
Le projet PIAAF

sémantisation et visualisation
orientée utilisateur
de métadonnées archivistiques

Florence Clavaud (Archives nationales de France)
Emmanuel Château-Dutier (UdeM, CRIHN)

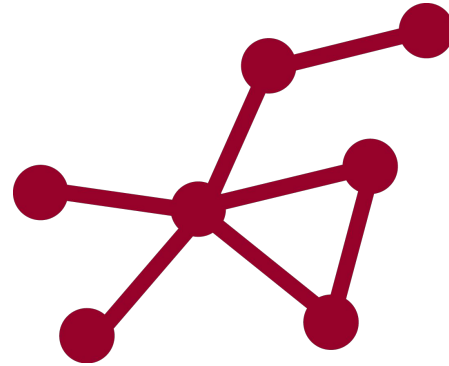


**ARCHIVES
NATIONALES**

{ BnF |  CRIHN

Une preuve de concept

De quoi s'agit-il ?



Livrables

- jeux de données & scripts de conversion
- bibliothèques de visualisation
- rapport critique

Le partenariat



**ARCHIVES
NATIONALES**

{ BnF |



CRIHN

- notices d'autorité
- instruments de recherche



Logilab



Démo

sources ss.
licence libre

45 000 €

Calendrier



2015 – cadrage & organisation

2016 – jeux de métadonnées & appel d'offres

2017 – réalisation en 3 itérations

automne 2017 – publication du prototype
+ jeux de données + logiciels

Contexte et enjeux

ARCHIVES NATIONALES

- 25 000 instruments de recherche XML/EAD
- 14 200 notices d'autorité XML/EAC-CPF
- plusieurs référentiels (vocabulaires)

Ouvrir le silo et libérer les données

Révéler le graphe sous-jacent

- 36 000 relations entre notices d'autorité
- 27 000 relations entre notices et IR

Pour un chercheur

- Évaluer l'**applicabilité des modèles archivistiques** en contexte de recherche
- Visualiser de façon diachronique les **organigrammes administratifs**
- Ménager un **accès analytique** à partir des données contenues dans les Instruments de recherche

Des enjeux stratégiques pour les trois partenaires

- qualité et expressivité des données
- croisement des données et établissement de relations
- accroître les possibilités
 - d'exploration,
 - de découverte
 - et de construction d'information

RiC-O

Records in Contexts-Ontology

- International Council on Archives
- EGAD group, mandaté 2013-2020
- Publication de l'ontologie OWL en novembre 2017

<http://www.ica.org/en/egad-ric-conceptual-model>

Articuler les quatre normes conceptuelles tout en (re)définissant les concepts, tenir compte des avancées de la réflexion et de la pratique archivistique et des besoins des utilisateurs, aller plus loin dans l'expressivité



International Council on Archives
Conseil International des Archives

Le défi de la data visualisation web

- densité informationnelle et variété des objets
- articuler données complexes et dimension temporelle
- granularité et lisibilité

NYPL Archives and Manuscripts. Beta Tools & Experiments



gilles du motier

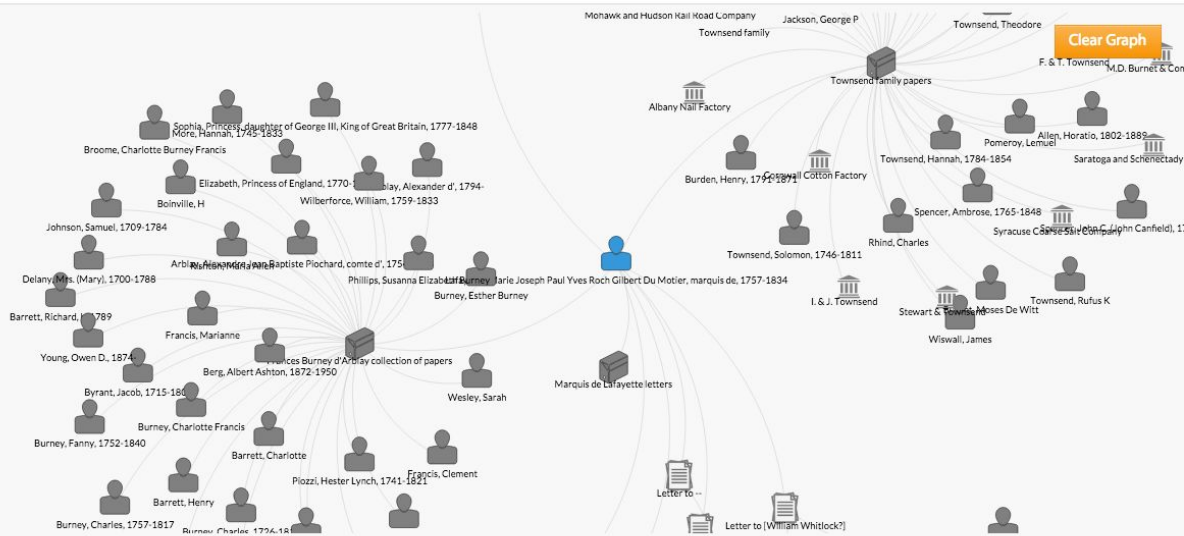


+ Add Person

+ Add Org.

+ Add Topic

+ Add Title



Kindred Britain

KINDRED BRITAIN

PEOPLE

CONNECTIONS

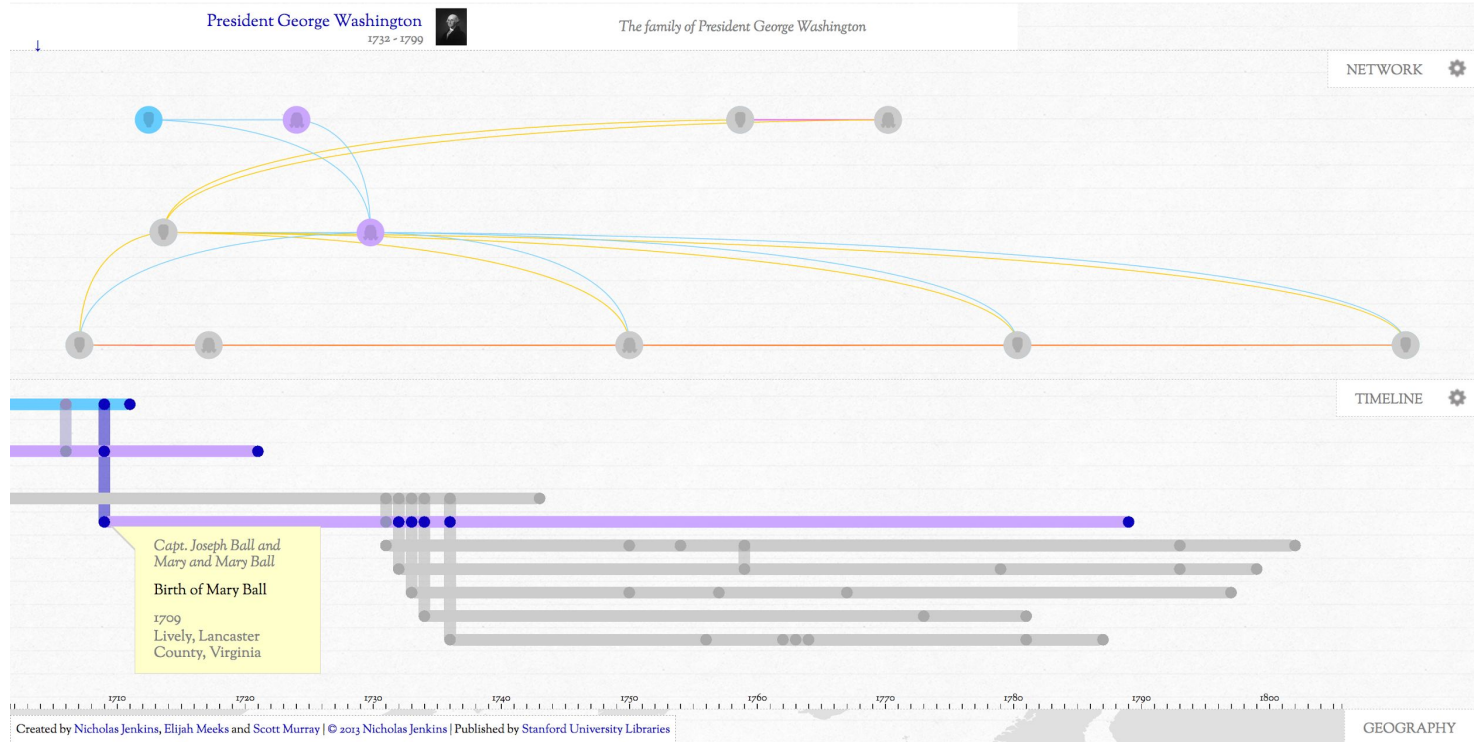
STORIES

🔍 🗺️ ?

President George Washington
1732 - 1799



The family of President George Washington



Réalisation

Définition des jeux de métadonnées

Réels, pertinents, maîtrisables et articulables

2 sous-domaines fonctionnels du ministère de la Culture

- gestion des Monuments historiques et des « Bâtiments civils »
- lecture publique et bibliothèques

Règles de description

Chacun travaille **selon sa perspective et avec ses outils**

Quelques règles générales adoptées par tous les partenaires

- formats d'encodage (XML/EAC-CPF et XML/EAD 2002)
- forme des noms des entités
- informations obligatoires (identifiants des notices, dates d'existence des entités, statut juridique, ...)
- saisie de l'identifiant ISNI, liens data.bnf.fr, Wikipédia
- travail de sous-catégorisation des relations entre entités

Production de vocabulaires communs

- Indexation des fonctions des entités
- Combinaison : domaine fonctionnel + activité
- Gestionnaire de référentiels Gingo (SKOS/RDF)

```
<function localType="piaafFunction">  
  <term vocabularySource="http://data.culture.fr/thesaurus/resource/ark:/67717/f14e8183-5885-46d6-8fc9-17ebd8f3c27e" piaaf:rank="1"  
    piaaf:conceptId="http://data.culture.fr/thesaurus/resource/ark:/67717/4002e166-eb11-446c-a777-cb73b6def641">objets culturels</term>  
  <term vocabularySource="http://data.culture.fr/thesaurus/resource/ark:/67717/51232822-adac-4a33-aa14-29e2c701a5ee" piaaf:rank="2"  
    piaaf:conceptId="http://data.culture.fr/thesaurus/resource/ark:/67717/b06a81ff-bbe0-4bf3-b33e-e51d26eb0431">conservation</term>  
</function>
```

Jeux de métadonnées sources

300 notices d'autorité

- dont 14 concernant des catégories de collectivités territoriales

41 instruments de recherche (IR)

2 vocabulaires SKOS

les complémentarités et les chevauchements sont effectifs

Classes RiC-O utilisées

RiC:Agent,
RiC:CorporateBody,
RiC:Person
RiC:GroupType
RiC:FunctionAbstract
RiC:Position
RiC:Mandate
RiC:Event
RiC:AgentName
RiC:Place
RiC:Relation
RiC:Description
RiC:Record
RiC:RecordSet
RiC:Proxy
RiC:FindingAid

Conversion vers RDF

Ne rien perdre

Individualiser les objets « immergés »

- **dans les notices d'autorité en EAC/CPF** : des noms, des événements, des postes, des fonctions, des lieux, des relations entre entités ;
- **dans les fichiers EAD** : via la hiérarchie XML, des relations d'agrégation et des séquences documentaires ; des relations de provenance ; des sujets

Problèmes traités

- quels objets / quels URI ?
- comment traiter les **relations entre agents** ?
- comment traiter les **catégories de collectivités** ?
- comment représenter correctement les fonctions résultant de la **combinaison de concepts** ?
- comment prendre en charge les **séquences d'objets** dans des agrégats documentaires ?

Relations n-aires dans RiC-O

The screenshot shows a web interface for the RiC-O ontology. On the left is a tree view of classes, with 'Relation' selected. The tree includes classes like AuthorityRelation, AgentControlRelation, AgentLeadershipRelation, SubordinationRelation, ManagingRelation, OwnershipRelation, RightsHoldingRelation, AuthorizationRelation, AuthorshipRelation, CharacterizationRelation, DenominationRelation, IdentificationRelation, SubjectRelation, MainSubjectRelation, TypeRelation, EvidenceRelation, FunctionFulfillmentRelation, LocationRelation, MembershipRelation, AgentMembershipRelation, AgentLeadershipRelation, PerformanceRelation, RecordsProvenanceRelation, RecordsAccumulationRelation, RecordsProductionRelation, ActivityProductionRelation, AgentProductionRelation, OccupationProductionRelation, and PositionProductionRelation. On the right, the 'Relation' class is detailed. It has an rdfs:comment in English: 'The top level relation class. Connects two entities. Has some data and object properties : - description (data property) like any thing - certainty (for 'certain', 'quite probable', 'uncertain', 'unknown') - evidence (object property) - begin date (data property ? ; if so, xs:datetime) (or, better, object property, since we have a Date class) - end date (data property ? ; if so, xs:datetime) (same reasoning as for begin date) - state (for « past », « present », « unknown ») - place (in some cases) (object property) - context (hasContext object property) - associates at most 2 things'. Below this, there is a 'Description: Relation' section with 'Equivalent To' and 'SubClass Of' lists. The 'SubClass Of' list includes properties like beginningDate max 1 rdfs:Literal, certainty max 1 rdfs:Literal, description max 1 rdfs:Literal, endDate max 1 rdfs:Literal, hasContext only owl:Thing, hasPlace only Place, relationAssociates exactly 2 owl:Thing, and state max 1 rdfs:Literal.

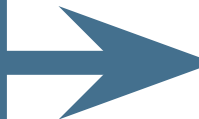
```
<rdf:Description rdf:about="http://www.piaaf.net/relations/FRAN_agent_relation_050789-d249e473">
  <rdf:type rdf:resource="http://www.ica.org/standards/RiC/ontology#AgentLeadershipRelation"/>
  <rdfs:label xml:lang="fr">Relation
    de "leadership" ou de direction entre les agents "Jeanneney, Jean-Noël (1942-....)" et "Radio France"</rdfs:label>
  <RiC:description xml:lang="fr">Jean-Noël Jeanneney a été président de Radio France et RFI</RiC:description>
  <RiC:beginningDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1982-01-01</RiC:beginningDate>
  <RiC:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1986-12-31</RiC:endDate>
  <RiC:leadershipBy rdf:resource="http://www.piaaf.net/agents/FRAN_person_050789"/>
  <RiC:leadershipOn rdf:resource="http://www.piaaf.net/agents/FRAN_corporate-body_005419"/>
  <!--coll='FRAN'-->
  <RiC:leadershipWithPosition rdf:resource="http://www.piaaf.net/positions/FRAN_position_050789-d249e473"/>
</rdf:Description>
```

Chaîne de traitement



XSLT 2

0-generateIdentifiers.xml
1-processFunctions.xml
2-processLegalStatuses.xml
3-processMandates.xml
4-processEvents.xml
5-processAgentNames.xml
6-processPlaces.xml
7-processPositions.xml
8a-8d-processRelations...xs
|
9-processGroupTypes.xml
10-processAgents.xml
11-12-recordsSets.xml
13-processProxies.xml



agent-names/
agents/
events/
group-types/
mandates/
ontology/
places/
positions/
proxies/
record-sets/
records/
relations/
vocabularies/
functions.rdf
legal-statuses.rdf

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <rdf:RDF xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:RiC="http://www.ica.org/standards/RiC/ontology#"
3 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:piaaf-onto="http://www.piaaf.net/ontology#"
4 xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:isni="http://isni.org/ontology#"
5 xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
6 <rdf:Description rdf:about="http://www.piaaf.net/agents/FRAN_person_050789">
7 <rdf:type rdf:resource="http://www.ica.org/standards/RiC/ontology#Person"/>
8 <rdfs:label xml:lang="fr">Jeanneney, Jean-Noël (1942-....)</rdfs:label>
9 <RiC:hasDenominationRelation
10 rdf:resource="http://www.piaaf.net/relations/FRAN_DenominationRelation_050789_d78e96"/>
11 <RiC:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1942-04-02</RiC:birthDate>
12 <RiC:history xml:lang="fr">Jean-Noël Jeanneney est né en avril 1942 à Grenoble. Après des études aux
13 lycées Montaigne et Louis-le-Grand à Paris, il est reçu à l'École normale supérieure de la rue d'Ulm où il est
14 élève de 1961 à 1966. Diplômé de l'Institut d'études politiques de Paris en 1964, agrégé d'histoire en 1965, il
15 consacre une année,
16 [...]
17 Il a été enfin président de la Bibliothèque nationale de France de mars 2002 à mars 2007, y conduisant en
18 particulier un combat pour la création d'une bibliothèque numérique européenne. Il est docteur honoris causa de
19 l'Université libre de Bruxelles.</RiC:history>
20 <owl:sameAs rdf:resource="http://data.bnf.fr/ark:/12148/cb11908748q#foaf:Person"/>
21 <owl:sameAs rdf:resource="http://fr.dbpedia.org/resource/Jean-No%C3%ABJ_Jeanneney"/>
22 <isni:identifierValid>0000000384320062</isni:identifierValid>
23 <RiC:headOf rdf:resource="http://www.piaaf.net/relations/FRAN_agent_re
24 <RiC:headOf rdf:resource="http://www.piaaf.net/relations/FRAN_agent_re
25 <RiC:headOf rdf:resource="http://www.piaaf.net/relations/FRAN_agent_re
26 <RiC:headOf rdf:resource="http://www.piaaf.net/relations/FRAN_agent_re
27 <RiC:associatedWithRelation
28 rdf:resource="http://www.piaaf.net/relations/FRAN_agent_relation_050789-d249e2
29 <RiC:associatedWithRelation
30 rdf:resource="http://www.piaaf.net/relations/FRAN_agent_relation_050789-d249e2
31 <RiC:associatedWithRelation

```

Jean-Noël Jeanneney vu dans les univers archivistiques de la BnF et des AN

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <rdf:RDF xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:RiC="http://www.ica.org/standards/RiC/ontology#"
3 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:piaaf-onto="http://www.piaaf.net/ontology#"
4 xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:isni="http://isni.org/ontology#"
5 xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
6 <rdf:Description rdf:about="http://www.piaaf.net/agents/FRBNF_person_0296">
7 <rdf:type rdf:resource="http://www.ica.org/standards/RiC/ontology#Person"/>
8 <rdfs:label xml:lang="fr">Jeanneney, Jean-Noël (1942-....)</rdfs:label>
9 <RiC:hasDenominationRelation
10 rdf:resource="http://www.piaaf.net/relations/FRBNF_DenominationRelation_0296_d61e109"/>
11 <RiC:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#Year">1942</RiC:birthDate>
12 <RiC:history xml:lang="fr">Jean-Noël Jeanneney est un haut fonctionnaire de l'Etat. Il a étudié à l'École
13 normale supérieure de 1961 à 1966, est diplômé de l'Institut d'études politiques de
14 Paris en 1964, agrégé d'histoire en 1965, docteur ès lettres en 1975. Il enseigne à
15 l'Institut d'études politiques de Paris à partir de 1968 et à l'Université de Paris X
16 [...].
17 ligne de près de 39 millions de documents numérisés représentatifs du patrimoine
18 historique et culturel de l'Europe. En 2007, atteint par la limite d'âge, il quitte
19 ses fonctions.</RiC:history>
20 <RiC:conformsToRule rdf:resource="http://www.piaaf.net/mandates/FRBNF_mandate_0296-d57e216"/>
21 <RiC:conformsToRule rdf:resource="http://www.piaaf.net/mandates/FRBNF_mandate_0296-d57e222"/>
22 <owl:sameAs rdf:resource="http://data.bnf.fr/ark:/12148/cb11908748q#foaf:Person"/>
23 <isni:identifierValid>0000000384320062</isni:identifierValid>
24 <RiC:headOf rdf:resource="http://www.piaaf.net/relations/FRBNF_agent_relation_0296-d57e272"/>
25 <RiC:agentFollows rdf:resource="http://www.piaaf.net/relations/FRBNF_agent_relation_0296-d57e293"/>
26 <RiC:agentPrecedes rdf:resource="http://www.piaaf.net/relations/FRBNF_agent_relation_0296-d57e314"/>
27 <RiC:hasProvenanceRelation
28 <RiC:hasProvenanceRelation
29 rdf:resource="http://www.piaaf.net/relations/FRBNF_ProvenanceRelation_0296-d57e329"/>
30 <RiC:hasProvenanceRelation
31 <RiC:hasProvenanceRelation
32 <RiC:hasProvenanceRelation_0296-d57e335"/>
33 <RiC:describedBy rdf:resource="http://www.piaaf.net/descriptions/FRBNF_description_0296"/>
34 </rdf:Description>
35 <rdf:Description rdf:about="http://www.piaaf.net/relations/FRBNF_DenominationRelation_0296_d61e109">

```

Premier bilan

- opération non triviale
- réalisable avec le niveau de qualité attendu
- encourageant pour la suite du projet, les projets comparables et pour valider l'ontologie RiC-O

Points d'attention

- limites du modèle EAC-CPF
- produire des ids pérennes en amont de la conversion
- revoir les pratiques d'encodage

Cahier des charges du prototype

- recherche experte SPARQL, et consultation par facettes
 - visualisations en graphes dynamiques diachroniques
 - alignements de données et inférences
-
- analyse de la qualité des jeux de données fournis
 - approche agile
 - développements *open source*

Réalisation en cours

1/3 itération (terminée)

- mise en place du triplestore + SPARQL endpoint
- import des données RDF
- première visualisation en graphe (partielle)

2/3 itération

- recherche guidée par facettes
- requêtes préparées SPARQL (pertinentes et typiques)
- traitement de la dimension temporelle
- ajout de fonctionnalité d'alignements internes/externes

Jeanney, Jean-Noël (1942-....)

http://www.piaaf.net/agents/FRAN_person_050789

Start layout

Graphes

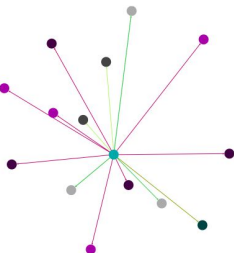
- urn:plafan

Ressources

- r:chconnus
- ric:Agent
- ric:CorporateBody
- ric:Person
- ric:Position
- ric:RecordSet

Relations

- associatedWithRelation
- hasProvenanceRelation
- headOf
- sameAs



Attributs

rdflabel	Jeanney, Jean-Noël (1942-....)
ric:birthDate	1942-04-02
richistory	Jean-Noël Jeanneney est né en avril 1942 à Grenoble. Après des études aux lycées Montaigne et Louis-le-Grand à Paris, il est reçu à l'École normale supérieure de la rue d'Ulm où il est élève de 1961 à 1966. Diplômé de l'Institut d'études politiques de Paris en 1964, agrégé d'histoire en 1965, il consacre une année, avant son service militaire, à faire le tour du monde grâce à une bourse Singer-Polignac. Il enseigne ensuite l'histoire contemporaine à l'université de Nanterre comme assistant puis maître-assistant. Docteur ès-lettres, en 1975, avec une thèse intitulée "François de Wendel en République, l'Argent et le pouvoir", il est nommé professeur des universités à Sciences Po en 1977. Il y enseigne depuis lors l'histoire politique, l'histoire culturelle et l'histoire des médias, sujets sur lesquels il a publié de multiples ouvrages, parmi divers livres touchant à d'autres domaines. Il est l'auteur de nombreux documentaires historiques pour la télévision, de 1973 à 2009. Il appartient à l'équipe fondatrice de la revue L'histoire. Il a été co-président fondateur du Festival international de Radio France et de Montpellier (1985-1986). Il a présidé, entre 1991 et 2000, le Conseil scientifique de l'Institut d'Histoire du temps présent. Il a été chroniqueur au Journal du Dimanche de 1999 à 2001. Il a été membre du Conseil d'administration (1987-1991 puis 1993-1995) puis de surveillance (1995-2002) des éditions du Seuil et président du Conseil d'orientation de la chaîne de télévision câblée « Histoire » de 1997 à 2004, membre du Conseil d'administration de l'École normale supérieure entre 1998 et 2004, membre du Haut Conseil de l'Institut du monde arabe de 2003 à 2008, membre du Conseil d'administration du Fonds d'action SACEM de 1982 à 2013. Il préside le Conseil scientifique des « Rendez-vous de l'Histoire de Blois » depuis 2003. Il est président d'honneur du Festival international du film historique de Pessac depuis 1990. Il préside le jury du Prix du livre d'Histoire du Sénat depuis 2007, les Rencontres de la photographie d'Arles depuis septembre 2009. Depuis mars 1998, il co-préside l'association Européennaires. Il produit l'émission hebdomadaire « Concordance des temps », sur France Culture, depuis 1999. Il a exercé diverses activités publiques : président de Radio France et de Radio France Internationale de 1982 à 1986, président de la Mission du Bicentenaire de la Révolution française et de la Déclaration des droits de l'homme et du citoyen (1988-1989), secrétaire d'État au Commerce extérieur (1991-1992) puis à la Communication (1992-1993) sous la seconde présidence de François Mitterrand. Il a été conseiller régional de Franche-Comté de 1992 à 1998. Il a été enfin président de la Bibliothèque nationale de France de mars 2002 à mars 2007, y conduisant en particulier un combat pour la création d'une bibliothèque numérique européenne. Il est docteur honoris causa de l'Université libre de Bruxelles.

Relations ayant cette ressource comme sujet

rdftype	http://www.ica.org/standards/RIC/ontology#Person
ow:sameAs	http://aida.bnf.fr/ark:/12148/cb11988748q#foaf:Person http://fr.dbpedia.org/resource/Jean-No%C3%A9_AJ_Jeanneney
ric:associatedWithRelation	http://www.piaaf.net/relation/FRAN_agent_relation_050789-d249e256 http://www.piaaf.net/relation/FRAN_agent_relation_050789-d249e279 http://www.piaaf.net/relation/FRAN_agent_relation_050789-d249e322 http://www.piaaf.net/relation/FRAN_agent_relation_050789-d249e322

Archives papier de Jean-Noël Jeanneney, Président-directeur général de Radio France (1964-1986)

Start layout

Graphes

- urn:plafan

Ressources

- ric:Person
- ric:RecordSet

Relations

- originatedBy

Attributs

ric:title	Archives papier de Jean-Noël Jeanneney président-directeur général de Radio France et de Radio France Internationale de 1982 à 1986 (1964-1986)
-----------	---

Relations ayant cette ressource comme sujet

rdftype	http://www.ica.org/standards/RIC/ontology#RecordSet
ric:mainSubjectOf	https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/FRAN_IR_050629
ric:originatedBy	http://www.piaaf.net/relation/FRAN_ProvenanceRelation_050789-d249e496

Relations ayant cette ressource comme objet

ric:originatedObject	http://www.piaaf.net/relation/FRAN_ProvenanceRelation_050789-d249e496
----------------------	---

Conclusion

Conclusion

- **automne 2017** : publication du démonstrateur à l'automne + logiciels + RDF
- **courant 2018** : publication d'un bilan critique et prospectif
- **la preuve de concept est dès à présent fructueuse**
- **le projet ouvre des perspectives à la fois pour les partenaires et pour d'autres acteurs du domaine**

La description des contextes archivistiques est l'opportunité d'un travail conjoint entre archivistes, bibliothécaires, et chercheurs.

Merci !

florence.clavaud@culture.gouv.fr
@FloClavaud

emmanuel.chateau.dutier@umontreal.ca
@emchateau