



**Digital Studies  
Africa  
Asia  
Middle-East**

Première école d'été du consortium Huma-Num DISTAM

## **Publier avec des données**

**5-8 juillet 2022**

### **LIEUX**

**BULAC :** Bibliothèque universitaire des langues et civilisations  
65 rue des Grands Moulins, 75013 Paris

- 5 et 6 juillet : Auditorium et foyer du Pôle des langues et civilisations (entrée sur la gauche depuis le hall d'accueil)
- 6 et 7 juillet : Salles de formation situées au rez-de-jardin des salles de lecture de la BULAC (entrée sur la droite depuis le hall d'accueil)

**Bibliothèque nationale de France. Site François-Mitterrand :**

Quai François-Mauriac, 75013 Paris  
Rendez-vous à l'entrée principale, hall Est, devant la maquette du bâtiment, pour rejoindre la salle 70 (à proximité du petit auditorium)

**Plan Libre :** 12 rue Cantagrel, 75013 Paris. Tiers-lieu situé à l'intérieur de la Cité du Refuge (bâtiment construit par Le Corbusier) : en haut à droite de l'escalier du perron, entrée au fond du hall d'accueil

**Restaurant administratif Buffon :** 17 rue Hélène Brion, 75013 Paris

**Transports en commun :** Ligne 14 et RER C : station Bibliothèque François Mitterrand. Tramway 3b : Porte de France. Bus : 62, 64, 89

**Distances :**

- BULAC ← → Restaurant Buffon : 500 m
- Restaurant Buffon → BnF : 900 m
- BULAC → Plan Libre : 250 m

### **ORGANISATEURS**

Marie Bizais (université de Strasbourg, GEO/USIAS)  
Benjamin Guichard (BULAC)  
Shueh-Ying Liao (DISTAM)  
Mercedes Volait (CNRS/InVisu)

### **INTERVENANTS**

Cécile Arènes (Sorbonne-Université)  
Marie Bizais (université de Strasbourg, GEO/USIAS)  
Ian Johnson (université de Sydney)  
Shueh-Ying Liao (DISTAM)  
Pierre Magistry (ERTIM)  
Louis Manière (géomaticien indépendant)  
Chowra Marakemi (IRIS, ERC Off-Site)  
Emmanuelle Morlock (HISoMA, Equipex Biblissima+)  
Natalia Pashkeeva (IRIS, ERC Off-Site)  
Jean-Christophe Peyssard (MMSH, ANR Shakk)  
Jean-Luc Pinol (consortium Huma-Num Paris Time Machine)  
Guillaume Porte (université de Strasbourg)  
Cécile Shaalan (CEAlex)  
Vladimir Tybin (service du dépôt légal numérique, Bnf)  
Dominique Roux (MSH Caen)  
Mercedes Volait (CNRS/InVisu)

## PROGRAMME

### Mardi 5 juillet : Vie du consortium DISTAM

Journée réservée aux membres des équipes et des établissements fédérés par DISTAM, sur invitation

#### LIEU BULAC, auditorium

Ouvert à tous les participants à partir de 18h30

**8h45** Accueil au foyer de l'auditorium de la BULAC (café et viennoiseries). Remise des dossiers

#### 9h-12h Réunion du bureau de DISTAM (auditorium)

**9h-11h** animateurs des Groupes de travail et observateurs-relais auprès des GIS Etudes africaines, Asie et Moyen Orient et mondes musulmans

Modalités de fonctionnement, point budgétaire et programmation des activités, dont résidence numérique et prochaine École d'été

**11h-12h** Restitution de l'étude sur le logiciel Access to Memory pour la production de descriptions d'archives et de manuscrits en EAD et outils alternatifs, avec Federico Nurra (SNR INHA), Mehdi Sakatni (DigiMagh) et Bruno Morandière (ResEFE)

**REPAS** Restaurant administratif Buffon (CROUS)

17 rue Hélène Brion, 75013 Paris (sur inscription préalable)

#### 14h-18h30 Assemblée générale (auditorium)

##### 14h-16h Session 1. Restitution d'expériences et résultats des travaux des groupes de travail

- HTR : expérimentations pour la reconnaissance automatique des graphies arabes manuscrites et futurs projets, avec Antoine Perrier (IREMAM), Marie Bizais, Chahan Vidal-Gorene (Calfa), et représentant de l'ANR LiPOL
- Publication enrichie de textes : exemple d'édition en ligne à partir d'éditions chinoises xylographiées (Marie Bizais et Shueh-Ying Liao)
- Résidence numérique d'Estelle Senna : géodésie (septembre 2022) : en vue d'un guide méthodologique de géoréférencement des cartes IGN en situation coloniale (cas de l'Indochine) et établissement d'un lien avec l'IGN

**16h-16h30** Pause café

##### 16h30-18h30 Session 2. Fonctionnement et perspectives

- Actualités : relations avec l'IR\* Huma-Num; Salon Innovatives SHS; appels en cours et à venir (dont DataLab-Bnf, IR\* Huma-Num); nouveaux projets CollEx Persée et future AMI CollEx Persée
- Fonctionnement de DISTAM (calendrier, prises en charge)
- Appel à résidence numérique 2022
- Activités des groupes de travail et présentation des livrables achevés et à venir
- Calendrier et thématique de l'École d'été 2023

#### 18h30-19h30 Inauguration de l'École d'été (foyer et auditorium)

- Conférence de Jean-Luc Pinol : Le cycle de vie d'un consortium à travers l'exemple de Paris Time Machine
- À partir de 19h30 : Apéro-dinatoire (Foyer de l'auditorium)

### Mercredi 6 juillet : Publier en liant ses données

#### LIEU Salles de groupe et formation de la BULAC (salle de lecture, niveau rez-de-jardin)

Accueil dans le hall d'entrée du Pôle des langues et civilisations. Attention, compte tenu des modalités d'accès à la BULAC avant 10h, ponctualité de rigueur!

**9h-10h45** Conférence plénière : Présentation de chaînes éditoriales associant des bases de données (dont Estrades et Métopes, articulées à Heurist et BaseX), avec Guillaume Porte et Dominique Roux (Salle RJ.24)

#### 11h-12h45 Ateliers « bases de données » au choix et en parallèle (inscription dans la matinée)

- Heurist : prise en main avec Ian Johnson et Marie Bizais (gérer les entités nommées, construire une structure, éditer des fiches) (Salle RJ21)
- BaseX : niveau avancé avec Guillaume Porte et Shueh-Ying Liao (organiser, lier, mettre en forme et publier des fichiers en XML) (Salle RJ22)

**REPAS** Restaurant administratif Buffon (CROUS)

#### 14h30-16h30 Ateliers « bases de données » au choix et en parallèle (inscription dans la matinée)

- Heurist : niveau avancé avec Ian Johnson et Marie Bizais (construire une structure, modifier massivement les données) (Salle R.J21)
- BaseX : niveau basique avec Guillaume Porte et Shueh-Ying Liao (réunir des fichiers en XML en vue d'une publication, contraintes éditoriales) (Salle R.J22)

**16h30-17h** Pause café

**17h-17h45** Prospection en ville et en campagne à l'aide d'une tablette nomade avec Easytrails, avec Cécile Shaalan (salle R.J24)

#### 17h45-19h15 Réunions ouvertes des groupes de travail

- GT1 : Encodage, publication et fouille de corpus textuels en graphie non-latine (RJ.23)
- GT2 : Signalement numérique pérenne et enrichi des ressources documentaires en études aréales (RJ.20)
- GT3 : Exposition de données visuelles, spatiales, sérielles enrichies (RJ.19)

**19h30-21h** Apéritif dînatoire (Foyer de l'auditorium)

### Jeudi 7 juillet : Étudier et publier les données du web

#### LIEUX Matinée : salles de formation de la BULAC (RJ.24)

Après-midi : Bibliothèque nationale de France, site François-Mitterrand, salle 70 (hall Est)

**9h** Café d'accueil (cafétéria du Pôle des langues et civilisations : 1<sup>er</sup> étage à gauche dans le hall)

**9h30-11h** Mise en contexte et documentarisation des sources du web en contexte autoritaire : retour d'expérience de l'ERC Off-Site, avec Natalia Pashkeeva et Chowra Marakemi (Salle RJ24)

**11h-12h30** Archivage des données du web : sauvegarde et citabilité pour la recherche, avec Jean-Christophe Peyssard (salle RJ24)

**REPAS** Restaurant administratif Buffon (CROUS)

#### Après-midi : Bibliothèque nationale de France, site François-Mitterrand

Rendez-vous à l'entrée principale, hall Est à partir de 14h15, devant la maquette du bâtiment, pour rejoindre la salle 70 (à proximité du petit auditorium)

**14h30-17h** Le dépôt légal de l'internet et les Archives de l'internet : collecte, modalités de consultation et collaborations avec les chercheurs (services, projets & Datalab), avec Vladimir Tybin

**17h-18h** Atelier dans les archives de l'internet : l'exemple du site Égypte d'antan, avec Mercedes Volait; visite guidée du BnF DataLab

### Vendredi 8 juillet : Gérer et rendre ses données publiques

#### LIEU Plan Libre. Cité du Refuge, 12 rue Cantagrel, 75013 Paris

**9h** Accueil

**9h30-10h45** Conférence plénière : « Les data papers : un retour d'expérience », avec Louis Manière

**11h-12h30** La publication des données dans un projet d'infrastructure numérique multipolaire : le rôle du plan de gestion de données, avec Emmanuelle Morlock

**REPAS** Déjeuner sur place (sur inscription préalable)

#### Après-midi : Publier des données visibles et identifiables

**14h-15h30** Le plan de gestion de données dans le cycle de vie d'un projet financé, avec Cécile Arènes

**16h-17h30** La pérennité des données, avec Pierre Magistry

**17h30-19h** Goûter conclusif

Prochaine école d'été : 4 au 7 juillet 2023, à l'université de Strasbourg