

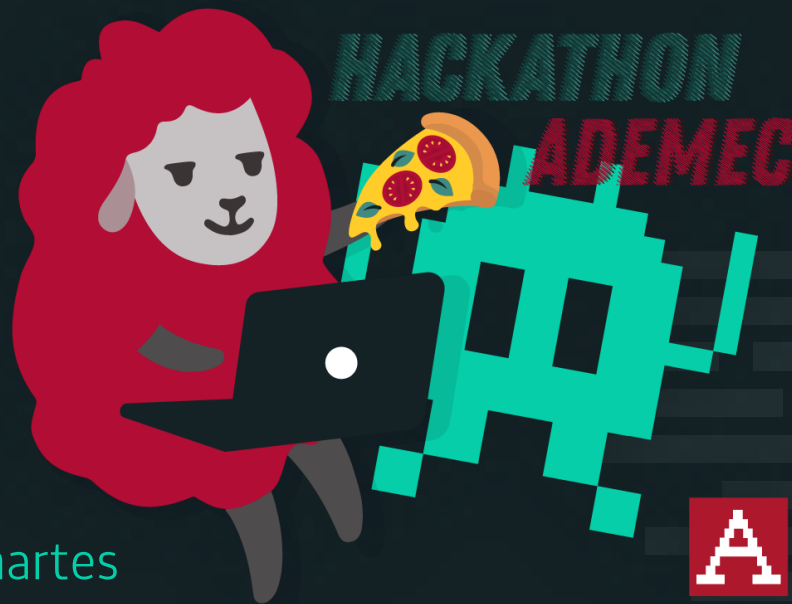


CARNET DE BORD

ValDa (VALorisation DAta)

Hackathon ADEMEC/Ecole nationale des chartes

24-26 septembre 2021



VAL-DA

l'équipe VALorisation des DAta

Lucie Rondeau du Noyer
Jean-Philippe Moreux
Jean-Damien Généro
Antoine Sylvestre de Sacy
Aurélien Conraux

Le pad collaboratif :
<https://demo.hedgedoc.org/euyefCUxRmOmMaqrG5LZA?both>

Le code :
<https://github.com/ademec/Valda>

VAL-DA : un projet agile et structuré

Découverte

Données biographiques,
scientifiques et
géographiques

Idéation

Géolocalisation des travaux
de recherches, travail
sur les photographies

Sélection

Lier les données, rendre
accessible les
informations

Formulation

3 directions de recherche à
intégrer : données,
textes, images

Réalisation

Une application FLASK
contenant un « data
paper » interactif

Perspectives

[à suivre...]

Idéation

Les territoires sont-ils
valorisés dans les projets
de recherche ?

Les chartistes sont-ils
tous Parisiens ?

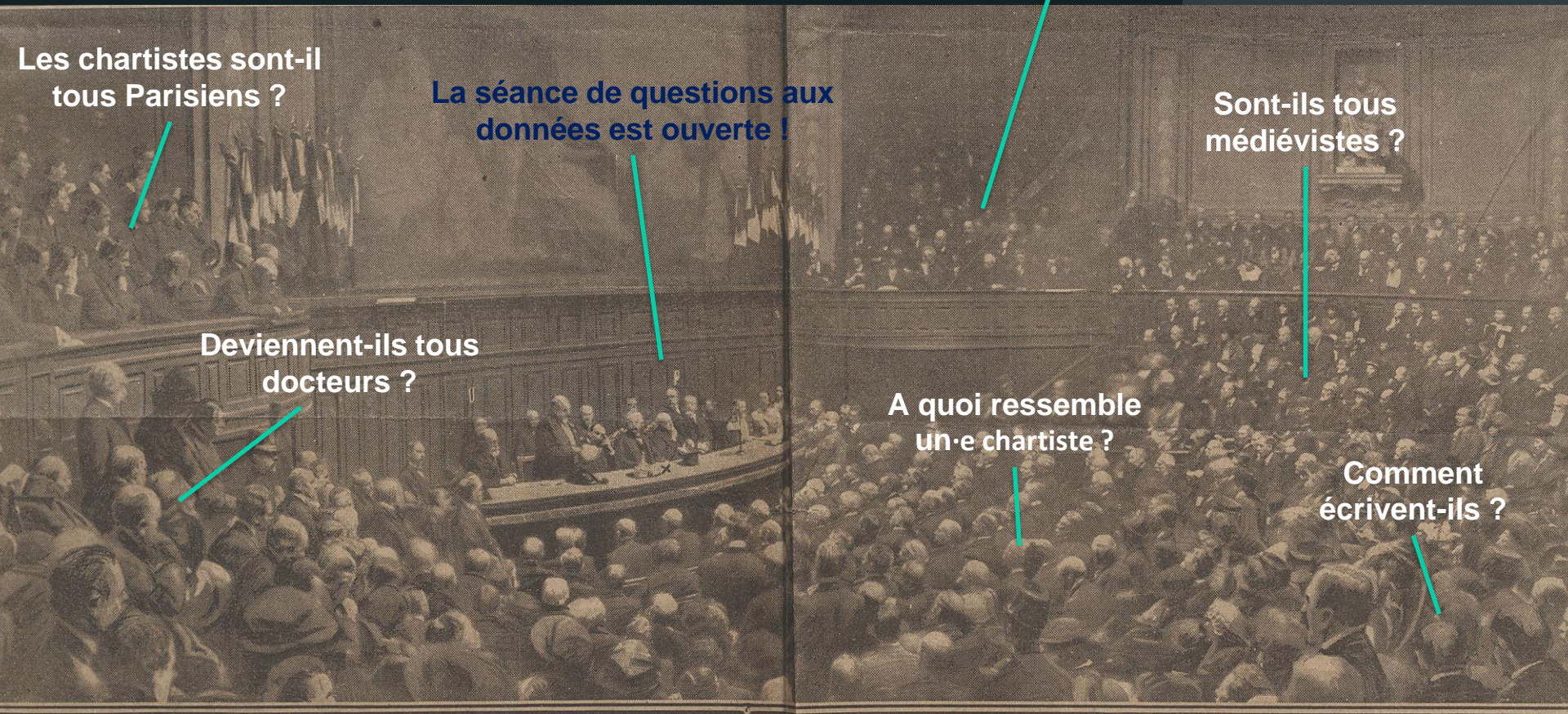
La séance de questions aux
données est ouverte !

Sont-ils tous
médiévistes ?

Deviennent-ils tous
docteurs ?

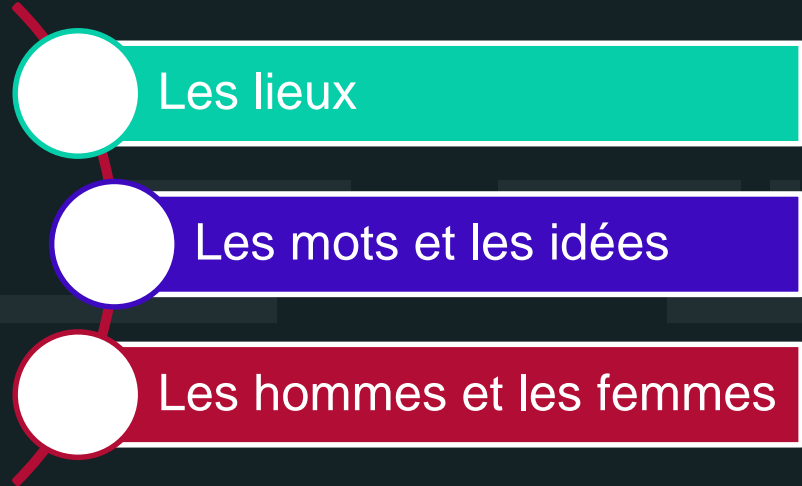
A quoi ressemble
un-e chartiste ?

Comment
écrivent-ils ?



NOTRE MISSION : extraire & représenter, de façon simple et attractive,
des ensembles de données liées aux chartistes,
en lien avec les individus et les ressources scientifiques

- Une démarche sociologique
- Une valorisation accessible au grand public
- L'accès direct aux sources primaires et secondaires



Extraire & représenter les lieux

À partir des
titres de thèse



Moyens mis en
oeuvre

Résultats

```
Paris ==== Le sanctoral de l'ordinaire de Saint-Victor de Paris. Edition critique et commentaire
Paris ==== Paris capitale de guerre (1792-1794)
Paris ==== Chapeaux, chapeliers et autres couvre-chefs à Paris (1550-1660). Aspects économiques, sociaux et
Paris ==== Pratiques sociales et charitables en France dans la seconde moitié du XVIIIe siècle. L'exemple d
Paris ==== Représentation du roi et pouvoir de « faire loy » : enquête sur l'activité « réglementaire » du
Paris ==== Des « messes théâtrales » au Concert spirituel : chanteuses laïques et musique religieuse à Pari
Paris ==== La « Réformation des Halles » : un lotissement royal à Paris (1543-1572)
Paris ==== Michel Villedo et Cie : entreprise et expertise à Paris au XVIIe siècle
Paris ==== L'Institut de psychologie de l'université de Paris (1920-1970)
Paris ==== La réglementation du marché et le contrôle du commerce à Paris et à Rome, fin XVIIe-début XVIIIe
Paris ==== Le ballet de l'Opéra de Paris sous le Second Empire
Paris ==== L'hydraulique à Paris à l'époque moderne. Adductions et distribution d'eau dans la ville (v. 161
Paris ==== À l'enseigne de La Cloche d'or. Les Le Guay, une dynastie de fondeurs de cloches à Paris sous l'
Paris ==== Souvenirs et impressions du siège de Paris. Edition critique et commentaire
Paris ==== Verre et verriers à Paris dans la seconde moitié du XVIe siècle (1547-1610)
Paris ==== « Aux plumes d'or ». Les maitres écrivains à Paris, calligraphes au XVIIe siècle
Paris ==== L'architecture religieuse de la Renaissance à Paris, dans le Paris et le Vexin
Paris ==== Les vigneron de Sannois au XVIIIe siècle : étude démographique et sociale d'un village du Par
Pau ==== L'historien Paolo Coimi, dit Paul Emile (1455-envron 1529)
Pau ==== La paroisse Saint-Paul des origines à l'année 1559
Pau ==== L'architecture militaire du xie au XVIe siècle dans la viguerie de Grasse et le bailliage de Saint
Pau ==== La participation de François de Meyronnes à la querelle de la Pauvreté (1322-1324)
Pau ==== Le temporel de l'abbaye Saint-Paul de Besançon des origines à 1333
Pau ==== Antoine le Paultre (1621-1679)
Pau ==== Pauvreté et assistance dans le district de Douai sous la Révolution (1788-an V)
Pau ==== Les activités politiques de Joseph Paul-Boncour de 1932 à 1936
Pau ==== Paul Lacroix (1806-1884), le « Bibliophile Jacob » : une passion, une profession
Pau ==== Paul-Etienne Devilliers du Terrage, directeur général de la police en Hollande d'après sa correspo
Pau ==== De la coexistence à l'exil, les fugitifs protestants béarnais à la révocation de l'Édit de Nantes
Pau ==== Victor Hugo - Paul Meurice : correspondance (1852-1855)
Pau ==== Paul Barrillon, ambassadeur de France en Angleterre (1677-1688)
Pau ==== Paul-Émile Victor, biographie d'un explorateur polaire (1907-1995)
Pau ==== Le quartier de l'hôtel Saint-Paul à Paris (1360-1550). Étude topographique, économique et sociale.
Pau ==== La Nouvelle Revue française, « entre Ordre et Aventure ». Jean Paulhan et Marcel Arland. Correspon
Pau ==== « Je rêve à toi presque chaque nuit et je ne suis jamais en prison ». Lettres clandestines de pris
Pau ==== Paul Hay du Chastelet (1592-1636) ou la quotidienneté du pouvoir
Pays-Bas ==== François de Valois, duc d'Alençon, d'Anjou, les Malcontents, les Pays-Bas (1554-1584)
Pays-Bas ==== Essai sur la politique religieuse de Philippe II dans les Pays-Bas
Pays-Bas ==== Essai sur la politique religieuse de Philippe le Bon dans les Pays-Bas
Pays-Bas ==== Recherches sur le monde rural dans les Pays-Bas méridionaux (1480-1630)
Pays-Bas ==== Joutes, tournois et pas d'armes dans les villes des Pays-Bas méridionaux à la fin du Moyen Âge
```

- Données de localisations extraites *via* Spacy
- + nettoyage des données sur Dataiku
- Scripts automatisés : récupération des coordonnées avec la librairie geopy

- Carte interactive avec filtres



```
:type places csv: str
:return: list with placenames
:rtype: list
"""
counter = -1
places = []
with open(places_csv, 'r', encoding='utf-8')
    cities = csv.reader(opening)
    for city in cities:
        places.append(city[1])
# print(places)
return places

def dict_title(placenames, titles):
    """
    :param placenames: list with placenames
    :type placenames: list
    :param titles: list with list ["titltes", "u
    :type titles: list
    :return: dict with {placename: }
    """
    results = {}
    for item in placenames:
        itemname, listname, list
```

Extraire & représenter les lieux

Perspectives

- Étendre aux positions de thèse
- Travailler sur la qualité des entités géographiques
- Intégrer des référentiels géographiques historiques (Austrasie, « Bourgogne du XVe siècle »...
- Gérer des niveaux de visualisation emboîtés, des clustering de points (ex. DicoTopo)
- Ajouter couches de métadonnées
- Affiner l'affichage (par date de soutenance, genre...)

Extraire & représenter les mots et les idées

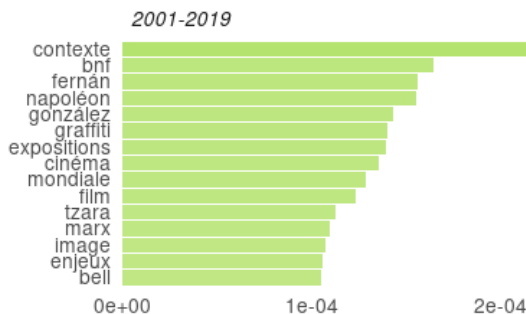
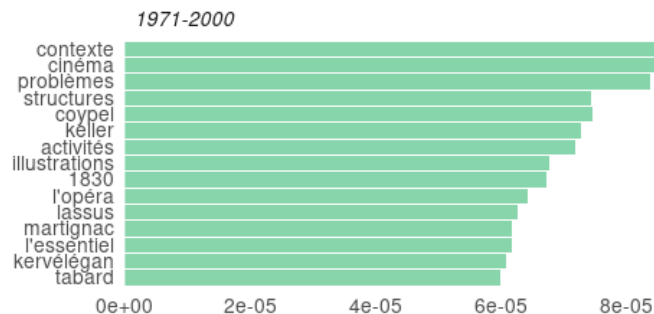
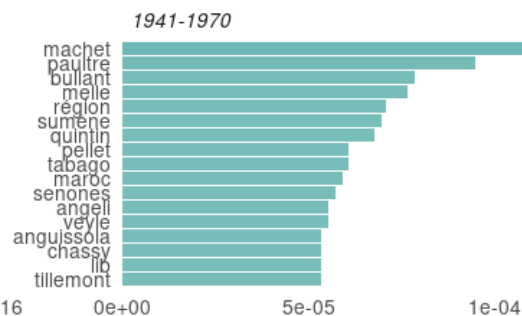
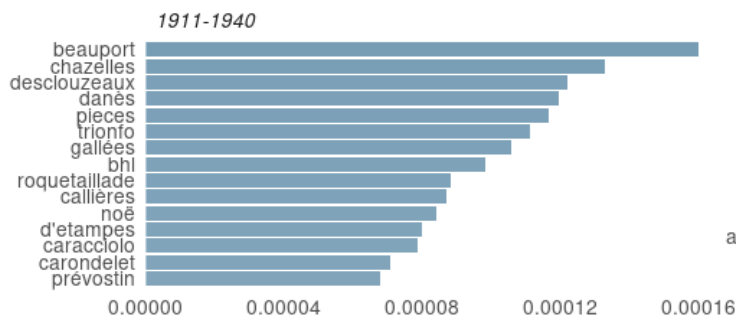
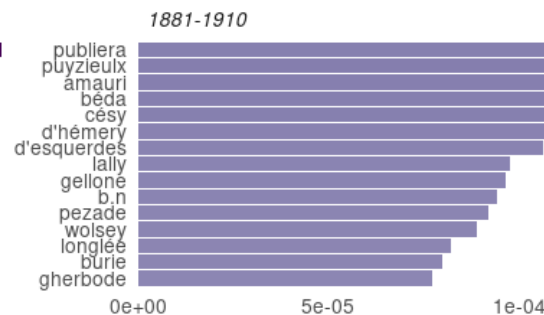
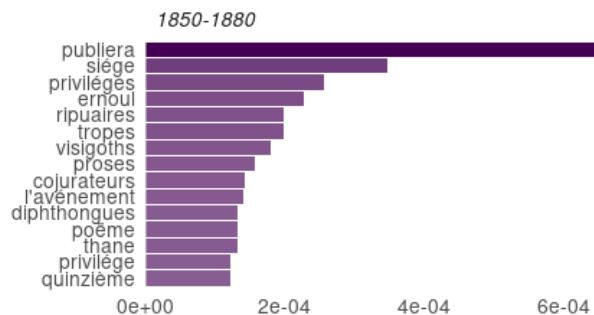
À partir des positions de thèse depuis 1849

- Technologie utilisée : script en langage de programmation « R »
- Identification des patterns stylistiques (lemmatisation, étiquetage morphosyntaxique)
- Application de règles de transformation (lexicales et syntaxiques)
- Temps de traitement : 1h 2' pour un corpus de 8 336 760 mots
- Analyses particulières
 - Structures impliquant un complément du nom
 - Énumérations de noms propres
 - Précision de la date/des dates d'événements

Perspectives d'approfondissements

- Étendre aux thèses pour un corpus plus représentatif du style des chercheurs
- Différentiation de sous-corpus chronologiques, thématiques ou disciplinaires pour affiner des ensembles plus compacts ou différenciés, et leur évolution
- Analyse poussée des langages idiosyncratiques au fil du temps : évolution de la langue et des référentiels scientifiques et historiques, apparition et disparition de structures/domaines/concepts

Les mots les plus discriminants par période sur les positions de thèses de l'ENC



tf-idf

Extraire et représenter les hommes et les femmes

À partir des photos de promotion et du référentiel wikidata

- Technologies utilisées : démonstrateur GallicaPix (BaseX + REST + Xquery + IIIF), modèles de détection d'objets (Yolo v4 et MTCNN), visualisateurs (OpenSeaDragon...).
- OCR par API Google Cloud Vision
- Analyse quantitatives à partir des détections
- Tests d'alignement des éléments identifiés (visages...) aux positions et aux identifiants IdRef et VIAF
- Tests de morphing de visages (à poursuivre)

Perspectives d'approfondissements

- Elargir la base d'images de promotions ou de photos individuelles disponibles
- Extension des tests d'association éléments d'images/listes faits manuellement à un corpus plus large (par exemple indexation participative des personnes dans les photos de promotion)

De l'ensemble...



620 visages des anciens élèves de l'Ecole nationale des Chartes (1904-1933). Hackathon du bicentenaire de l'ESCS
Image affichée avec [Luminoso](#)

... aux individus



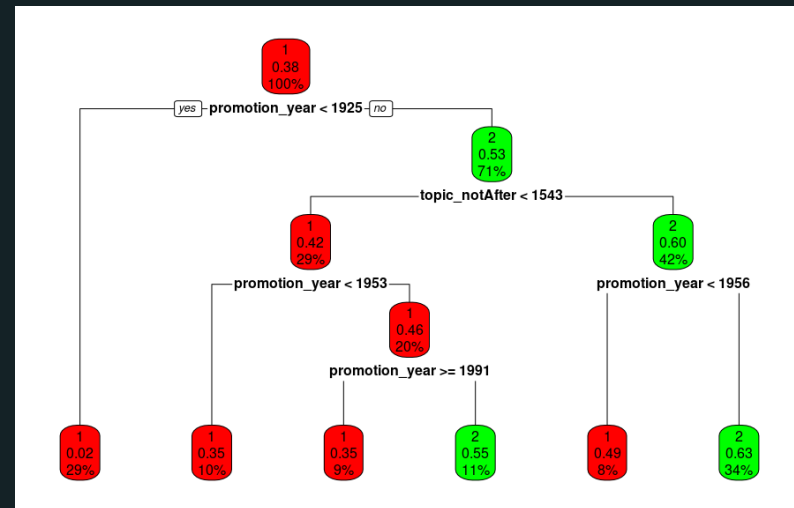
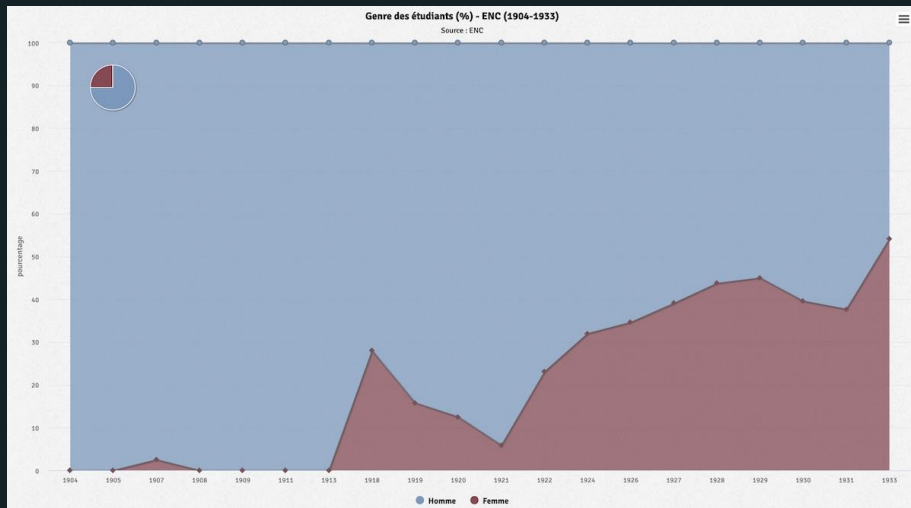
... au détails vestimentaires



Reigler Chantreac Girard Tr. den Bouchat L. Berger
Guigue Dormingham Baudry S. Gélén-Dontalis Dulong Gaucheron Vallot
Braisant

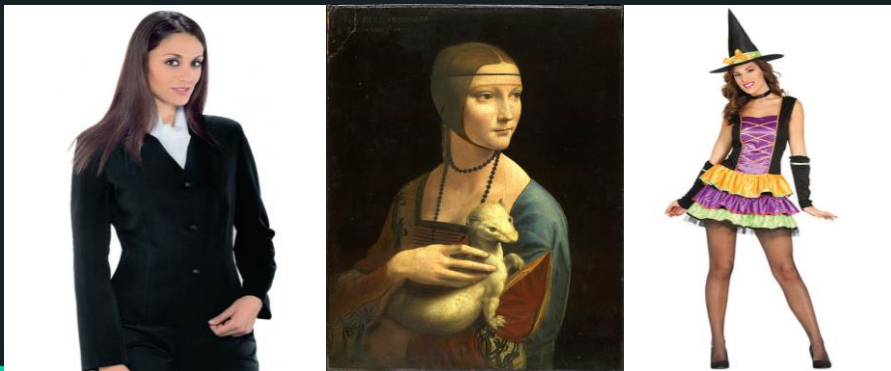


Utilisations dérivées : sociologiques...



... ou ludiques (diapos suivantes)

À quoi ressemblent l'*Homo Chartisticus* moyen, la *puella chartistica* type ?



?

IMAGE MORPHEE MOYENNE

À partir des
photographies des
promotions 1905 à
1934

(à intégrer si réussie)

?

Pipologie

Notre coup de
cœur

“J’aime beaucoup le visuel
“Cravates”

« Tiens, les chartistes du
XXI^e siècle étudient
massivement les graffitis ?”

“C’est normal le Dauphiné en
Anjou sur la carte ?”

L’organisation
par l’ADEMEC et
l’accueil par
l’École des
chartes !

Merci à toute
l’équipe et à
l’institution