



HAL
open science

Un projet expérimental à la Maison Asie Pacifique (Marseille): Online digital sources and annotation system for the social sciences

Laurent Dousset

► To cite this version:

Laurent Dousset. Un projet expérimental à la Maison Asie Pacifique (Marseille): Online digital sources and annotation system for the social sciences. Collecter, organiser, valoriser les archives de la recherche en sciences humaines et sociales, 18-19 Septembre, 2008, Aix-en-Provence, France. hal-00351609

HAL Id: hal-00351609

<https://hal.science/hal-00351609v1>

Submitted on 9 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ODSAS ¹

Online digital sources and annotation system for the social sciences

Sources numériques et système d'annotation en ligne pour les sciences sociales

Laurent DOUSSET

CREDO (UMR 6574)

Centre de Recherche et de Documentation sur l'Océanie, Marseille

<http://www.pacific-credo.fr>

Une version plus courte de cet exposé fut présentée dans le cadre du colloque *Collecter, organiser, valoriser les archives de la recherche en sciences humaines et sociales*, MMSH Aix en Provence, 18 et 19 Septembre 2008.



L'ambition du projet ODSAS est de proposer une plateforme de sauvegarde et d'archivage multifonctionnelle des données de terrain que rapportent les chercheurs, en particulier les anthropologues, de leurs terrains. Cette ambition répond à une demande forte ressentie parmi les anthropologues de notre centre. Trop de données ethnologiques finissent inexploitées et souvent inconnues dans les « cartons », car les chercheurs n'emploient que rarement la totalité de leurs données dans leurs analyses et leurs publications. Les publications n'offrent guère la possibilité de reproduire, et donc de pérenniser les données brutes, mais se limitent aux analyses, aux conclusions et aux propositions théoriques. Une grande quantité de données originales sont ainsi souvent perdues pour les générations de chercheurs à venir. Mais l'ambition d'offrir une plateforme de sauvegarde des données ethnologiques répond également à une transformation du paysage politique de la recherche ethnologique. Les groupes et sociétés dans lesquels les anthropologues enquêtent développent une prise de conscience considérable par rapport aux données culturelles que les chercheurs recueillent et stockent. Il s'agit de la portée que les données culturelles peuvent exercer sur l'existence politique et économique des groupes concernés. Les données ethnographiques deviennent des

¹ Avec mes remerciements à Pascale Bonnemère, Louise Bertaux-Pichard et Pierre Lemonnier pour leurs commentaires au sujet de ce texte.

preuves potentielles de l'existence différentielle de sociétés ou sous-sociétés au sein d'Etats et d'Etats-Nations qui paraissent homogènes mais qui se construisent sur un arrière plan hétérogène. Dans le contexte des rapports de force entre les administrations centrales et étatiques et les communautés locales dans lesquelles les anthropologues enquêtent, les données ethnographiques deviennent un argument potentiel mis au service d'une exigence de la différence culturelle reconnue et légitimée. Il s'agit, pour les communautés locales et parfois les administrations centrales, de mettre l'ethnologie au service de l'aspiration à l'autonomie politique des communautés. Le développement d'une plateforme technologique sur laquelle des données ethnologiques brutes peuvent être sauvegardées, et donc réutilisées dans d'autres contextes que ceux de la science, participe ainsi au rapatriement de ces données vers les groupes et sociétés dont elles proviennent ; rapatriement que ces groupes et sociétés exigent d'une manière de plus en plus soutenue.

L'archivage numérique : besoins et problèmes

Le développement d'une telle plateforme, et l'archivage de matériel ethnographique en général, rencontre toutefois des problèmes considérables. Ces difficultés doivent être surmontées afin de rendre un tel projet viable et pérenne. Le premier problème, le plus considérable aussi, est celui de la confidentialité et de la « sensibilité » des données. Les matériaux ethnographiques bruts contiennent pour leur grande partie des informations personnelles. Ils mentionnent et citent des individus particuliers, identifiables et localisables, avec leurs discours, leurs opinions, leurs défauts et leurs côtés positifs. Ils rapportent les histoires de vie des acteurs, font le détail des conflits et des alliances qui les lient à leurs communautés, ou énumèrent et chiffrent les échanges commerciaux qu'ils pratiquent. L'archivage, et surtout la mise à disposition de ces données, ne peuvent être considérées comme des procédures simples, sans enjeux. La divulgation de ces informations risque de mettre les individus en danger réel, danger social ou même de vie. Elle risque également de perturber la confiance établie entre le chercheur et les personnes qui se livrent et qui livrent ainsi des informations qui les concernent personnellement. Une plateforme archivant de telles données ne peut faire l'impasse sur ces problèmes de confidentialité et doit prendre en compte la nécessaire protection des personnes.

Le problème est similaire pour ce que nous appelons la « sensibilité » des données. Nous employons ici une traduction littérale du terme anglais, « sensitive information », qui est devenu monnaie courante pour exprimer ces difficultés. Archiver et rendre public des informations sont des procédures qui peuvent nuire aux ambitions politiques et économiques de la communauté hôte lorsque ces données deviennent accessibles à des opposants. Les Etats-Nations et leurs administrations, mais aussi d'autres groupes locaux rivaux, peuvent s'emparer de ces données, les analyser à leur manière et enfin les mettre à leur propre service dans une ambition d'anéantissement des intentions d'autonomie des communautés. En cherchant et en soulignant des contradictions internes à la société ou à certaines déclarations individuelles qui surviendraient dans les données ethnographiques elles-mêmes, les groupes opposés mettent ainsi indirectement l'anthropologue au service de conflits locaux, régionaux et même nationaux. Ces processus ne créent pas seulement un danger potentiel mais ont été observés et vécus par nous-mêmes dans de nombreux cas où les données sociologiques furent trop librement mises en circulation. Archiver – et donc d'une manière ou d'une autre rendre un jour public – des matériaux ethnographiques, c'est ainsi avoir des responsabilités considérables. Celles sur la confidentialité des données personnelles et la protection des individus, et celles qui concernent plus globalement l'autonomie et le développement désiré par les communautés hôtes. N'oublions pas que, au contraire des historiens (ou du moins d'une grande partie d'entre eux), les sujets sur lesquels les anthropologues enquêtent sont contemporains. Une plateforme technologique ne peut faire l'impasse sur ces difficultés.

D'autres problèmes, plus techniques ou mécaniques, doivent également être résolus avant de pouvoir créer une plateforme d'archivage ethnologique. Le premier concerne la faisabilité du travail d'archivage lui-même. Afin que des données ethnographiques soient compréhensibles et exploitables, aujourd'hui comme demain, l'archivage ne peut être entrepris que par le chercheur lui-même, celui qui a rassemblé les données et qui les a rapportées sous une forme ou une autre. Il n'y a guère opération plus difficile que celle d'essayer de lire, de comprendre et d'exploiter les carnets de terrains, par exemple, d'un collègue anthropologue. Chacun possède son codage, chacun rédige à sa manière des phrases ou des mots qui, à l'évocation, lui permettent de reconstruire des observations dont il n'était pas possible de noter l'intégralité *in situ*. La mise en contexte, la rédaction de légendes aux images et d'explications aux graphiques, le création de l'ordre chronologique des documents et des données, l'identification des documents et leur mise en relation nécessitent des

connaissances fines sur le groupe hôte, le lieu de l'enquête et les processus qui ont guidé l'enregistrement de ces données. Une personne tierce ne pourra que difficilement accomplir cette tâche de manière satisfaisante. Bref, il est préférable, sinon nécessaire, que le chercheur puisse lui-même archiver ses propres données. Or, le travail du chercheur n'est pas celui d'archiver pour les générations futures les données qu'il a récoltées, mais de les exploiter, de les analyser et d'en publier les conclusions. Faute de temps – c'est aussi simple que ça –, le chercheur donnera la priorité à l'analyse, à la théorisation et à la publication de ses données plutôt qu'à l'activité qui consiste à les préparer pour les générations futures.

Autre problème, matériel encore, est celui de l'hétérogénéité des matériaux. Les anthropologues rapportent de leurs terrains des données dont les supports sont disparates, car cette discipline s'intéresse moins au support des données sociales qu'à leur contenu qualitatif, et ce contenu qualitatif mérite d'être enregistré sous ses diverses formes d'apparition. Images numériques ou papier, films, enregistrements sonores, notes manuscrites, notes tapées et électroniques, copies d'archives, coupures de journaux, d'ouvrages, de catalogues, cartes surdimensionnées, croquis et plans etc. sont des exemples de support que l'ethnologue récolte et utilise dans ses analyses et qui peuvent tous renvoyer à un et même phénomène social que l'anthropologue analyse. Tous font partie de ces données qui, pour être exploitables dans le futur, doivent être archivées ensemble, en relation les unes avec les autres, sur une même plateforme, cohabitant de manière connectée et sans rupture. Un rituel, pour ne citer qu'un exemple, est décrit (notes écrites), filmé, enregistré (extraits sonores). L'espace sur lequel il a lieu fait l'œuvre d'un croquis et des parties ou objets particuliers sont photographiés. Afin de le replacer dans son contexte historique et culturel, ce rituel doit être documenté par des archives et par des analyses anthropologiques précédentes. Enfin, les acteurs du rituel, afin de mieux comprendre leurs fonctions, doivent être décrits dans le cadre de leur champ social par le biais de généalogies, par exemple. Nous voyons donc que le rituel en question ne peut être convenablement documenté et analysé que si les processus permettent une multiplicité des modes d'enregistrement, de support et de restitution.

Avec ODSAS, nous tentons de répondre au mieux à ces différents problèmes en développant une plateforme qui, tout en archivant les données, est orientée vers les responsabilités et ambitions des chercheurs et de leurs groupes hôtes plutôt que vers une logique d'archivage et d'identification documentaire neutre et objective. Les réponses que

nous apportons à ces différents problèmes sont relativement simples. En voici, brièvement, les principes.

Au dépositaire, qui est dans la grande majorité des cas également l'anthropologue qui a récolté les données, est attribué un espace privatif sur la plateforme, dans lequel il gère lui-même et sans intervention extérieure autre que technique si nécessaire, ses archives. Il y gère lui-même les autorisations d'accès qu'il pense possible ou utile à accorder à des tierces personnes. En d'autres termes, le dépositaire en reste le propriétaire légal, intellectuel et matériel. C'est à lui que revient la responsabilité et le droit de juger quelles sont les données et documents qui peuvent, ou qui, au contraire, ne doivent pas, être rendus publics. C'est à lui d'estimer lesquelles de ses données sont sensibles ou confidentielles ou, au contraire, lesquelles peuvent être accessibles sans causer de préjudice. C'est à lui également qu'incombe la tâche de réévaluer ses choix en fonction de situations qui évoluent, et de rendre public des documents qui ne l'étaient pas auparavant ou, au contraire, de limiter l'accès à des documents qui jusque-là étaient publics. Il peut ainsi, au sein de sa ou ses collections, créer des ensembles de documents qu'il jugera privés et confidentiels, d'autres pour lesquels il pourra attribuer des droits d'accès limités, par exemple à d'autres chercheurs ou à la communauté hôte, et d'autres encore qu'il rendra totalement publics. La conséquence de ce droit absolu et irrévocable du chercheur par rapport à son propre matériel est une frustration indéniable du visiteur public par rapport à ces collections et archives en ligne. Les collections sont annoncées, le visiteur sait qu'elles existent, mais il ne peut y accéder sauf autorisation explicite du propriétaire. Cette frustration, compréhensible, est la conséquence immédiate et inévitable d'un compromis entre nécessité d'archivage et nécessité de protection et de confidentialité des données et des acteurs.

Afin de rendre l'archivage effectif et productif, et afin que le chercheur s'engage dans cette procédure qu'il considère par ailleurs être une perte de temps par rapport aux définitions premières de son activité, il est nécessaire que cet archivage, et donc la plateforme que nous développons, apporte une réelle plus-value scientifique. Il est nécessaire que l'investissement en temps et les efforts occasionnés par l'archivage et le renseignement des données archivées sur une plateforme numérique entraîne par ailleurs un gain d'efficacité dans le domaine de l'analyse, de l'exploitation et de la publication des données. L'archivage ne doit pas constituer une procédure dont la finalité est l'archivage lui-même. Il faut donc que

la plateforme propose des outils de manipulation, d'organisation, de regroupement, de mise en relation des objets qui ne sont pas aisément faisables par le chercheur lui-même. Il est nécessaire que la plateforme reflète une pré-analyse scientifique, ou propose le support de cette pré-analyse, tout en répondant aux interrogations personnelles (et imprévisibles) du chercheur ; et sans que ces manipulations modifient pour autant la souche brute des données, et donc l'archive en elle-même. Ces manipulations sont ce que nous appelons des annotations, domaine sur lequel nous reviendrons plus loin. Notons pour le moment que l'annotation est une information qui constitue une couche sémantique à l'objet en question sans remplacer ou modifier l'objet original. Notons aussi que chaque objet peut être associé à diverses annotations, à diverses couches sémantiques, et que chaque annotation peut elle-même devenir un objet source qui peut lui-même être annoté. Nous y reviendrons.

Afin de rendre l'archivage « complet » et relationnel – au sens où c'est la relation entre les informations qui procure du sens à celles-ci –, il fallait que la plateforme puisse gérer de manière transparente et sans trop diversifier l'expérience de navigation de l'utilisateur d'un plan à un autre une interface identique pour tout type de support d'information. Films, sons, images, graphiques ou textes sont traités de manière différentielle par la plateforme en arrière-plan, mais sont rassemblés à un niveau identique dans le contexte de son interface avec l'utilisateur, le chercheur en particulier. Enfin, dernier point, il fallait que la plateforme puisse également fonctionner comme un outil de mutualisation des efforts. L'anthropologie moderne est une science en réseau et de réseau. Ces réseaux scientifiques, d'échanges et de collaborations sont nationaux et internationaux. Le nombre de chercheurs qui travaillent en collaboration avec des collègues à l'autre bout du monde est croissant. Il fallait donc que la plateforme puisse proposer un outil de travail personnalisable à distance, mutualisé, dans lequel des chercheurs même distants puissent collaborer, discuter, manipuler et organiser le matériel sans que la plateforme perde son objectif premier, celui d'être un lieu d'archivage de données ethnologiques. Ici encore, ces collaborations doivent pouvoir s'exprimer sans modifier la source brute des données et tout en conservant les travaux qui résultent de ces collaborations sous forme de couches sémantiques qui enrichissent le sens donné à un objet particulier, à un moment donné par des chercheurs donnés.

Réalisation d'ODSAS (version bêta)

Le fait que le titre de cette sous-partie de notre exposé contient la mention « version bêta » est significatif. Il existe deux types, deux modèles de développement informatique et donc de développement d'une plateforme telle que nous l'envisagions. Ces modèles ne se retrouvent que rarement dans leur état pur et la réalité est toujours une combinaison de leurs principes. Mais la tendance vers l'un ou l'autre de ces modèles se fait ressentir dès les toutes premières phases des projets et se maintient en arrière-plan tout au long de leur durée de vie.

Le premier modèle, que nous appellerons « planificateur », se fonde sur une pré-étude, une étude de faisabilité, l'articulation et la rédaction des besoins et des résultats escomptés. Les premières étapes de ce mode de mise en place des outils informatiques se construisent sur papier, par la réflexion et la discussion, avant d'écrire les premières lignes de code. Le second modèle, que nous appellerons « évolutif », mais que l'on pourrait aussi appeler « intuitif », commence au contraire par des lignes de codes à partir d'une idée brute et approximative. Les discussions et réflexions préalables sont limitées. Le programmeur crée du code et des ébauches de plateformes sur lesquelles les utilisateurs accèdent dès les premiers jours d'existence du projet. Ils renvoient leurs critiques, leurs besoins, leurs idées au programmeur qui adapte la plateforme au fur et à mesure des problèmes, questions et besoins. La conception et la programmation sont donc simultanées, et la conception est le produit de l'utilisation par des utilisateurs « tests ». La plateforme est ainsi évolutive dès sa naissance. A aucun moment du développement elle est un produit définitif, même si elle est déjà disponible et accessible.

Les deux modèles ont des avantages et des inconvénients. Le modèle planificateur est plus professionnel en ce sens qu'il fait l'état des besoins à un moment donné et propose des solutions envisageables. Même si les discussions et études préliminaires prennent parfois un temps considérable, le développement de la plateforme elle-même est normalement plus rapide et aboutit d'emblée à un résultat tangible et exploitable. Il se fonde sur un cahier des charges contre lequel le résultat peut être évalué. La visibilité publique du résultat entraîne d'emblée, dès sa mise en ligne, une satisfaction considérable, même si sa mise en ligne est plus tardive qu'avec le modèle « évolutif ». Le produit, lorsqu'il est visible, est un produit fini

et utilisable. Ce type de modèle permet également d'intégrer un nombre de participants considérable dans la pré-étude et la conception de l'outil. Les inconvénients de ce modèle sont sa rigidité et l'absence d'interaction continue entre le groupe de concepteur, les utilisateurs et la mise en place ou programmation. Une fois le projet défini sur papier, les acteurs impliqués dans les discussions et les études préliminaires sont déconnectés de la réalisation de la plateforme à proprement parler. Les échanges entre concepteurs et développeurs sont souvent minimales. La seule référence pour les développeurs est le cahier des charges. Or, souvent, les idées et les besoins potentiels surviennent lorsque le projet évolue de manière palpable et utilisable. Le modèle planificateur, aussi économique et professionnel qu'il puisse paraître, produit souvent une plateforme qui ne satisfait pas totalement car rigide et peu évolutif. Le risque est qu'il est rapidement dépassé et que sa durée de vie est donc courte.

A l'inverse, le modèle « évolutif » ou « intuitif » est à la fois plus chaotique et déconcertant, mais aussi plus adaptatif et plus satisfaisant à terme. Il permet un échange et une réévaluation permanente entre les besoins, les idées, les nécessités et les contraintes de leurs réalisations. Il ne vit que par le travail simultané et constant entre utilisateurs « test » et développeurs. La distinction entre concepteur et développeur n'est plus une distinction dans le temps ; elle n'est pas une distinction entre les étapes de la réalisation du projet, mais devient une distinction dans la nature de l'implication et du point de vue des acteurs. Les avantages sont considérables car ce modèle permet une réactivité immédiate et une satisfaction rapide des visions et besoins des utilisateurs. Ce modèle permet également une mise à disposition rapide et crée immédiatement la preuve de l'existence publique du projet. Il permet aussi de démarrer les travaux et le développement sans devoir être certain d'avoir à disposition toutes les ressources financières, logistiques et humaines pour l'accomplir, puisqu'il peut se permettre d'aller aussi loin qu'il pourra sans risque de faire disparaître la totalité du projet : ce qui aura été mis à l'œuvre sera disponible, même si ce qui est disponible n'est pas tout ce qui avait été envisagé. Les inconvénients sont aussi considérables que les avantages. Le travail de développement est accru. Le programmeur devra adapter, reprendre et modifier des fonctions et procédures logicielles à plusieurs reprises. Les aspects visuels et les éléments de l'interface qui a un moment furent pensés efficaces s'avèrent à des étapes ultérieures peu ergonomiques ou même logiques et doivent donc souvent être entièrement reconçues. Ce modèle nécessite ainsi des échanges constants entre concepteurs, utilisateurs et développeurs, et donc une

disponibilité certes intermittente mais longue dans le temps. Il ne peut fonctionner convenablement que si les relations humaines entre les concepteurs, les utilisateurs et les développeurs sont productives, et si les uns et les autres comprennent les contraintes liées à la fois au développement technique de la plateforme et aux exigences scientifiques qui en sont la finalité.

Le modèle que nous avons choisi pour développer ODSAS, que nous avons en quelque sorte été contraints de choisir pour des raisons diverses sur lesquelles il n'est pas nécessaire de revenir ici, est ce second modèle « évolutif ». Le développement à proprement parler est plus long, mais en concertation avec l'évolution des besoins des utilisateurs premiers, c'est-à-dire les anthropologues qui ont accepté d'archiver dès les premiers jours d'existence du projet leur matériel. Il est par contre plus réactif et dès les premiers jours de travail sur le projet, il fut visible et accessible sur le réseau. C'est donc la raison principale pour laquelle la mention « version bêta » accompagne la description de l'état de développement de notre plateforme. Elle est disponible en ligne, elle est accessible, elle propose déjà un grand nombre de fonctions, mais elle ne peut nullement être comprise comme un produit fini, correspondant à un état des lieux que nous aurions décidé en amont du projet. La plateforme évolue tous les jours.

A la réalisation de la plateforme ont participé diverses personnes à des niveaux différents. Nous distinguons deux projets, même s'ils constituent deux facettes de la même ambition. Soulignons que les deux projets sont co-financés par le TGE Adonis. Le premier porte lui-même le nom ODSAS et concerne le développement informatique à proprement parler. Ce projet est encadré par le CREDO (Centre de Recherche et de Documentation sur l'Océanie). L'auteur de cet exposé est aussi le responsable et le programmeur de cet aspect de notre plateforme. Le second volet, nommé ADAP (Archives Digitales Asie-Pacifique), concerne plus particulièrement la numérisation du matériel et la collaboration des chercheurs dans la conception et le développement de la plateforme. ADAP est encadré par la MAP (Maison Asie Pacifique) avec les participants principaux suivants :

Louise Bertaux-Pichard (MAP, collections Asie)

Pascale Bonnemère (CREDO, collections PNG)

Lorenzo Brutti (CREDO, collections PNG et co-responsable ADAP)

Laurent Dousset (CREDO, collections Australie et co-responsable ADAP)

Maurice Godelier (CREDO, collections PNG)

Pierre-Léonce Jordan (CREDO, collections filmiques)

Pierre Lemonnier (CREDO, collections PNG)

Helena Meininger (CREDO, collection Equateur)

Serge Tcherkézoff (CREDO, collections cartes)

Marc Tabani (CREDO, collection Vanuatu) nous a rejoint il y a quelques semaines.

Dans le cadre d'ADAP, près de 30 000 objets ont été numérisés depuis février 2008, dont la grande majorité est disponible en ligne sur la plateforme (<http://odsas.pacific-credo.fr>). Des dizaines d'heures de films ethnographiques ont également été numérisés et seront rendus disponibles dès le premier trimestre 2009.

Les fonctionnalités existantes

Cette partie de notre exposé liste les fonctionnalités disponibles dans la version bêta qui se trouve en ligne. Elle fait également allusion aux fonctionnalités qui sont en cours de préparation. Nous commencerons par illustrer la structure générale de la plateforme.

1) Structure générale

La structure générale de la plateforme est composée de trois parties. La première est le stockage brut des objets originaux. Sous objet original doit être entendu tout fichier numérique non-modifié de sa source et comprenant des images, films, sons et fichiers textuels sauvegardés sur le serveur. Chaque objet reçoit un nom unique et est stocké dans une organisation hiérarchique perméable. La perméabilité signifie ici que le matériel est physiquement organisé et regroupé selon des collections (par exemple Australie, Viet Nam etc.) qui contiennent elles-mêmes des ensembles (tel ouvrage, telle expédition etc.), qui eux-mêmes contiennent les objets originaux. La hiérarchie est perméable car tout objet d'un ensemble ou d'une collection peut-être référencé dans un autre ensemble ou dans une autre collection sans pour autant modifier son emplacement physique sur le serveur. La hiérarchie

répond donc à une organisation de stockage physique, et non pas à une organisation sémantique, logique ou historique du matériel.

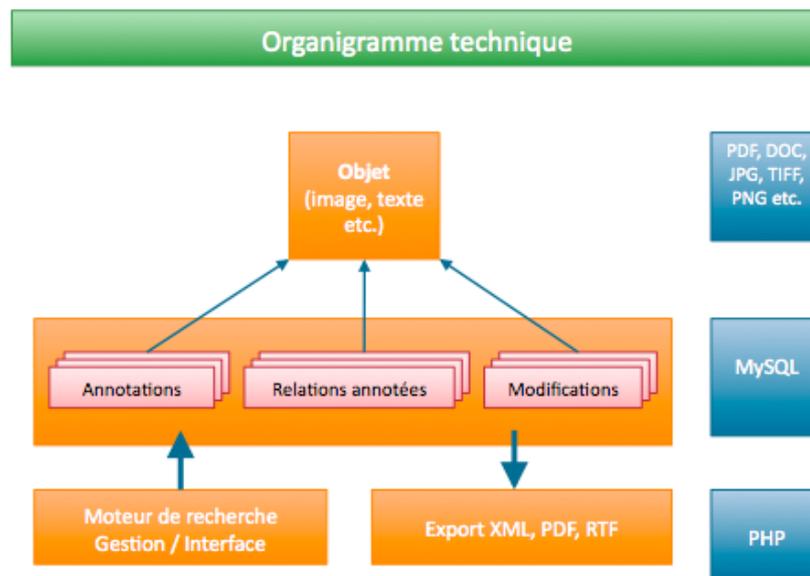


Figure 1 : structure générale de la plateforme avec ses trois parties : les objets physiques, les annotations (MySQL) et les modules d'interface (PHP)

La seconde partie est la base de données, dans notre cas MySQL. Elle comporte un certain nombre de tables qui enregistrent les annotations concernant les objets originaux. En d'autres termes, c'est dans une base de données relationnelles que la sémantique contextualisant les objets est stockée. Ces annotations concernent les données suivantes :

- *Les données descripteurs de base de l'objet* : il s'agit de renseignements primitifs, comme le lieu, la date, l'identité, la localisation physique, la localisation géographique etc. de l'objet en question. La structure de ces informations répond aux méta-tags du Dublin Core. Ces descripteurs primitifs peuvent être édités par le propriétaire de l'objet ou par ceux auxquels le propriétaire aurait donné un accès total à son matériel. Ces champs descripteurs contiennent également les coordonnées géographiques de sorte à pouvoir localiser l'objet ou l'ensemble d'objets sur une carte satellitaire via une interface avec GoogleEarth/GoogleMap.
- *Les annotations à proprement parler* : chaque utilisateur qui en a l'autorisation peut

ajouter des informations et descripteurs libres qui intéressent l'objet en question. Ces annotations peuvent être publiques : tous ceux qui ont accès à l'objet peuvent les lire et y répondre, ou y ajouter une couche sémantique supplémentaire en rectifiant (sans supprimer les version précédentes, les informations, par exemple) ; ou elles peuvent être privées, et dans ce cas seul l'auteur peut visionner, chercher et retrouver les objets en fonction de ses annotations.

Ajouter une annotation au document: Doc: 23932 / Set: 213 / Coll: 13

Si vous n'êtes pas l'auteur de cette annotation, merci d'ajouter son nom ici:

Les danseurs tournent autour des tambours centraux en secouant leurs pieds, sur lesquels sont encrochés des bracelets de graines musicales.

Visible par d'autres: Moi seulement:

Figure 2 : Mise à jour des annotations (extrait d'écran). L'auteur peut décider si cette annotation est une note personnelle ou si elle peut publiquement accompagner l'objet.

Id	Doc id	Author	Year	Title	Doc identification	location	
247	16664	Pascale Bonnemère	1994	Idzi beri	:	Ikundi	
248	16665	Pascale Bonnemère	1993	Idzadze wiew et sa maman	:	Ayakupna'wa	
249	16666	Pascale Bonnemère	1993	Idzadze wiew	:	Ayakupna'wa	

Figure 3 : Dès que des descripteurs de base ont été renseignés par le propriétaire/utilisateur de l'objet ou de l'ensemble, une table des matières devient disponible pour l'ensemble en question.

- *Les annotations géométriques* : nous appelons annotation géométrique toute information que l'utilisateur crée par la simple utilisation de l'interface. Il s'agit de l'historique des modifications graphiques de la représentation des objets dans l'interface. Nous considérons que les manipulations graphiques d'un objet sont en elles-mêmes des annotations à partir du moment où elles sont identifiables (qui a fait quoi avec quel objet à quel moment ?). Si l'utilisateur crée une rotation de l'objet, s'il agrandit l'objet, s'il crée un négatif/positif de l'objet etc., ces manipulations sont enregistrées et participent à la création d'un environnement sémantique qui accompagne l'objet lui-même.

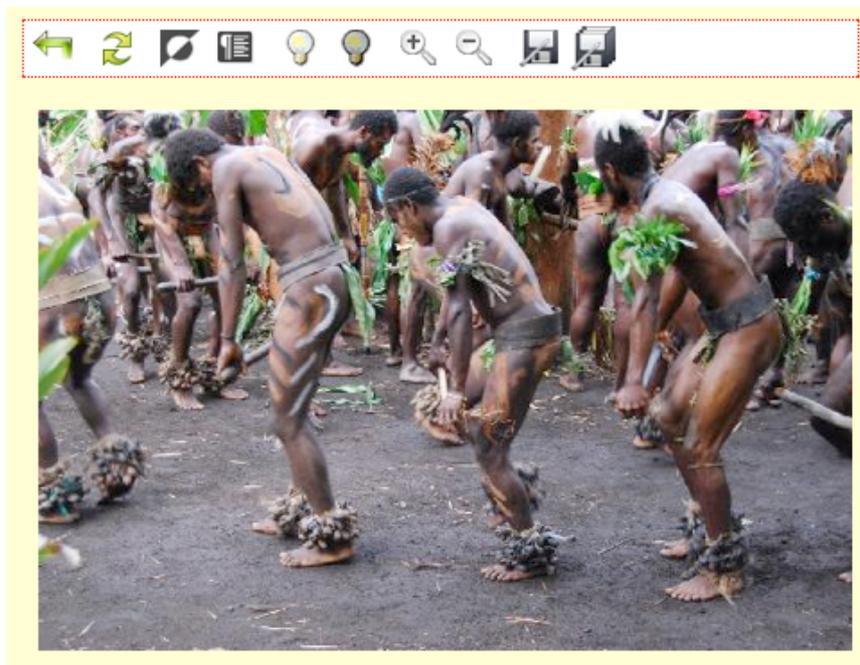


Figure 4 : Les fonctions de manipulation géométrique de l'objet (extrait d'écran). De gauche à droite : revenir à l'image d'origine, rotation, noir et blanc, négatif/positif, plus clair, plus sombre, plus grand, plus petit, enregistrer cette configuration, enregistrer toutes les images de cet ensemble selon cette configuration. D'autres annotations géométriques sont en cours de préparation (Photographie Laurent Dousset, Lamap [Malekula], Vanuatu, 2008).

- *Les annotations relationnelles* : chaque objet peut être relié à n'importe quel autre

objet, et cette relation peut être accompagnée d'une annotation sémantique. La relation entre deux objets se fonde sur un descripteur de lien qui illustre une sémantique relationnelle et qui enrichit le contexte de chacun des objets reliés. Ici encore, ces liens peuvent être publics ou privés. S'ils sont privés, seul l'auteur du lien les apercevra (les autres utilisateurs verront l'objet sans qu'il y ait la relation en question). La mise en relation des objets permet une navigation facile et personnalisable entre les objets, les ensembles et les collections ; navigation entièrement créée par l'utilisateur pour ses propres besoins ou pour un public plus général, selon son choix.

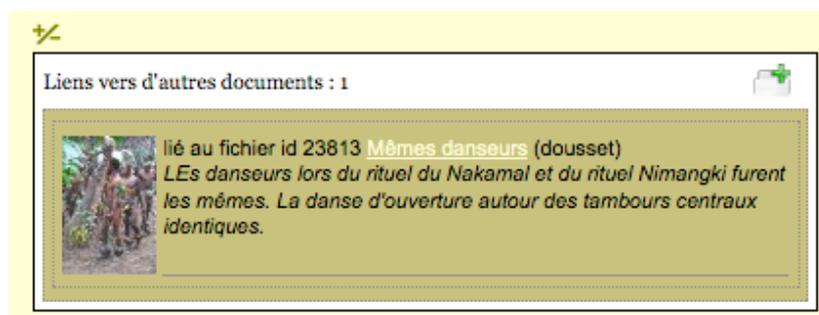


Figure 5 : Deux fichiers ont été reliés (extrait d'écran). Cliquer sur l'image ou le titre de l'annotation liante permet de naviguer d'un objet à l'autre quelqu'en soit la collection ou l'ensemble. Cliquer sur l'onglet à droite permet d'ajouter une nouvelle relation vers un autre objet, texte, ou séquence de film/son.

- *Les annotations temporelles sur des fichiers diachroniques* : ce type d'annotations n'est utilisable que dans le cas d'objets pour lesquels le facteur diachronique est important. Il s'agit des films ou des fichiers sons qui constituent une entité composite d'information séquentielle, or chaque séquence peut à elle seule être significative. Les descripteurs y reçoivent ainsi, en plus de l'identification et du contenu sémantique, un descripteur temporel qui permet de retrouver la séquence particulière qui concerne l'information. Lorsque le film est visionné, les annotations enregistrées par les divers utilisateurs apparaissent sur l'écran au moment voulu, tel un sous-titrage, au choix du visiteur. Ces annotations permettent également, au travers du moteur de recherche, de retrouver le moment et le lieu exact de l'occurrence d'un mot ou d'une annotation complète dans un film ou un son sans devoir visionner la totalité du film.

reliant les objets. Il pourra ainsi tirer profit de son travail en ligne en vue d'une publication, par exemple. La seconde sera la préparation d'une duplication de la base MySQL dans des répertoires XML de sorte que l'information puisse être moissonnée par des moteurs XML, par exemple ceux qui obéissent aux normes de l'OpenArchives.

The screenshot shows the ODSAS interface. At the top, the logo reads "ODSAS Online Digital Sources and Annotation System for the social sciences". The URL is "http://odsas.pacifico-credo.fr". There are language options for "Français" and "English". Below the header is a toolbar with icons for "filename" and "annotation" search, and navigation arrows. The main area features a large image of a group of people in traditional attire performing a ritual. To the left is a table of contents with a grid of numbers. To the right is a metadata panel for document ID 23932, titled "Photographies d'un rituel de graduation (Nimangki) à Lamap, Malekula, Vanuatu". Below this is an "Annotations disponibles pour ce document" section with one annotation by Laurent Dousset, dated 2009-01-03, titled "Nimangki". The annotation text describes dancers around a drum. At the bottom right is a "Liens vers d'autres documents" section with one link to document ID 23813.

Figure 7 : Vue d'ensemble de l'écran d'un objet. De gauche à droite : Première ligne : Ajouter un objet à l'ensemble, ouvrir l'objet dans une nouvelle fenêtre à sa taille d'origine, chercher dans les noms de fichiers, chercher dans les annotations, aller à l'objet précédent, aller à l'objet suivant, aller à la table des matières de l'ensemble. Seconde ligne (bloc de gauche) : Table des matières iconographique de l'ensemble et naviguer dans cette table des matières, (bloc central) l'objet lui-même et les fonctions géométriques déjà décrites, (bloc de droite) colonne des annotations sémantiques avec en haut les descripteurs de base (qui s'étendent au fur et à mesure des renseignements fournis), au milieu les annotations textuelles et en bas les liens vers d'autres objets.

2) Gestion des utilisateurs

Comme nous l'avons souligné, une bonne gestion des utilisateurs et de leurs droits d'accès est une fonction essentielle afin de répondre aux problèmes de confidentialité et de « sensibilité » des données que nous avons exprimés en introduction à cet exposé. Quatre types généraux d'utilisateurs sont disponibles dans la plateforme ODSAS.

- a) le *superuser* : il s'agit de l'administrateur général qui crée les utilisateurs de niveau b.
- b) les propriétaires : chacun de ces utilisateurs est défini comme propriétaire d'une ou de plusieurs collections, ou ensembles ou objets. Il gère les objets, peut en ajouter de nouveaux, gère les descripteurs de base, peut ajouter, éditer, modifier ou supprimer ses annotations et crée les utilisateurs de niveau c.
- c) Les utilisateurs autorisés : ces utilisateurs ont un accès « d'auteur » d'annotations et de relations entre objets à des collections, ensembles ou objets selon les autorisations fournies par les propriétaires de niveau b. Ils peuvent participer à l'annotation des objets. Ils constituent le groupe de travail temporaire ou permanent sur un ensemble d'objets.
- d) Utilisateurs publics : ils n'ont pas de droits d'annotation ou de mise en relation d'objets, mais peuvent seulement visionner et manipuler les objets (sans pouvoir enregistrer ces manipulations) si les objets sont rendus publics par les utilisateurs de niveau b.

La définition et l'attribution des droits se font par l'interface publique. Dès que le visiteur s'est enregistré, des fonctions supplémentaires apparaissent sur l'interface lui permettant d'atteindre des fonctions particulières qui permettent de gérer l'utilisation et les droits.

3) Gestion des collections et des ensembles

Les utilisateurs de niveau a et b peuvent modifier les informations générales qui accompagnent et décrivent les collections et les ensembles. Ces informations sont disponibles au moteur de recherche, tout comme les annotations lorsqu'elles sont publiques. Il s'agit là de descripteurs de base de type Dublin core enrichis par des descripteurs sémantiques.

Dans la version actuelle d'ODSAS, seule l'utilisateur de niveau *a* peut instaurer une nouvelle collection et créer des ensembles au sein de cette collection. Les versions prochaines de la plateforme permettront aux utilisateurs autorisés de créer eux-mêmes de nouvelles collections ou de nouveaux ensembles au sein de collections existantes et de télécharger des documents originaux afin de peupler les ensembles (pour le moment ils ne peuvent qu'ajouter des objets individuels à des collections et ensembles existants). La gestion des objets eux-mêmes deviendra ainsi totalement décentralisée et autonome. Cette fonction sera mise en ligne dès que l'ensemble des filtres de conversion et de vérification des divers formats des fichiers originaux téléchargés par les propriétaires seront disponibles.

4) Gestion des objets et des annotations

Les utilisateurs de niveau *a*, *b* et *c* peuvent manipuler les objets et les annoter. Les manipulations sont géométriques en ce qu'elles modifient la géométrie interne des fichiers, et obéissent à un certain nombre de fonctions qui doivent améliorer le visionnage de l'objet ou personnaliser la manière dont l'objet doit être montré à l'utilisateur. Les fonctions actuelles sont la rotation, la transformation de l'objet en noir et blanc, éclaircir ou assombrir l'objet, la transformation en positif/négatif, l'agrandissement ou la miniaturisation. D'autres fonctions, comme l'annotation graphique (pouvoir surligner en couleur des parties de l'objet, par exemple) seront ajoutées. Ces modifications et manipulations peuvent être enregistrées de sorte que l'utilisateur auteur de ces manipulations puisse visionner à chaque visite l'objet tel qu'il l'a défini. Mais cet enregistrement est lui-même une couche d'annotation et ne modifie pas l'objet original ni l'objet tel qu'il peut être vu par d'autres visiteurs.

Dans ce même registre figurent les annotations textuelles à proprement parler. Les utilisateurs de niveau *a*, *b* et *c* peuvent informer l'objet en renseignant les descripteurs prédéfinis, ou en l'enrichissant d'un texte libre (une annotation sémantique). Ce texte participe à la constitution d'un champ sémantique qui enveloppe l'objet et qui permettra au moteur de recherche de l'identifier. D'autres utilisateurs peuvent ajouter de nouvelles annotations qui répondent à des annotations déjà existantes et construire ainsi une pyramide ou un système de couches d'annotations sémantiques autour de l'objet. Il en est de même avec les relations

entre les objets, une fonction qui fut déjà décrite dans le chapitre précédent. L'ensemble de ces mises à jour et renseignements utilisent le protocole AJAX de sorte que la mise à jour d'une partie de l'écran n'implique pas le rechargement de toute la page, accélérant considérablement l'expérience de navigation pour l'utilisateur.

The screenshot shows a web interface with a form titled "Descripteurs de base du document". The form is highlighted with a red dashed border. It contains the following fields and text:

- Doc id: 23932 | Set id: 213
- Author.....
- Year.....
- Document title.....
- Identification type..... (write something like "page", "serial number" etc.)
- Identification..... (THE page numer, THE serial number etc.)
- Location / Place name.....
- Copyright infos.....

At the bottom of the form, there are two buttons: "insérer / éditer" and "ANNULER". The background shows a partial view of a document page with a photo of a person and some text.

Figure 8 : Formulaire de mise à jour des descripteurs de base (extrait d'écran). De nouveaux champs et descripteurs peuvent être globalement définis par les utilisateurs de niveau a.

Protection et diffusion

Nous avons commencé notre exposé en essayant d'expliquer les problèmes fondamentaux qui accompagnent l'archivage de matériaux ethnographiques. Il engage la responsabilité du chercheur en matière de confidentialité et de « sensibilité » des données. Les données ethnographiques sont différentes des données historiques (histoire ancienne, s'entend) ou des données de la sociologie quantitative. Nous avons souligné que ces problèmes sont l'une des causes majeures de la résistance des chercheurs eux-mêmes à mettre à disposition leurs sources et données. Il était donc nécessaire que la plateforme développée puisse répondre à ces problèmes, qu'elle puisse offrir un moyen d'archivage et de mise à disposition au moins potentielle du matériel tout en répondant aux questions déontologiques

associées à l'archivage ethnographique lui-même.

Afin de pouvoir répondre à ces questions déontologiques, notre choix fut de définir le dépositaire comme étant également le propriétaire du matériel, et de lui confier le droit et même l'obligation de déterminer les droits d'accès au matériel. En d'autres termes, nous avons décidé que ce n'est pas à la plateforme ou à ses gestionnaires de décider de la nature politique et personnelle, confidentielle et sensible du matériel. La plateforme dispose donc, pour les utilisateurs propriétaires (utilisateurs de niveau b), de différents réglages d'attribution des droits d'éditions et de visibilité. Le propriétaire peut en premier lieu décider si une collection entière, si un ensemble entier ou si un objet singulier peut être vu par le public anonyme, ou, de manière plus restreinte, s'il est accessible à un groupe d'utilisateurs particulier et défini. Le propriétaire peut constituer un « groupe de travail », c'est-à-dire un ensemble de personnes, a priori des chercheurs collègues ou collaborateurs sur un projet collectif, auquel il attribue le droit de visionner un corpus donné. Le propriétaire d'une collection ou d'un ensemble peut donc créer et définir de nouveaux utilisateurs auxquels il pourra attribuer certains droits.

add right to see	SET: 66	Sumware
add right to see	SET: 67	Dessins Topogrpahiques
remove right -->	SET: 69	Aamen distribution
remove right -->	SET: 70	Aamen Mumu
remove right -->	SET: 71	Aamen Benium Edils

Figure 9 : Les utilisateurs propriétaires peuvent par simple click attribuer des droits d'accès à leurs ensembles à d'autres utilisateurs, et révoquer ces droits à tout moment. Ils restent « maîtres » de leurs ensembles et collections.

Il s'ensuit que le propriétaire étant maître de sa collection ou de son ensemble, il est aussi le seul à décider quels sont les objets qu'il veut archiver. Aucune sélection, aucun choix préalable et aucun filtrage de la part des concepteurs, programmeurs ou responsables de la plateforme n'est effectué. Car nous avons décidé qu'il n'y a pas a priori de documents bons à archiver et d'autres qui ne le seraient pas. Il n'y a pas de bonnes archives ou de mauvaises

archives. Il n'y a que des questions auxquelles l'archive peut, ou au contraire ne peut pas, apporter des éléments de réponse. C'est au chercheur de décider ce qu'il pense nécessaire d'archiver, et ceci sans présumer des questions et des thématiques que les chercheurs de demain pourront construire au sujet du matériel. Le chercheur, à partir du moment où il est reconnu comme tel et qu'il dispose d'un espace qui lui a été attribué par le *superuser* sur le serveur, peut disposer comme il l'entend de ses archives. La seule restriction à laquelle il se soumet est qu'il ne peut ni techniquement ni moralement révoquer un objet, un ensemble ou une collection entière. Autrement dit, il est dans l'obligation de les préserver. Même s'il décide de n'attribuer aucun droit d'accès à aucun utilisateur, il ne peut retirer ses contributions de la plateforme et ne peut demander leur destruction. Ses contributions scientifiques sont ainsi préservées et sont assujetties au concept même d'archives. En fonction de la législation en vigueur en France et de celle des pays dans lesquels le matériel fut récolté, ainsi qu'en fonction des décisions testamentaires du propriétaire, il deviendra public à partir d'un certain nombre d'années.

L'auteur de chaque annotation ou relation entre objets pourra également décider si l'annotation ou le lien entre des objets sont personnels ou publics. Dès lors qu'ils ont été rendus publics, ils gardent ce statut. Si l'auteur définit l'annotation ou le lien comme étant personnel, alors lui seul verra apparaître cette annotation ou ce lien sur l'écran de l'objet en question, et lui seul pourra décider de ce qu'il doit advenir de sa contribution. Cette particularité peut s'avérer utile lorsque le propriétaire travaille sur un ensemble particulier et que les annotations qu'il produit ne sont que temporaires, des aide-mémoires qui doivent servir à la constitution d'un corpus plus élargi. Les annotations, si le propriétaire ou auteur en décide ainsi, peuvent par la suite et à tout moment devenir publiques.

La contrepartie à cette protection forte du matériel, des objets et des annotations est une visibilité amoindrie. La plateforme disponible sur un serveur liste une grande quantité d'objets, d'ensembles et de collections dont la majeure partie reste pour le moment inaccessible par l'utilisateur lambda. Afin de pallier cette frustration évidente, toutes les informations d'un objet et d'une collection ne sont pas retenues ou cachées. En effet, habituellement, lorsqu'un utilisateur n'a pas accès à une partie d'une base de données ou d'une page, l'accès lui est simplement refusé *dans son intégralité* : soit vous avez accès, soit vous n'avez pas accès. Nous avons choisi d'autoriser l'utilisateur à accéder à la page en

question, celle qui héberge l'objet, mais de masquer les informations sensibles ou confidentielles : l'objet principal et les annotations privées. Le visiteur peut, malgré les restrictions, visionner le contexte physique et sémantique dans lequel se trouve l'objet qu'il désire visualiser (voir les deux figures ci-dessous).

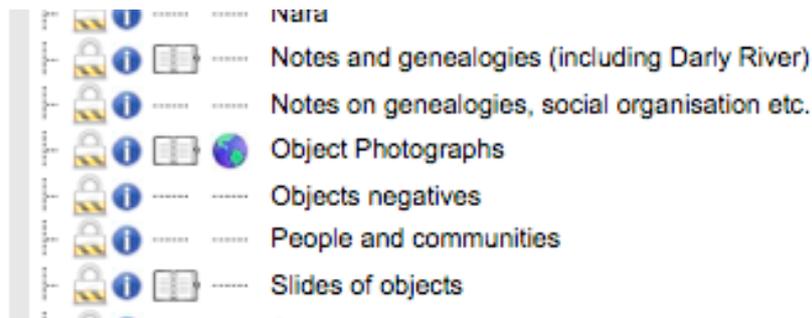


Figure 10 : Liste des ensembles d'une collection (extrait d'écran). Lorsqu'un utilisateur lambda consulte la liste des collections et ensembles, cette liste est accompagnée de plusieurs icônes. La première, le cadenas, lui signifie qu'il n'a pas accès à cet ensemble et que cet ensemble est verrouillé. Afin de pouvoir y accéder, il doit contacter le propriétaire et négocier avec lui les droits d'accès. Si l'ensemble lui est accessible, le cadenas est remplacé par un lien d'accès à l'ensemble. La seconde icône, le « i », affiche des informations générales sur l'ensemble, les descripteurs de base (variables selon le taux de renseignement par le propriétaire/auteur). Ces informations sont disponibles même si l'ensemble n'est pas accessible. La troisième icône, lorsqu'elle s'affiche, renvoie le visiteur vers la table des matières de l'ensemble. Cette table des matières est construite automatiquement dès le moment où le propriétaire informe les objets par des annotations ou descripteurs de base. Cette table des matières est disponible même lorsque le visiteur n'a pas l'autorisation nécessaire pour visionner l'ensemble. La quatrième icône, le globe, apparaît lorsque le propriétaire ou les auteurs ont renseigné des coordonnées géographiques pour l'ensemble en question. Ce lien renvoie vers une carte GoogleEarth en temps réel. Il est également disponible lorsque le visiteur ne possède pas tous les droits nécessaires.

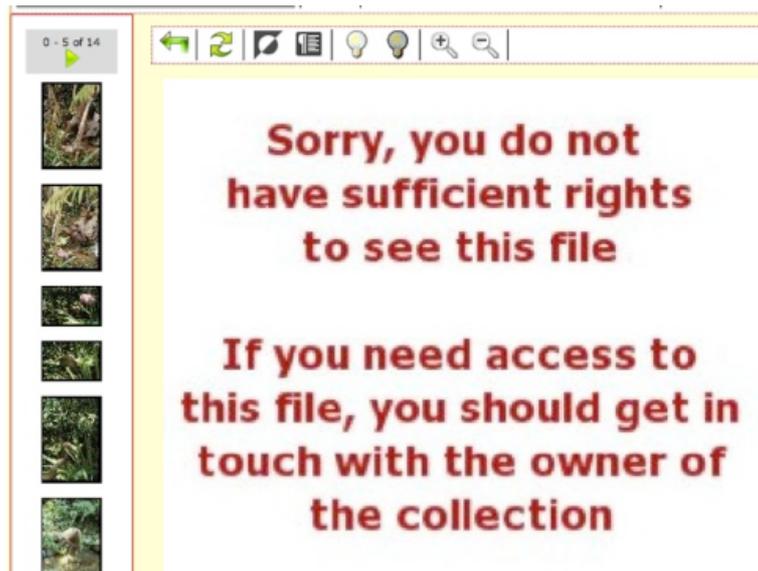


Figure 11 : Lorsqu'un utilisateur lambda navigue vers la page d'un objet « privé », l'objet lui-même n'est pas montré. A la place apparaît une image-type d'excuse. Ceci permet au visiteur d'apercevoir le contexte dans lequel se trouve l'objet en question. En particulier, il peut visionner les images miniatures de la table des matières iconographique (bloc de gauche), et il peut voir les annotations et liens vers d'autres objets si le propriétaire ou les auteurs ont décidé qu'ils sont publics.

En conclusion de cette partie, soulignons que nous avons tenté autant que possible de trouver un compromis acceptable entre déontologie anthropologique et visibilité publique. La plateforme permet au visiteur de comprendre qu'un objet avec son champ sémantique construit existe, mais ne lui permet pas d'accéder aux données les plus sensibles de cet objet.

Plus-value pour le chercheur

Un autre problème évoqué en début d'exposé fut celui de la plus-value pour le chercheur. Le temps nécessaire pour archiver et informer/organiser ces archives est considérable. Les chercheurs ne peuvent s'adonner à cette tâche si elle n'apporte pas une réelle plus-value organisationnelle ou analytique. Il fallait donc développer une plateforme dans laquelle l'archivage lui-même produit un gain de temps pour ou dans l'analyse du matériel.

Nous avons ainsi conçu une plateforme qui permet au propriétaire d'organiser son matériel, de le mettre en relation, de créer des pages analytiques virtuelles qu'il pourra réutiliser. L'archivage est en même temps un moyen d'organiser l'information et de l'enrichir en vue de l'analyse et de la publication. Notons à ce sujet les fonctions particulières suivantes :

- L'utilisateur peut facilement partager des ensembles ou des objets individuels avec d'autres chercheurs ou les mettre à disposition aux groupes et sociétés hôtes et contribuer ainsi au rapatriement des données ou documents. Il peut constituer un groupe de travail temporaire ou permanent qui annote les objets ou qui surimpose des annotations à des annotations existantes, créant une pyramide sémantique d'annotations en couches.
- L'utilisateur peut modifier les objets géométriquement et enregistrer ses préférences. Il peut par exemple décider, à partir de son domicile où la connexion au réseau est plus lente, que les objets soient reproduits à taille plus modeste, accélérant de fait considérablement le temps de travail. Il peut améliorer la qualité de l'image sans modifier l'original archivé.
- L'utilisateur peut annoter des objets et associer ainsi ses photographies avec ses notes de terrain ou d'autres sources d'information. Il peut attribuer des termes et des phrases libres qui lui permettent de retrouver des objets ou de regrouper des ensembles d'objets de manière temporaire ou permanente selon sa propre classification terminologique et ses propres typologies.
- L'utilisateur peut annoter des films et retrouver les séquences exactes qui correspondent à une chaîne sémantique, ce qui implique un gain de temps considérable pour l'analyse des documents filmiques et pour leur réutilisation dans leur mise en perspective avec d'autres objets ou textes.
- L'utilisateur peut relier de manière temporaire ou permanente des objets de collections ou ensembles différents ou identiques; il peut annoter la signification de cette mise en relation et rechercher ces relations, ou les regrouper.
- L'utilisateur peut, enfin, exporter les pages ainsi constituées dans des fichiers de format RTF et les réutiliser dans un traitement de texte dans le cadre de la préparation d'une publication, par exemple (cette fonction est en cours de finalisation)

Conclusion

Notre plateforme, en constante évolution, suscite un grand intérêt dans la communauté scientifique. Des chercheurs nationaux comme internationaux de grande renommée ont proposé (ou sont en cours de discussion) de faire d'ODSAS la plateforme d'archivage de leur matériel scientifique. Pour citer un professeur australien qui va rejoindre ODSAS : « Nous n'avons chez nous aucun système qui nous permette d'archiver et de réutiliser le matériel tout en garantissant la confidentialité quand elle est nécessaire ». Si la progression suit la même courbe qu'elle a connue au cours de cette première année d'existence d'ODSAS (depuis janvier 2008), la plateforme hébergera d'ici peu des centaines de milliers d'objets uniques et irremplaçables pour les SHS en général et l'anthropologie et son histoire en particulier. Ce côté positif de notre projet doit toutefois être contrebalancé par ce qui risque d'être une prévision négative, comme nous allons essayer de l'expliquer ci-dessous.

Telle qu'elle est disponible en ligne, la plateforme ODSAS est une version bêta, donc une version fonctionnelle mais toujours en évolution. D'autres fonctions viendront s'ajouter ou seront modifiées et améliorées. Nous pensons tout particulièrement à l'intégration de fonctions géométriques supplémentaires, par exemple la possibilité de surligner virtuellement en transparence des parties d'images ou de texte, d'ajouter des éléments graphiques sur des images (des cercles, du texte, des lignes etc.) et ceci toujours sans modifier l'original lui-même. Nous pensons également à un module qui construirait une pré-analyse à partir de la carte sémantique virtuelle encadrant les ensembles ou les collections. Les possibilités d'amélioration et d'innovation sont quasi illimitées. Cette évolution au fur et à mesure que les idées et besoins voient le jour n'est possible que parce que nous avons choisi un modèle de développement qui n'est pas « planificateur » mais « évolutif » ou « intuitif ». Seul ce modèle permet de produire et de maintenir un outil qui pourra rester en phase avec les besoins de la recherche et l'évolution technique. Seul ce modèle possède la capacité d'adaptation nécessaire à la pérennisation des projets. La contrepartie de ce modèle est bien évidemment la nécessité d'efforts considérables et quasi permanents, et d'un échange constant entre les utilisateurs, les propriétaires des objets et la programmation elle-même. En d'autres

termes, il n'est pas possible de désaliéner une telle plateforme de ses concepteurs et programmeurs.

Ce modèle n'est donc applicable que si le groupe de concepteurs, programmeurs et utilisateurs « tests » est limité en nombre au sein de petites ou moyennes équipes, et si le système d'information global dans lequel les multiples plateformes numériques des SHS s'insèrent est lui-même décentralisé. Il est nécessaire que des projets comme ODSAS puissent, tout en s'insérant dans des programmes plus généraux de numérisation dans les SHS, maintenir leur indépendance créative et leur réactivité.

Les premières décennies de traitement automatique de données ont pêché par la volonté d'imposer aux SHS des modes de fonctionnement généraux auxquels nos disciplines ne pouvaient se plier sans perdre ce qui souvent fait leur spécificité, à savoir la « normalité » de données marginales, difficiles à faire entrer dans les cadres de traitement établis, mais cruciales pour nos disciplines. Tel fut le cas en histoire (travaux de Marcel Couturier à l'EHESS), mais aussi en anthropologie (groupe « Anthropologie et calcul » autour de Philippe Richard).

La mise en place de méta-portails partagés qui contraindraient des petits projets (mais dont le potentiel de croissance est majeur, comme pour ODSAS) à se conformer à des structures et politiques générales et imposées limiterait lourdement la capacité d'évolution de la plateforme, et la rendrait vite obsolète car trop lourde et peu adaptative. Un méta-portail, aussi nécessaire qu'il soit, ne peut être que fédérateur et fédératif, et non pas centralisateur s'il ne veut pas étouffer les initiatives et les spécificités des uns et des autres.

