créées pendant la Révolution française, les Archives nationales conservent les archives publiques de l'État depuis le VIIe siècle, les minutes des notaires parisiens et de nombreux fonds d'archives privées de particuliers ou d'associations. Les Archives nationales permettent l'accès gratuit de tous les citoyens aux archives publiques et offrent ainsi la garantie de la transparence de l'État et de la démocratie.

Ce sont près de 350 kilomètres linéaires de fonds et 50 téraoctets d'archives numériques, qui constamment s'enrichissent, soit des millions de documents précieux tant pour leur contribution à la connaissance historique et à la mémoire individuelle et collective que pour leur intérêt patrimonial unique et inestimable.

Le département de la maîtrise d'ouvrage du SI...

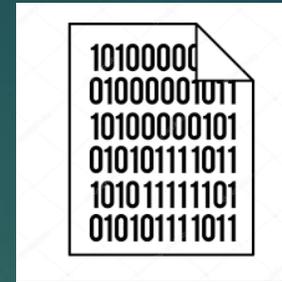
Ses missions

- La maîtrise d'ouvrage SI des projets et applications métiers.
- Pilotage de l'évolution des outils de gestion archivistique et documentaire (SIA/SIV)
- Pilotage des évolutions du SIA vs ADAMANT,
- Pilotage de projets documentaires (migrations des bases externes vers le SIA, numérisation et encodage des IR, ...)
- Appui technique et scientifique auprès des départements utilisateurs,
- Accompagnement des utilisateurs sur les usages et les bonnes pratiques (notamment en matière d'informatique documentaire)

Un enjeu majeur : Accompagner la transition numérique des Archives nationales avec le développement de l'offre numérique vers les partenaires ministériels et vers les publics.

...et un
hackathon,
c'est quoi ?

- **Des données**, mises à disposition par des institutions culturelles, librement réutilisables (archives numérisées, images ; bases de données, inventaires) ;
- **Des gens**, de tous horizons, qui viennent avec leurs compétences, leurs idées, leur bonne humeur, et travaillent en petit groupe pendant un temps donné sur un projet de réutilisation de ces données ;
- **Un lieu**, pour accueillir les jeux données et les hackathoniens dans un espace convivial avec tables de travail, prises de courant et wifi, et des espaces annexes pour alimenter les troupes en boissons et nourriture.



[Rex BNF 2016](#)

Gallicarte 2016

En savoir plus

The screenshot shows the Gallicarte interface with a map of Paris. Three orange circles are overlaid on the map, indicating search results counts for different areas: 185 in the north, 387 in the west, and 442 in the east. The interface includes the BnF logo, search bar, and navigation elements.

The screenshot shows the Gallicarte interface with search results for 'Bibliothèque Atget'. The interface includes the BnF logo, search bar, and navigation elements. A dropdown menu is open, highlighting the 'carte' option. The search results list includes details such as 'Type de document', 'Documents consultables en ligne (5 162)', and 'AFFINER' options.

Gallicarte : géolocaliser les résultats de recherche de Gallica

Hackathon Open Cultural Data (Suisse)

[En savoir plus](#)



Reconnaissance faciale et annotation collaborative
(collection de plusieurs milliers de photographies des assemblées dans les Archives de la SDN alias ONU)

Fonds photo mis en ligne mais sans métadonnées [en savoir plus](#)

Le Hackathon des AN

THEMATIQUE : archives et citoyenneté

Objectifs du hackathon des AN

Les Archives nationales attendent de cet événement des projets de développement de solutions innovantes pour :

- identifier des services numériques correspondant à de nouveaux usages pour les publics ;
- améliorer l'exploitabilité des données pour les rendre interopérables (dans la perspective d'alignement de données avec d'autres institutions) ;
- faciliter l'accès aux ressources (amélioration de l'expérience utilisateur, data-visualisation, géolocalisation, etc.).

Jeux de données sélectionnés

1. Base Léonore
2. Instruments de recherches référençant des actes des 6 notaires de paris (XVIIIe – début XXe)
3. Discours et reportages photographiques de De Gaulle et Rocard
4. Fonds de la restauration de terrain de montagne
5. Actes de naturalisation
6. Export de tous les IR publiés dans la SIV
7. Référentiels en xml eac-cpf et xml
8. Registres des entrées

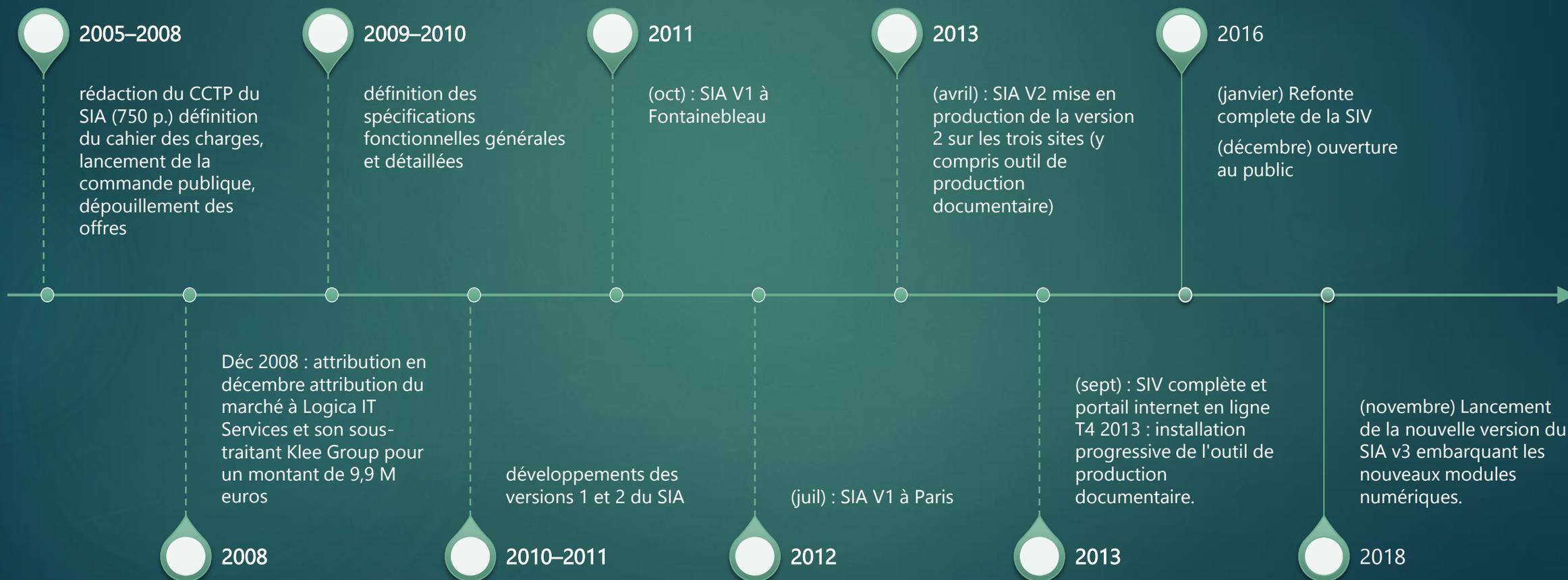
Formats de données à exploiter

- ▶ XML – EAD
- ▶ Extraction en CSV
- ▶ Images en jpg
- ▶ Métadonnées Dublin Core pour le jeu de données numéro 6 (OAI)

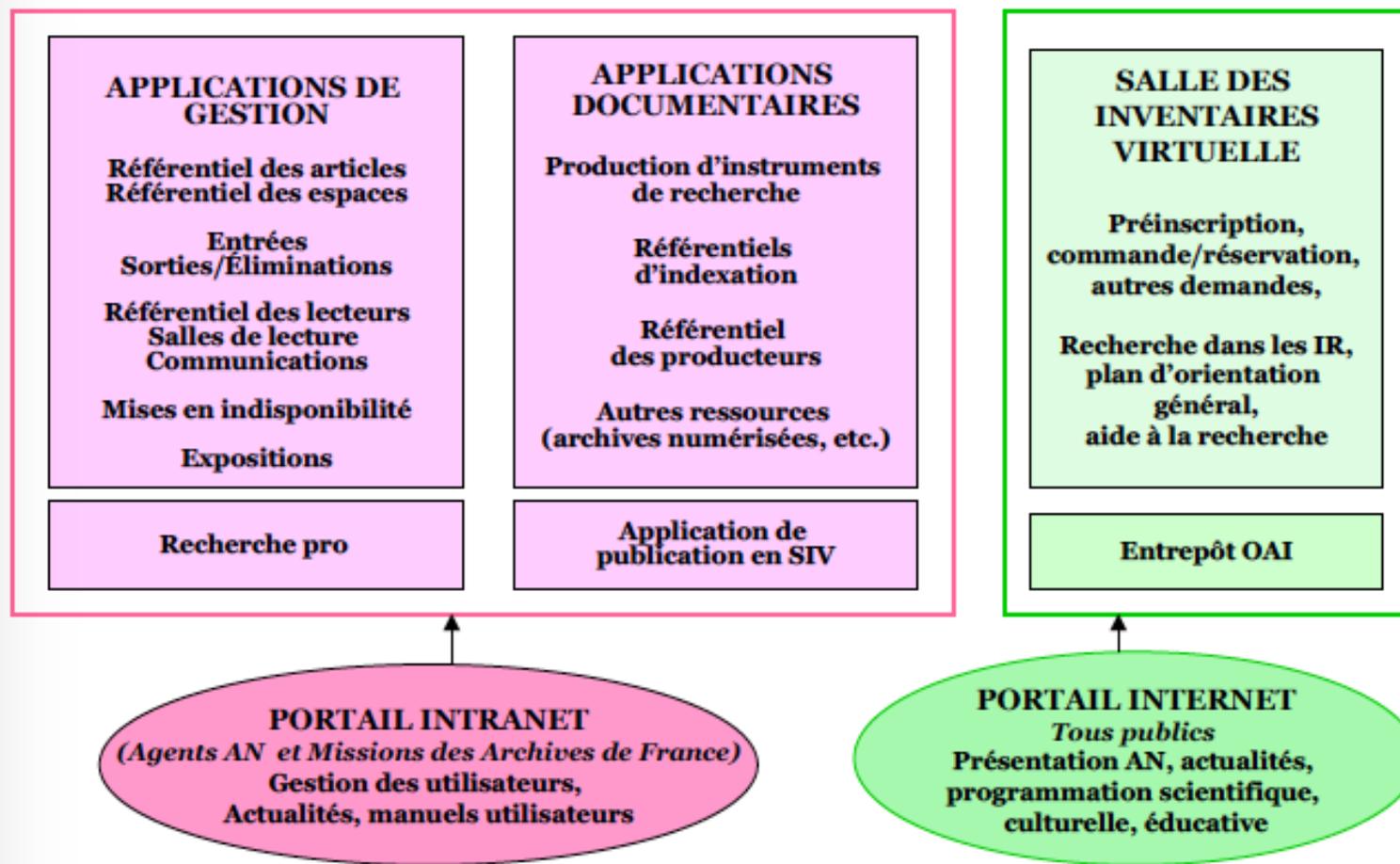
SIA



Calendrier



Architecture fonctionnelle du SIA



Architecture applicative

- **La solution s'appuie sur l'intégration de développements spécifiques et de solutions progiciels du marché, conformes aux standards :**
 - Application de gestion : plateforme Java / JEE ;
 - Production documentaire : client riche d'édition XML (Oxygen) basé sur Eclipse RCP ;
 - Référentiels documentaire et multimédia : instances Alfresco ;
 - Portails (intranet et internet) : publication et gestion de contextes (CMS Liferay) ;
 - Application de gestion des utilisateurs (Meibo), en synchronisation avec l'annuaire LDAP du ministère ;
 - Plateforme d'échanges (web services) : BPM, médiation, MoM et annuaire UDDI ;
 - Moteur de recherche Sinequa pour la SIV.
- **Une solution en client léger**
 - ▶ Accès aux applications par le navigateur internet (ex. Mozilla Firefox) avec identifiant/mot de passe.
 - ▶ Seule l'application documentaire nécessite l'installation d'un logiciel sur le poste de travail.

Les objets documentaires du SIA sont :

Les instruments
de recherche
(IR)

Les notices des
producteurs
(NP)

Le plan
d'orientation
générale (POG)

Les référentiels
d'indexation
(RI)

Les modèles de
documents

Les archives
numérisées

Les documents
annexes

Les ressources
multimédia

Les IR /XML [Définitions]



Instruments de recherche [IR]



Dans les archives, un instrument de recherche est un ouvrage, un fichier ou une base de données qui décrit le contenu d'un ensemble de documents conservés par un service d'archives

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/css" href="ead.css"?> <!DOCTYPE ead SYSTEM "ead_sia.dtd">
<ead>
<eadheader countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b"
relatedencoding="DC" repositoryencoding="iso15511" scriptencoding="iso15924">
<eadid>FRAN_IR_054486</eadid>
<filedesc>
<titlestmt>
<titleproper>Fonds Édouard Daladier. Supplément : internement et procès de Riom (1940-1945).</titleproper>
<subtitle>Répertoire numérique détaillé de la sous-série 496AP (496AP/68-496AP/72).</subtitle>
<author>Par Vanessa Aspart, secrétaire de documentation et Virginie Grégoire, secrétaire de documentation.
Sous la direction d'Isabelle Aristide-Hastir, conservateur général, responsable du département des Archives
privées.</author>
</titlestmt>
<editionstmt>
<edition>Première édition électronique</edition>
</editionstmt>
<publicationstmt>
<publisher>Archives nationales (France)</publisher>
<address>
<addressline>Pierrefitte-sur-Seine</addressline>
</address>
```

Langage informatique de description (des données à l'aide de balises balisage)

Il permet de décrire et structurer un ensemble de données selon un jeu de règles (personnalisable) et des contraintes définies. On peut par exemple utiliser ce type de langage pour décrire l'ensemble des livres d'une bibliothèque, ou encore la liste des chansons d'un CD, etc. Parmi les langages de description les plus connus, on peut citer le SGML, le XML ou encore le HTML.



- Méta-langage [car il permet de définir d'autres langages avec chacun leur grammaire propre]
- XML ne propose pas un jeu de balises prédéfini.
- Lisible pour une machine et pour un humain
- Aujourd'hui utilisé pour toute représentation et échange de données structurées
- Associé à une CSS pour la présentation : XML ne s'occupe que de la structuration de l'information
- Contrôlé par une DTD (ou par Schéma XML ou Relax NG) qui définissent sa structure

LES NP / XML EAC (CPF)



Le référentiel
des
producteurs et
les Notices
d'autorité aux
Archives
nationales

**Construction du référentiel
et chantier de description
des producteurs :**

**enjeu majeur des AN inscrit
dans le PSCE 2017-2020**

Définitions adoptées aux AN

Le Référentiel des producteurs est une base de connaissances rassemblant toutes les informations dont les AN disposent sur la provenance de leurs « fonds ».

Il contient des notices d'autorité sur les producteurs

Une NP

Il s'agit d'une notice d'autorité associant une forme autorisée du nom à des éléments d'information qui permettent d'identifier et de décrire cette entité, et qui peuvent la mettre en relation avec d'autres notices d'autorité (Authority record).

Le réf aujourd'hui constitué d'un ensemble de notices, toutes stockées sous la forme de fichiers dans le même format XML, EAC-CPF. Il y a un fichier XML par notice.

Environ 14600 notices validées et publiées en août 2018 (sans compter environ 300 notices à intégrer et valider par mission référentiels, en permanence, et autant en cours de rédaction chez les rédacteurs et les responsables scientifiques)

Environ **14,600** notices validées et publiées en avril 2018 (sans compter environ 300 notices « à valider », en permanence, et autant en cours de rédaction)



Noyau du référentiel : **3,426** notices sur les études et notaires de Paris du XVe siècle à nos jours (issues du projet ETANOT, mises à jour en continu depuis 2013) + **9,700** notices environ issues de la base **PRIAM** du Centre des Archives contemporaines (nomenclature créée pour la gestion des versements entrés aux AN entre 1970 et 2012)



Environ **1,400** nouvelles notices, créées directement dans le SIA, depuis le déploiement des modules documentaires du SI en 2013

Suite

Près de 72 % des notices décrivent des collectivités



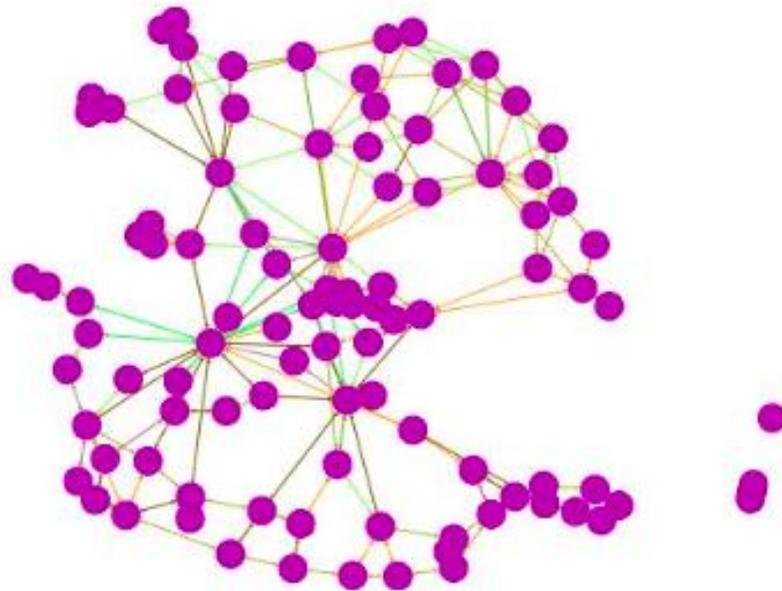
Le référentiel n'est pas une simple collection de notices car il existe plus de 36000 relations entre notices et plus de 27000 relations avec des unités de description de documents au sein des instruments de recherche.

Graphes

- urn:plaaaf:an (390/1889)
- urn:plaaaf:bnf (0/466)
- urn:plaaaf:data (0/526)
- urn:plaaaf:slaf (0/818)

Resources

- <inconnu> (0/167)
- Agent (0/312)
- ★ Collectivité (101/438)
- Fonction (concept) (0/91)
- Catégorie de collectivité (0/0)



Dates

1568 : 2016

GRAPHE

Votre Rôle

Hackathoniens les 8 et 9 décembre pour proposer des prototypes.



Informations pratiques sur l'événement

► Un prix à gagner !

Un prix sera attribué à l'équipe gagnante du hackathon, par un jury spécialement composé à cet effet. Il sera d'un montant de 4000 euros.

► Public attendu

Cet événement est largement ouvert et s'adresse à ceux qui connaissent les archives comme à ceux qui les ignorent ; il doit toucher des publics issus de mondes divers et complémentaires : développeurs, data scientists, ergonomes, designers, chercheurs, généalogistes, venant d'horizons variés, institutions culturelles, de recherche ou d'enseignement, humanités numériques, open data, entreprises de service du numérique, etc.

► Dates, horaires, lieu

L'ouverture de l'événement se fera le vendredi 7 décembre à 18 h aux Archives nationales, site de Paris. Le hackathon se déroulera à Paris au Liberté Living Lab (9 rue d'Alexandrie 75002 Paris), et sans interruption, du samedi 8 décembre à 13h jusqu'au dimanche 9 décembre à 18h.



Le lieu : Liberté Living Lab
9 rue d'Alexandrie 75002

Merci de votre attention !

▶ Frédéric ZAMARRENO

frederic.zamarreno@culture.gouv.fr

▶ Gaetano PIRAINO

gaetano.piraino@culture.gouv.fr

