BULAC

[도서관] [शिक्षक] [ဘာသာစကား] [ᠰᡳᠯᡤᡳ]

Bibliothèque universitaire
des langues et civilisations

LA GESTION DES TRANSCRIPTIONS HTR DANS OMEKA S

Conflits d'usages au sein d'une installation complexe

Anne BUGNER · anne.bugner@bulac.fr · Omeka S 4.0.4 · hébergement & administration locale sur système Ubuntu 22.04

ENS de Lyon | Journées Omeka - NumaHop | 21 novembre 2024

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC

[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

Valoriser des projets portés par la BULAC

Projet TariMa

- Juillet 2022 - Juillet 2024
- 34 textes en arabe maghrébin (XVI^e-XIX^e siècle)
- 19 manuscrits, 9 imprimés, 6 lithographies
- 8229 pages numérisées | 7821 transcrites

Projet Chi-Know-Po Corpus

- Septembre 2022 – Janvier 2025
- 12 ouvrages de la Chine classique (III^e-XI^e siècle), éditions du XVI^e au XX^e siècle, textes glosés en colonnes
- 20+ rouleaux (8 à la BULAC)
- 33 900 pages transcrites

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC

[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

Usages des transcriptions dans la BiNA

- **Diffusion / téléchargement** des fichiers xml-alto comme médias d'un Contenu

- **Visualisation en surbrillance** et en texte brut du texte transcrit

- **Indexation** par le moteur de recherche en texte intégral
  Exemple : recherche des occurrences du nom de personne " الخين لميرو"

Activer l'option dans les Paramètres généraux

Saisir le texte au curseur sur l'image

Coller le texte en barre de recherche

Résultats de la requête

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC
[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

Un affichage enrichi via le protocole IIIF

Affichage minimal: visionneuse Omeka S par défaut

Affichage enrichi: visionneuse IIIF Mirador

Protocole IIIF = structuration d'un ensemble d'informations relatives à 1 contenu (URL d'accès aux fichiers, résolution, description bibliographique et matérielle, conditions d'utilisation, annotation ajoutées…) dans 1 document, le *Manifest IIIF*, en vue de leur réutilisation par une visionneuse IIIF (ici *Mirador*).

Exemple: Contenu Omeka et Manifest IIIF

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

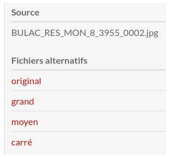
T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

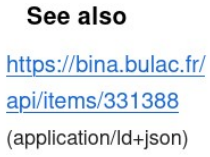
B U L A C

[도서관] [शिक्षक] [ဘာသာစကား] [ᓴᐃᓇᖅ]

Bibliothèque universitaire
des langues et civilisations

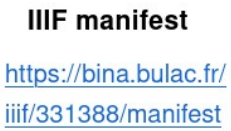
4 modules pour 1 workflow





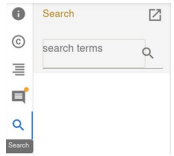
3.6.18

Image Server : **copie du fichier source en plusieurs résolutions,** assignation des fonctions (miniature, support zoom…) paramétrage avancé de l'affichage (rotation, couleur, sélection locale…), production de l'URL d'accès au fichier



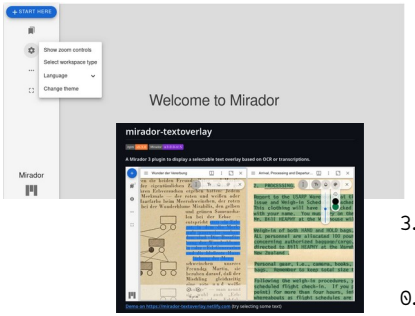
3.6.21

IIIF Server : récupération, ajout et **structuration de toutes les informations** nécessaires à la diffusion du Contenu et de ses fichiers média ; **production du manifest IIIF**



3.4.6

Iiif Search : indexation du contenu textuel du manifest IIIF dans le **moteur de recherche** de l'installation



3.4.9

0.3.8

Mirador Viewer : **fenêtre d'affichage personnalisable** du média et des informations descriptives, duplicable et ouverte au chargement de manifests extérieurs ; fonctionnalités avancées (téléchargement des ressources, annotation graphique)

Plugin *Text Overlay* de Mirador Viewer : affichage personnalisable de la transcription en **surbrillance sur l'image** et en volet latéral, sélection au curseur du texte

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

Total: 58 625 fichiers ajoutés à Omeka S (dont 58 082 pour Mirador Viewer)

BULAC

[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

De l'xml-alto au manifest IIIF (API Presentation 2)

Section d'un fichier xml-alto : première ligne de texte de l'ouvrage

```
-<Layout>
  -<Page WIDTH="982" HEIGHT="1224" PHYSICAL_IMG_NR="0" ID="page_0">
    -<PrintSpace HPOS="0" VPOS="0" WIDTH="982" HEIGHT="1224">
      -<TextBlock ID="TR_0" HPOS="93" VPOS="102" WIDTH="701" HEIGHT="916" TAGREFS="TYPE_2">
        -<Shape>
          <Polygon POINTS="93 102 794 102 794 1018 93 1018 93 102"/>
        </Shape>
        -<TextLine ID="TL_0" HPOS="261" VPOS="315" WIDTH="150" HEIGHT="59" BASELINE="262 363 412 362" TAGREFS="TYPE_1">
          -<Shape>
            <Polygon POINTS="261 315 411 315 411 374 261 374 261 315"/>
          </Shape>
          <String ID="SE_0" CONTENT="اسنطين" HPOS="273" VPOS="315" WIDTH="131" HEIGHT="59" WC="0.9015">
```

Section du manifest IIIF

```
@context:   "http://iiif.io/api/presentation/2/context.json"
@id:        "https://bina.bulac.fr/iiif/331388/annotation-page/398982/line"
@type:      "sc:AnnotationList"
resources:
  0:
    @id:        "https://bina.bulac.fr/iiif/331388/annotation-page/398982/line/l1"
    @type:      "oa:Annotation"
    motivation: "sc:painting"
    resource:
      @type:    "cnt:ContentAsText"
      format:   "text/plain"
      chars:    "اسنطين"
    on:         "https://bina.bulac.fr/iiif/331388/canvas/398982#xywh=261,315,150,59"
```

Mapping en bref

Layout = sequence
Page = canvases 0, 1, 2, 3 .....
PrintSpace/TextBlock = otherContent/sc:AnnotationList
TextLine = resources 0, 1, 2, 3 /cnt:ContentAsText
hpos, vpos, width, height = on
String/content = format/chars

Informations extraites et recomposées par le module **IIIF Server** *(/src/Iiif/AnnotationPage.php)*

```
$xpath = $imageNumber
    ? "/alto:alto/alto:Layout/alto:Page[@PHYSICAL_IMG_NR='$imageNumber']//alto:TextLine"
    : '/alto:alto/alto:Layout//alto:TextLine';
```

Bloc ciblé par le plugin **Text Overlay** :
resource/@type:cnt:ContentAsText + contenu

Requirements for supported IIIF manifests

- Line-level annotations with either one of:
  - a `motivation` that is `supplementing` (IIIF v3)
  - a resource that has a `@type` that is `cnt:contentAsText` (IIIF v2)

(Consignes d'utilisation sur le dépôt de code)

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

B U L A C

[도서관] [शिक्षक] [ဘာသာစကား] [ᏣᎳᎩ]

Bibliothèque universitaire
des langues et civilisations

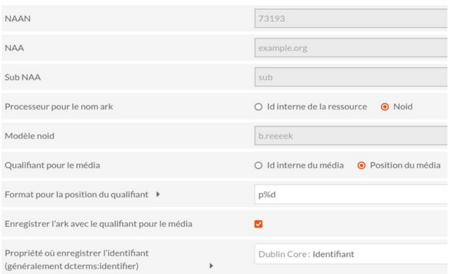
Objectif parallèle : pérenniser le signalement des contenus

Utiliser des URL aussi peu susceptibles de changer que possible ===> recours à des identifiants **uniques** et **non signifiants** (prévenir les bugs de modules / la migration des installations / les recotations / etc)

Choix des identifiants ARK : **gratuité, adaptabilité** (niveau collection, contenu, média…) et déjà largement utilisés

Attribution d'ARK dans Omeka : module *Ark*, option *Noid* pour une génération aléatoire mais non ambiguë à la lecture humaine (exclusion des voyelles, des majuscules, du chiffre 1…)

Modification de l'URL d'accès : module *Ark Url*



Ark Url

Ark Url is a companion module for Ark or Quark that replace resources URLs by their ARK URL. So instead of `/s/site/item/1` you will get `/s/site/ark:/99999/bapZs2`.

It replaces URLs for items, item sets and media. It doesn't change anything on the admin side.

Total : 158 fichiers ajoutés à Omeka + 1 librairie

Modèle d'un identifiant ARK : ark:/ 73193/ bmsbw4 / p1 .thumbnail

type — naan — name — qualifiants divers

Évolution de la syntaxe des URL BiNA :
bina.bulac.fr/s/tarima/item/331388 => bina.bulac.fr/tarima/BULAC_RES_MON 8_3955 => bina.bulac.fr/s/tarima/ark:/73193/bmsbw4

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC
[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]
Bibliothèque universitaire
des langues et civilisations

Favoriser le réemploi des manifests IIIF

- **Rebond depuis les autres canaux de signalement** des ressources de la BULAC : **Calames,** Nakala
  Contrainte : évolution récente, mise à jour *globale* nécessaire

dcterms:requires	dcterms:URI	https://bina.bulac.fr/iiif/ark:/73193/bmsbw4/manifest

Mention du manifest IIIF dans le dépôt des fichiers xml-alto du BULAC RES MON 8 3955 sur Nakala

	Identification		Contenu des Elments et attributs à mettre à jour dans un opnlet spécifique			
	cle	id	ribut TITLE du 1er daoloc	l'attribut HREF du 1er daoloc type	éventuel daodesc aut	aleur de l'attribut HREF du 1er daoloc type manifest_i
5	4333	Calames-20230209152457555	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/bb8hc9	ark:/73193/bb8hc9	https://bina.bulac.fr/iiif/ark:/73193/bb8hc9/manifest
6	471	Calames-201644194214231	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/b9gjf0	ark:/73193/b9gjf0	https://bina.bulac.fr/iiif/ark:/73193/b9gjf0/manifest
7	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/b9gjf0	ark:/73193/b9gjf0	https://bina.bulac.fr/iiif/ark:/73193/b9gjf0/manifest
8	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/bvdp9n	ark:/73193/bvdp9n	https://bina.bulac.fr/iiif/ark:/73193/bvdp9n/manifest
9	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/b05qrv	ark:/73193/b05qrv	https://bina.bulac.fr/iiif/ark:/73193/b05qrv/manifest
10	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/b3xsx3	ark:/73193/b3xsx3	https://bina.bulac.fr/iiif/ark:/73193/b3xsx3/manifest
11	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/b7pw3z	ark:/73193/b7pw3z	https://bina.bulac.fr/iiif/ark:/73193/b7pw3z/manifest
12	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/bcc31k	ark:/73193/bcc31k	https://bina.bulac.fr/iiif/ark:/73193/bcc31k/manifest
13	471	Calames-201644194214232	Consultable en ligne sur la BiNA	https://bina.bulac.fr/s/bina/ark:/73193/bcc31k	ark:/73193/bcc31k	https://bina.bulac.fr/iiif/ark:/73193/bcc31k/manifest

Extrait du tableur ABES-BULAC pour l'ajout en masse de liens externes dans Calames

- **Usages spontanés :** potentiellement n'importe quel Contenu

Images

BULAC and BnF Images are available through the libraries' IIIF server. For the list of IDs (images and documents), see the `list-images.tsv` file. To request an image, please use the following URL template:

BULAC library (BINA)

```
https://bina.bulac.fr/iiif/2/{image_ID}/{region}/{size}/{rotation}/{quality}.{format}
```

Consigne aux utilisateurs pour retrouver les images de travail – dataset du projet TariMa sur GitHub.com

Dans tous les cas : rendre possible la déduction de l'URL du manifest depuis l'URL du Contenu

https://bina.bulac.fr/s/tarima/ark:/73193/bmsbw4 => https://bina.bulac.fr/iiif/ark:/73193/bmsbw4/manifest

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC

[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

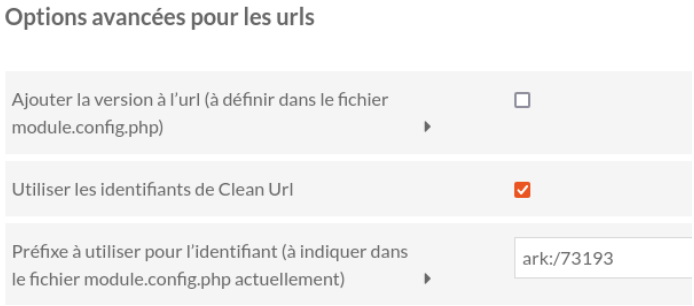
Bibliothèque universitaire
des langues et civilisations

Intégration de l'ARK dans le manifest IIIF

Module *Clean Url* : indique une **syntaxe fixe** pour la valeur de la propriété ***dcterms:identifier***, qui sera appelée **en lieu et place de l'identifiant interne** par les autres scripts dans Omeka S

Total : 69 fichiers ajoutés, dont
1 au **niveau racine** de l'application

Paramétrage dans IIIF Server



9 scripts prévoient l'usage de Clean Url.

Modèle d'un identifiant selon ces paramètres : *prefix + name*

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

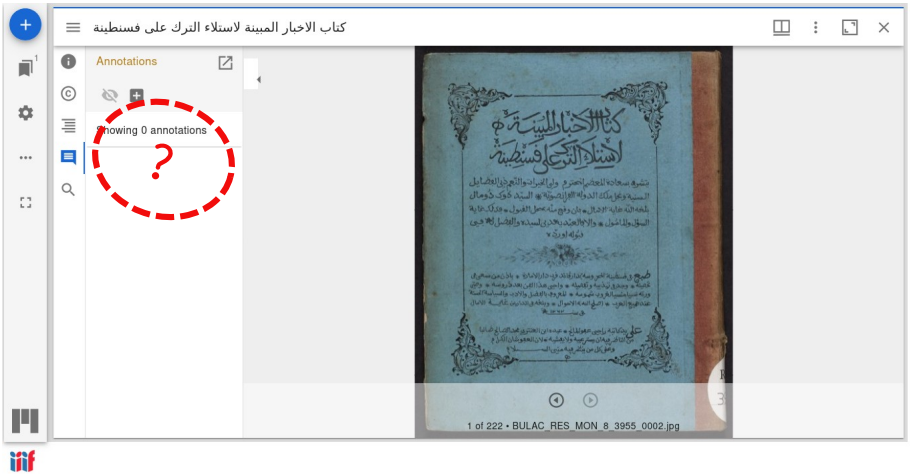
B U L A C

[도서관] [शिक्षक] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

Interactions dans IIIF

- Identifiant ARK utilisé pour l'URL du manifest (API Presentation) ☑
- Identifiant ARK non utilisé pour l'URL image (API Image) ☒
  (… aucune modification apportée à Image Server)
- Disparitions des Annotations, et donc de Text Overlay ☒



IIIF manifest

https://bina.bulac.fr/iiif/bmsbw4/manifest

See also

https://bina.bulac.fr/api/items/331388
(application/ld+json)

- Propriété *resource/"type":"cnt:ContentAsText"* non générée ☒

```
status:              "error"
message:             'Entité Omeka\\Entity\\Item avec critère {"id":"bmsbw4"} non trouvée'
reason:              "error-controller-cannot-dispatch"
display_exceptions:  true
controller:          "IiifServer\\Controller\\PresentationController"
controller_class:    null
```

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

BULAC

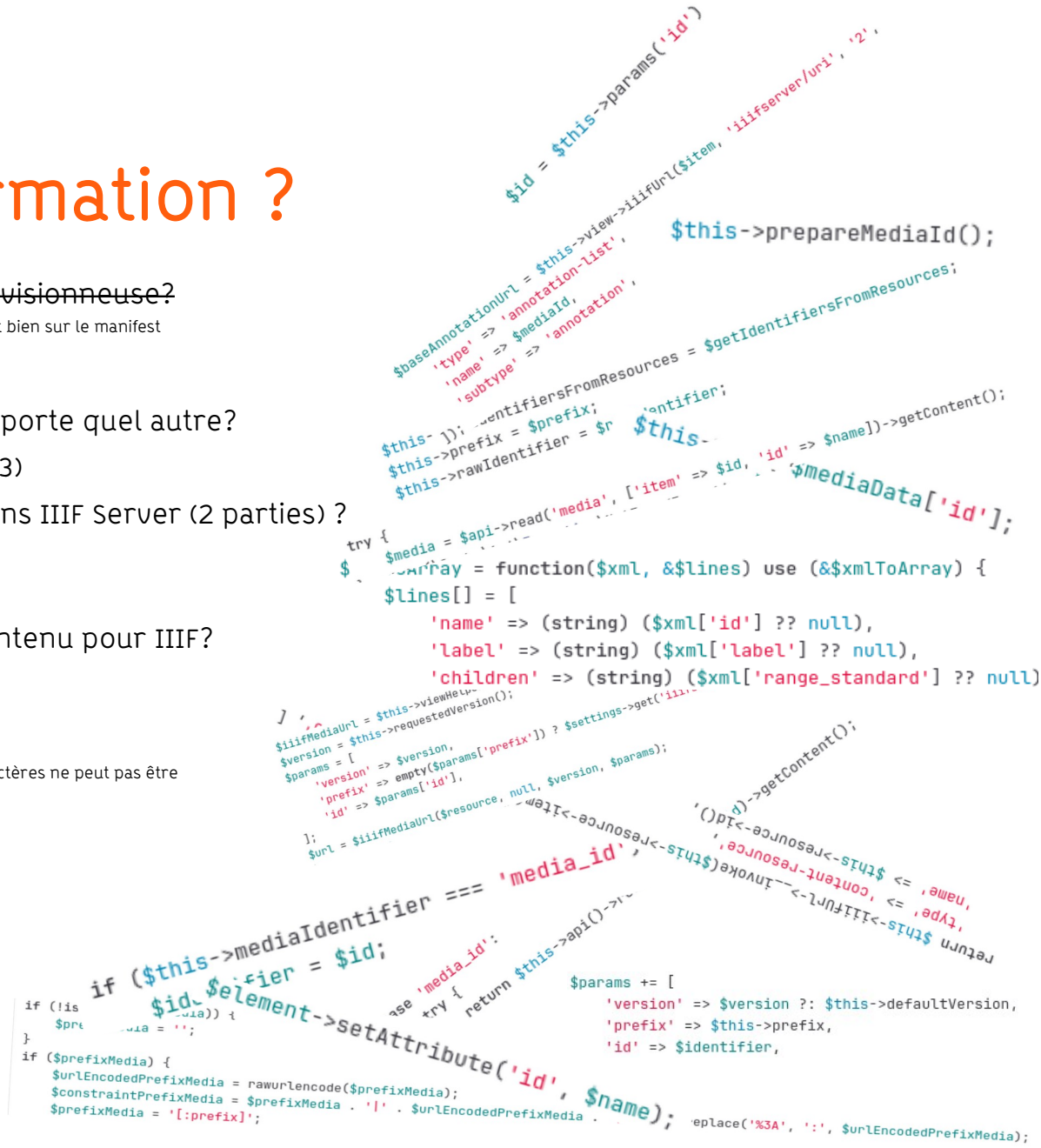
[도서관] [शिक्षक] [ဘာသာစကား] [ᓴᓇᔨᖅ]

Bibliothèque universitaire
des langues et civilisations

… Où s'est bloquée l'information ?

- Base de données Omeka ? Scripts PHP ? ~~Javascript de la visionneuse?~~
  Le message d'erreur était bien sur le manifest
- Serveur virtuel ? Serveur institutionnel? Navigateur?..
- Module IIIF Server ? Clean Url ? Ark? Common?.. Ou n'importe quel autre?
- Quel impact de la version de l'API IIIF Presentation? (2 | 3)
- Incompatibilité identifiant ARK (3 parties) // Clean Url dans IIIF Server (2 parties) ?
- Méthode d'appel d'un id non-interne manquante ?...
  - Combien de fois doit-on appeler l'identifiant d'un Contenu pour IIIF? Clean Url est-il bien convoqué pour toutes?
- Les slashs ont-ils dénaturé quelque chose ?
  - Pourquoi un *name* en chaîne de caractères ne peut pas être un *name* en chaîne de caractères ?
- Que faut-il vérifier entre l'*identifier*, l'*id*, le *prefix*, le *raw identifier* et le *name* ?
  - Les CORS?
  - Qu'est-ce qu'un *Controller*?
    - Ou était-ce le *Trait* ?
- Que choisir entre la citabilité et la performance ?

● ● ●



GIP BULAC
65 rue des Grands Moulins
F-75013 Paris

T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr

Conclusion

Tenter d'implémenter des transcriptions HTR réutilisables au sein d'un manifest IIIF à l'identifiant unique et pérenne dans Omeka S pointe **deux difficultés du modèle :**

- **Une application « clés en main », prête à la personnalisation, mais dont les outils d'enrichissement se multiplient de manière non concertée ;**
- **Une accumulation d'extensions qui peut aboutir à une dispersion de l'information.**

Comment former les gestionnaires non-développeur·euses au diagnostic des problèmes dans un environnement aussi foisonnant qu'une application Web ?

Comment améliorer la **compatibilité entre extensions,** sachant que les usages se traduisent souvent par des **combinaisons locales de versions différentes,** parfois sur des systèmes différents ?

Quelles méthodes de travail envisager au sein de la **communauté de développement,** au regard des autres impératifs comme la montée en performance ?

GIP BULAC
65 rue des Grands Moulins
F-75013 Paris
T + 33 1 81 69 18 00
contact@bulac.fr
www.bulac.fr