



Le traitement des archives numériques d'architectes

Le fonds de l'architecte Adrien Fainsilber
Fonds déposé au Centre d'archives d'architecture du xx^e siècle,
Cité de l'architecture et du patrimoine



Adrien Fainsilber examinant le plan de la ZAC Richter de Montpellier (cliché Jean-Jacques Sarciat). Extrait de J.-J. Sarciat, « Fainsilber dans son Richter », *Montpellier Plus*, n° 209, 21 novembre 2006, p. 2.

2014

La gestion archivistique des fichiers numériques produits par les architectes dans le cadre de leur activité soulève un grand nombre de questions auxquelles les archivistes n'ont pas encore de réponses précises.

Le Centre d'archives d'architecture du xx^e siècle (Cité de l'architecture et du patrimoine) et la société PASS Technologie ont obtenu du Service interministériel des archives de France une subvention de recherche sur le sujet.

La recherche a porté, de septembre 2013 à juin 2014, sur l'un des premiers fonds d'archives de ce type reçu au Centre d'archives d'architecture, celui de l'architecte Adrien Fainsilber.

Le présent document, rapport final de cette recherche, expose les problématiques soulevées (en dehors des questions liées aux droits associés aux archives numériques) et les réponses proposées. Celles-ci constituent une méthodologie de traitement, conçue pour ce fonds précis (caractérisé par le choix du logiciel qu'avait effectué l'agence à la fin des années 1980) et pour les outils documentaires particuliers du Centre d'archives.

Au-delà de ce cas particulier, cependant, cette recherche apporte des informations concernant l'ensemble des fichiers numériques de dessin assisté par ordinateur, et ses conclusions peuvent éclairer tout chantier d'archivage de ce type d'archives.

Sommaire

1.	Présentation de l'étude.....	6
1.1	Contexte de l'étude.....	6
1.2	Contenu de l'étude.....	6
1.3	Suivi scientifique, réunions.....	7
1.4	Modalités de travail.....	7
1.5	Structure du présent rapport.....	8
1.6	Rédacteurs.....	8
1.7	Contributeurs.....	8
1.8	Glossaire.....	9
2.	Le fonds Fainsilber.....	11
2.1	Description sommaire du fonds.....	11
2.2	Description technique des archives numériques.....	11
2.3	Description matérielle des archives numériques.....	12
2.4	Classement des fichiers par « Objets ».....	12
3.	Formats.....	14
3.1	Formats présents dans le fonds Fainsilber.....	14
3.2	Formats et sort final des fichiers.....	15
3.2.1	Repérage des formats présents dans le fonds.....	15
3.2.2	Analyse des formats utilisés.....	17
3.3	Formats de pérennisation : examen des formats candidats.....	18
3.3.1	Formats techniques.....	19
3.3.2	Formats complexes (conteneurs).....	19
3.3.3	Grille d'analyse.....	20
3.4	Formats de pérennisation retenus.....	21
3.4.1	Format de présentation : PDF/A ou PDF/E vs SVG.....	21
3.4.2	Format d'utilisation : DXF vs DWG.....	22
3.5	Conclusion.....	23
4.	Traitement du fonds.....	24
4.1	Choix de projets pour l'expérimentation.....	24
4.2	Types de données.....	24
4.3	Structure d'un document technique finalisé (arbre de fichiers liés).....	26
4.3.1	Cascades de liens.....	27
4.3.2	Bibliothèques d'objets.....	27

4.4	Spécificités d'Arc+	28
4.4.1	La bibliothèque ARCALIB	28
4.4.2	Les fichiers liés et la structure d'une présentation	29
4.4.3	Les fichiers de configuration (.par).....	29
4.4.4	Fonction d'export DWG/DXF.....	32
4.4.5	Synthèse et nettoyage.....	35
4.5	Les doublons.....	36
4.5.1	Repérage des doublons par hachage ou par comparaison de métadonnées.....	37
4.5.2	Analyse des doublons à partir du hachage.....	37
4.5.3	Un exemple : les doublons des sauvegardes (zip) du projet de Colmar	39
4.5.4	Les doublons dans les fichiers de bibliothèque (ARCALIB).....	40
4.5.5	Conclusion : principes de gestion des doublons	40
4.6	Sélection de formats de pérennisation	41
4.7	Modalités des exports vers PDF et DXF.....	41
4.7.1	Exports vers PDF.....	41
4.7.2	Exports vers DXF.....	42
4.7.3	Futurs formats PDF.....	43
4.8	Procédure complète de conversion	43
5.	Système de gestion et de description des fichiers.....	44
5.1	La base de données du Centre d'archives.....	44
5.2	Organisation des fichiers numériques selon l'arborescence de la base de données	44
5.2.1	Rappel et mise en cause des principes définis en 2011	44
5.2.2	Une conversion partielle, uniquement pour les fichiers de présentation... ..	45
5.2.3	...Mais une double conversion	46
5.2.4	Une réorganisation seulement partielle des fichiers	46
5.3	Module de traitement des archives numériques dans la base de données	47
5.3.1	Import des métadonnées dans le formulaire général.....	47
5.3.2	Sous-formulaire des métadonnées	48
5.3.3	Sous-formulaire Doublons.....	50
5.3.4	Attribution du dossier de destination	50
5.3.5	Finalisation du traitement d'un objet	51
5.3.6	Récapitulatif des opérations de traitement des fichiers numériques pour un objet donné	51
6.	Conclusions.....	52

6.1	Outillage spécifique pour le traitement d'un fonds d'archives numérique	52
6.1.1	Outil d'identification des formats	52
6.1.2	Outils de conversion pour atteindre des formats de pérennisation	52
6.1.3	Outils de vérification des formats de pérennisation	53
6.2	Compétences et connaissances nécessaires	53
6.3	Archivage préliminaire dans l'agence	53
6.4	Bilan du traitement des archives Fainsilber	54
6.4.1	Comparaison entre le traitement des archives numériques et des archives sur papier 54	
6.4.2	Priorités de traitement pour les fichiers techniques du fonds Fainsilber	55
6.4.3	Formats de destination pour les fichiers techniques du fonds Fainsilber.....	56
6.4.4	Quelques éléments d'évaluation (temps et coût) du traitement du fonds Fainsilber..	56
6.4.5	Ce qui peut être généralisé	58
6.4.6	Précautions ou recommandations	59
7.	Webographie	60
7.1	Cité de l'architecture et du patrimoine	60
7.2	SIAF	60
7.3	Formats, normes et standards	60
8.	Bibliographie.....	62
8.1	Cité de l'architecture et du patrimoine	62
8.2	Autres	62
9.	Annexes	63
9.1	Étude de faisabilité du traitement du fonds, juillet 2011	63
9.2	Charte graphique transmise par l'agence	63
9.3	Utilisation d'Arc+ à l'agence	63
9.4	Liste complète des types de fichiers	63
9.5	Tableau de grilles d'analyse	64
9.6	Analyse des 5 CD Zip de Colmar	64
9.7	FAINAD-A-2 : Analyse	64
9.8	Bibliothèque Arc+	64
9.9	Procédure de conversion des fichiers Arc+ et PLT	64
9.10	Gestion et description des archives numériques : rappels	64
9.11	Manuel ArchiVecture : Guide de saisie	64
9.12	Matériel, systèmes et logiciels de l'agence Fainsilber	64

1. PRÉSENTATION DE L'ÉTUDE

1.1 CONTEXTE DE L'ÉTUDE

Parmi les fonds conservés au Centre d'archives d'architecture du xx^e siècle de la Cité de l'architecture et du patrimoine (dorénavant « Centre d'archives » et « Cité »), le fonds d'archives Adrien Fainsilber (présenté *infra*, voir 2.1) est le second fonds arrivé avec un ensemble de fichiers numériques natifs. Reçues quelques années plus tôt, les archives de l'architecte Pierre Riboulet avaient fait l'objet d'une étude exploratoire en 2005-2006, résumée lors du colloque Gau:di sur les archives numériques¹.

La première intervention sur les archives numériques Fainsilber a consisté dans la copie des CD d'origine sur le serveur du Centre d'archives et dans une réflexion sur les modalités de leur futur traitement. Cette étude a été confiée à Cassandra Éveillard-Kervarrec en 2011 (citée ci-dessous comme « étude de 2011 » ou « rapport de juillet 2011 »), qui proposait une première esquisse de procédure, n'a pu déboucher sur le traitement des archives².

L'année suivante, en 2012, la société PASS Technologie³ et la Cité de l'architecture et du patrimoine ont demandé au Service interministériel des Archives de France (SIAF) une subvention de recherche pour une étude portant sur la pérennisation et le traitement archivistique des archives numériques natives du fonds Adrien Fainsilber. Cette subvention, accordée, a été intégrée au budget 2013 de la Cité, qui a passé commande à PASS Technologie pour sa part de l'étude. Un apport en industrie a été intégré par PASS Technologie à sa participation, et la Cité (Centre d'archives) a également fourni un important apport en industrie non prévu à l'origine.

À l'intérieur de la Cité, c'est le Centre d'archives qui, avec la direction des systèmes d'information de la Cité, a coordonné la recherche.

1.2 CONTENU DE L'ÉTUDE

À l'intérieur du fonds numérique Fainsilber, l'étude porte plus précisément sur les fichiers de DAO issus du logiciel Arc+ (version 14.03) et sur le traitement à leur appliquer en vue de leur pérennisation et de leur communication au public. Elle prolonge l'étude de 2011, qui avait axé l'esquisse de procédure élaborée sur la recherche du temps de traitement minimal des fichiers (notamment en

¹ Étude menée par Florence Wierre au Centre d'archives d'architecture du xx^e siècle. Voir F. Wierre, « Le traitement des archives numériques de l'agence Pierre Riboulet : de nouveaux rapports au temps », *Architecture et archives numériques. L'architecture à l'ère numérique : un enjeu de mémoire*, Gollion : InFolio, 2008 (actes du colloque Gau:di de décembre 2007), p. 235-244.

² Plus précisément, cette mission avait pour objectifs : 1/ d'assurer la sauvegarde du fonds numérique (pas encore transféré des CD sur un serveur à cette date) ; 2/ d'effectuer un repérage de cette partie du fonds ; 3/ d'élaborer une proposition de méthode de traitement pour la pérennisation, l'organisation et la description des fichiers ; 4/ de réaliser une première estimation financière pour l'ensemble de l'opération de traitement. L'étude a révélé, entre autres, le manque d'informations cohérentes sur la structure des fichiers, leur organisation et la nature des liens entre ces fichiers. (Cf. § 9.1 **Étude de faisabilité du traitement du fonds, juillet 2011.**)

³ Titulaire du marché passé par la Cité de l'architecture et du patrimoine pour la maintenance et les évolutions de l'application ArchiWebture.

recherchant une approche permettant de n'ouvrir chaque fichier, dans l'idéal, qu'une seule fois au cours de la chaîne de traitement).

La présente étude s'est décomposée en deux volets, étudiés parallèlement :

1. L'évaluation des qualités et des défauts des différents formats de sortie et de pérennisation possibles pour des fichiers natifs Arc+;
2. La description archivistique dans la base de données ArchiVecture des archives numériques d'un ensemble de projets de l'architecte Adrien Fainsilber, permettant de valider, de remettre en question ou de compléter les étapes définies par le rapport de juillet 2011.

1.3 SUIVI SCIENTIFIQUE, RÉUNIONS

L'étude a été menée essentiellement par Nader Boutros (directeur général, PASS Technologie) et Sonia Gaubert (adjointe, responsable du système d'information, Centre d'archives d'architecture du ^{XX}^e siècle).

Le contrôle et l'avis scientifique ont été effectués et apportés :

Pour le volet 1, par Michel Jacobson, chef de projet, à l'époque, sur l'archivage électronique, Service interministériel des archives de France (SIAF), bureau des traitements et de la conservation ;

Pour le volet 2, par David Peyceré, responsable du Centre d'archives, et Sonia Gaubert ;

Pour l'ensemble, par Renaud Sagot, responsable des services d'information, Cité de l'architecture et du patrimoine.

Le comité scientifique réunissait le SIAF (notamment Françoise Banat-Berger, alors sous-directrice de la politique interministérielle et territoriale pour les archives traditionnelles et numériques), la Cité de l'architecture et du patrimoine, PASS Technologie et les Archives départementales de l'Hérault (Cassandra Éveillard-Kervarrec). Ce comité s'est réuni en décembre 2013, février et juin 2014.

1.4 MODALITÉS DE TRAVAIL

De nombreuses réunions ont eu lieu entre Nader Boutros, Sonia Gaubert et David Peyceré au cours de la mission.

Un accès « connexion bureau à distance » a été créé par la DSI de la Cité de l'architecture et du patrimoine pour PASS Technologie le 27 juin 2013. La Cité de l'architecture et du patrimoine a acquis une licence Arc+. Le travail s'est fait au Centre d'archives et à distance sur le poste dédié à Arc+ du Centre d'archives.

Le travail à distance a cependant posé problème pendant la plus grande partie de l'étude : après de nombreux échanges avec le support d'Arc+ et l'acquisition d'une clé réseau, une solution performante de travail à distance depuis les locaux de PASS Technologie n'a pu être mise en place qu'en mai 2014.

1.5 STRUCTURE DU PRÉSENT RAPPORT

Le chapitre 2, « Le fonds Fainsilber », présente le fonds, les modalités de dépôt et la méthodologie de classement primaire appliquée en 2011.

Les chapitres 3, « Formats », et 4, « Traitement », portent sur le volet 1. Leur diffusion relève d'abord (et non exclusivement) du SIAF.

Le chapitre 5, « Système de gestion et de description des fichiers », porte sur le volet 2. Destiné d'abord à l'usage interne du Centre d'archives, il peut cependant être utilement diffusé à l'extérieur.

Enfin, ce rapport se termine par une série de conclusions et plusieurs annexes, dont un *manuel de traitement* détaillant les étapes du traitement d'un dossier d'archives numériques dans le contexte du Centre d'archives (serveur du Centre d'archives, base de données ArchiVecture). D'autres éléments ont été produits lors de cette recherche, notamment des descriptions archivistiques des fichiers numériques traités, dans la base de données ArchiVecture, et un ensemble de fichiers numériques créés dans le cadre de l'étude par conversion des fichiers du fonds.

La Cité, PASS Technologie et le SIAF disposent des droits complets de diffusion et d'exploitation sur l'ensemble de ces rendus.

1.6 RÉDACTEURS

Cité de l'architecture et du patrimoine (CAPA)

- Sonia Gaubert, adjointe, responsable du système d'information, centre d'archives
- David Peyceré, responsable du centre d'archives

PASS Technologie (PASS-TECH)

- Nader Boutros, directeur général

1.7 CONTRIBUTEURS

Cité de l'architecture et du patrimoine (CAPA)

- Renaud Sagot, direction des systèmes d'information

Service interministériel des archives de France (SIAF)

- Françoise Banat-Berger, sous-directrice de la politique interministérielle et territoriale pour les archives traditionnelles et numériques
- Michel Jacobson, chef de projet sur l'archivage électronique, sous-direction de la politique interministérielle et territoriale pour les archives traditionnelles et numériques ; bureau des traitements et de la conservation

Archives départementales de l'Hérault (AD Hérault)

- Cassandre Éveillard-Kervarrec, responsable des archives d'architecture

PASS Technologie (PASS-TECH)

- Pierre Dittgen, directeur technique

- Julien Bellion, développeur senior

Agence Fainsilber

- Yves Chauvin, chargé de la gestion informatique (à l'époque du versement)

1.8 GLOSSAIRE

TERME OU ACRONYME	DÉFINITION
Base de données 2D	Ce terme est souvent utilisé pour désigner le répertoire contenant l'ensemble des fichiers d'un projet. Cette notion de « base de données du projet » reflète bien l'étendue et l'importance des informations inscrites dans les documents techniques pour l'agence d'architecture.
BIM	Building Information Model(ing) (Maquette numérique du bâtiment)
CAO	Conception assisté par ordinateur
DAO	Dessin assisté par ordinateur
ddd, icn, iii	Les extensions d'un document Arc+
Document	Un document métier peut être représenté par un ou plusieurs fichiers informatiques. Un document Arc+ est en général représenté par au moins 2 fichiers (.iii et .ddd) et peut faire référence à plusieurs autres documents Arc+. Les fichiers représentant ces documents liés par référence doivent conserver les mêmes chemins relatifs d'accès afin que le document original puisse être reconstitué.
Document [ou fichier] de présentation	Mise en page d'un ou plusieurs documents de travail afin de constituer une planche complète, y compris le cartouche et les intitulés des différentes vues sur les documents de travail.
Document [ou fichier] de travail	Document technique graphique contenant une partie d'une planche de présentation.
DXF	Drawing eXchange Format, format créé par Autodesk servant à échanger des fichiers DAO ou CAO entre systèmes CAO n'utilisant pas le même format de fichier natif
DWG	AutoCAD Drawing (Fichier AutoCAD)
Espace objet (EO)	Espace objet : terme lié à AutoCAD pour décrire l'espace de travail standard du logiciel, dit aussi « modèle »
Espace papier (EP)	Espace papier : terme lié à AutoCAD pour décrire l'espace de travail dédié à la mise en page, dit aussi « présentation »
Fichier	Un fichier informatique peut être autonome, faire partie d'un ensemble de fichiers liés ou contenir plusieurs fichiers

Information graphique en image <i>raster</i>	Information graphique transcrite dans un fichier image sous forme d'une matrice de pixels. Chaque pixel contient une information sur la couleur et/ou la transparence. La taille de la matrice définit la qualité de l'image. Plus la matrice est grande plus l'image contient des informations plus détaillées. L'impression et l'échelle de ces informations dépend de la résolution maximale de l'imprimante ou du traceur et de la résolution souhaité par l'utilisateur en fonction de la qualité du papier utilisé. Cette résolution est exprimée en général en PPP (points par pouce) ou DPI (dots per inch). Plus la taille de la matrice augmente plus le poids du fichier augmente. Différentes techniques de compression permettent de maintenir un poids de fichier raisonnable pour une qualité correcte.
Information graphique vectorielle	Information graphique transcrite dans un fichier informatique sous une forme mathématique permettant la reconstitution de cette information à l'échelle souhaitée par l'utilisateur sans perte ou dégradation. Il s'agit en général de fichiers techniques issus de logiciel de CAO/DAO ou BIM.
Objet	Le terme « Objet » désigne, dans le présent rapport, la représentation en base de données d'un objet architectural ; en général, il s'agit d'un projet architectural.
PLT	Hewlett Packard Vector Graphic Plotter File (Fichier traceur de données graphiques vectorielles)
Projet	Le terme « Projet » désigne, dans le présent rapport, un projet architectural (réalisé ou non).

2. LE FONDS FAINCILBER

2.1 DESCRIPTION SOMMAIRE DU FONDS

Le fonds d'archives de l'agence Adrien Fainsilber et associés (AFA) a été donné à l'État par l'architecte Adrien Fainsilber en décembre 2007, et déposé à la Cité de l'architecture et du patrimoine (Centre d'archives d'architecture du xx^e siècle). Celui-ci l'a pris en charge matériellement au printemps 2008.

Né en 1932, l'architecte Adrien Fainsilber a ouvert en 1970 son agence, qui est devenue AFA en 2000. Connu avant tout pour son principal projet parisien, la Cité des sciences et la Géode, il est l'auteur de très nombreux équipements publics et d'importants aménagements urbains. On renvoie ici à la notice d'ArchiWebture (http://archiwebture.citechailot.fr/fonds/FRAPN02_FAINAD) sur l'architecte et à la bibliographie qui lui est consacrée, en grande partie mentionnée dans cette notice⁴.

Le fonds couvre, avec d'importantes lacunes (pas de pièces écrites pour beaucoup de projets), l'activité de l'architecte depuis ses débuts, dans les années 1960, jusqu'au début des années 2000. Le *terminus ad quem* du fonds est assez flou, compte tenu du fait que l'agence a continué son existence sous le nom AFA même après le départ en 2008 d'Adrien Fainsilber et de son épouse Julia Fainsilber, associés de la société AFA. Certains projets en cours sont donc restés à l'agence AFA, aussi bien pour les archives sur supports traditionnels que pour les fichiers numériques.

Cas jusque-là exceptionnel pour le Centre d'archives, les archives numériques de l'agence ont également été versées, concernant certains projets menés de 1993 à 2003 environ. Une répartition remarquable des archives de projets ou de réalisations doit être notée : pour certains projets les archives sont essentiellement des impressions sur papier, pour d'autres projets essentiellement des fichiers numériques. Cet état de fait n'a pas trouvé d'explication lors des échanges entre le Centre d'archives et l'agence.

2.2 DESCRIPTION TECHNIQUE DES ARCHIVES NUMÉRIQUES

La majorité des fichiers concernait les documents techniques des projets réalisés avec le logiciel de dessin assisté par ordinateur Arc+. L'agence utilisait le logiciel Arc+ pour ses fonctions de dessin 2D, en exploitant très peu les fonctions 3D du logiciel ou les fonctions décrivant les éléments architecturaux. Certains fichiers reçus ou échangés sont dans des formats liés au logiciel AutoCAD. Certains de ces fichiers sont des conversions des fichiers Arc+ pour les besoins d'échange avec les entreprises et les bureaux d'études. Parallèlement à l'usage d'Arc+, l'agence a progressivement adopté AutoCAD, et le passage s'est fait progressivement ; les projets les plus récents contiennent de plus en plus de fichiers créés avec AutoCAD.

⁴ Notamment deux monographies : Madec (Philippe), Pélissier (Alain), *Adrien Fainsilber. La virtualité de l'espace. Projets et architecture. 1962-1988*, Paris, Electa Moniteur, 1988 ; Desmoulin (Christine), *Adrien Fainsilber et associés. 1986-2002*, Paris, éditions Alternatives, 2003.

Lors d'entretiens avec David Peyceré en 2008, avec Cassandre Éveillard-Kervarrec en 2011 puis avec Nader Boutros en 2013, l'architecte qui jouait au sein de l'agence, le rôle de responsable informatique à l'époque du don d'archives, Yves Chauvin, a donné notamment les informations suivantes :

- L'agence s'est dotée d'une « charte graphique » en 2006, donc postérieurement à l'élaboration des fichiers numériques contenus dans le fonds d'archives. Par « charte graphique », il faut entendre en l'occurrence une charte de description des règles de classification, de nommage et de présentation du contenu des documents techniques. Basée sur les méthodes acquises par l'expérience antérieure, elle posait les règles pour une utilisation massive du logiciel AutoCAD au sein de l'agence (Cf. § 9.2, **Charte graphique transmise par l'agence**).
- Les noms des fichiers sont en partie imposés par le logiciel Arc+, et en partie contraints par le système d'exploitation (DOS puis Windows) imposant une longueur limitée des noms des fichiers (au maximum 8caractères.3caractères). L'agence a ainsi été amenée à codifier les noms des fichiers afin d'identifier plus ou moins facilement leur contenu dans la masse importante de fichiers créés pour chaque projet (Cf. § 9.3 **Utilisation d'Arc+ à l'agence**).

En revanche, plusieurs interrogations, restées sans réponse, ont orienté la présente étude, entre autres :

- Comment garantir l'utilisation de la bonne bibliothèque d'objets associée à chaque projet voire à chaque document Arc+ (§ 4.5.1) ?
- Comment garantir la cohérence du contenu des documents Arc+ s'il manque des objets placés (§ 4.5.2, 5.2.1) ?
- Comment identifier les bons fichiers de définition des couleurs et des épaisseurs de traits (*.par) associés à chaque document Arc+ (§ 4.5.3) ?
- La question la plus critique (§ 5.2.1) : que faut-il conserver ou pérenniser de cette masse de documents non renseignés ? Comment identifier les bonnes versions de chaque phase du projet quand l'organisation des répertoires et des fichiers ne les signale pas explicitement ?

2.3 DESCRIPTION MATÉRIELLE DES ARCHIVES NUMÉRIQUES

Les archives numériques ont été données par l'agence Fainsilber en 2008 sous la forme d'un ensemble de 100 CD. Les mentions inscrites sur les CD donnaient une indication probable et approximative sur le contenu du CD. Les fichiers de certains projets occupaient plusieurs CD tandis que d'autres projets se trouvaient regroupés sur un même CD.

En 2011, l'ensemble des CD a été copié sur un serveur de fichiers de sauvegarde répliqué régulièrement. À ce moment, trois ans après le don du fonds d'archives – mais les modalités de création des CD sont inconnues et leur date peut être bien antérieure au don –, deux CD étaient (devenus ?) illisibles et leur contenu n'a pas pu être récupéré. L'ensemble des CD a été détruit à la fin de l'opération de sauvegarde.

2.4 CLASSEMENT DES FICHIERS PAR « OBJETS »

Depuis cette sauvegarde de 2011, le fonds numérique Fainsilber est accessible sur le serveur de fichiers du centre d'archives. Les fichiers numériques natifs sont enregistrés de la même manière que

les fichiers numériques produits par le Centre d'archives par numérisation de documents : dans un partage dédié du serveur, subdivisé en répertoires par fonds, eux-mêmes divisés en sous-répertoires par projet. Les identifiants de ces sous-répertoires, appelés « références Objets⁵ », correspondent à la principale unité documentaire des inventaires en base de données (base de données ArchiVecture). Il existe donc une corrélation étroite entre l'organisation des fichiers sur le serveur et la structure des inventaires en base de données, permettant, via leurs référencements, de faire des liens et d'appliquer des traitements spécifiques aux fichiers depuis la base.

Certains fichiers de bibliothèque, de présentation de l'agence, de logiciels ou de configuration ne concernent pas un objet particulier, et sont transversaux au fonds : ils ont été regroupés dans un objet spécifique du chapitre « Pièces personnelles », intitulé « Documentation professionnelle – bibliothèque » (référence objet FAINAD-A-2), et donc enregistrés dans un répertoire spécifique dans l'arbre des objets.

⁵ Le terme « objet » désigne les projets (d'architecture, d'urbanisme, etc., réalisés ou non) et les autres unités de classements, de type « pièces personnelles » (documents du fonds n'appartenant pas à des projets).

3. FORMATS

Un fonds d'archives numériques d'architecte comprend un ensemble complexe de documents de toutes natures : documents de bureautique (traitement de texte, tableaux, présentations), documents multimédias (images, illustrations, films), photographies, dessins scannés, images de synthèse (directes ou retraitées sous un logiciel de traitement d'image), animations virtuelles, vidéos, documents techniques, sources DAO, sources 3D, exports des sources dans différents formats d'échanges et/ou de présentation.

Parmi l'ensemble de ces typologies, nous nous concentrons dans la présente recherche sur les sources DAO et les exports associés, qui constituent en général la majorité des fichiers. Les autres types de documents nécessitent d'autres recherches spécifiques.

Les fichiers sources et fichiers d'exports, identifiés dans le fonds, peuvent se détailler ainsi :

- Sources (DAO, 3D, etc. ; dans le cas du fonds Fainsilber : fichiers Arc+ puis AutoCAD)
- Exports servant aux échanges avec les bureaux d'études et les entreprises (fichiers DWG)
- Exports servant à la communication et/ou à la contractualisation (uniquement une production papier servait pour ces actions et parfois des DWG)
- Impression virtuelle (dans des fichiers) servant au tirage de plans (fichiers PLT, PLZ)

Enfin, les archives numériques peuvent comprendre des répertoires compressés dans des fichiers ZIP, qui contiennent des sauvegardes partielles ou finales, donc de nombreux doublons, mais aussi, parfois, des documents uniques non présents dans les autres dossiers du projet.

3.1 FORMATS PRÉSENTS DANS LE FONDS FAINSILBER

Dans les archives numériques du fonds Fainsilber, PASS Technologie⁶ a identifié les formats techniques suivants :

Éditeur	Logiciel producteur dans le contexte du fonds	Extensions	Ouvert/fermé	Libre/propriétaire	Pourcentage des fichiers uniques ⁷ du fonds
Arc Technologie	Arc+	.ddd, .iii, .icn, .rpt	DDD, III Fermés RPT format texte ouvert	Propriétaire	81,47%
Autodesk	AutoCAD	.dwg, .dxf, .bak, .ctb	DWG Ouvert (via ODA ⁸), DXF Ouvert par Autodesk	Propriétaire	4,71%

⁶ À partir de l'identification des fichiers uniques dans l'ensemble du fonds, Pierre Dittgen et Nader Boutros de PASS Technologie ont établi un tableau récapitulatif des formats, le nombre de documents uniques et leur pourcentage dans le fonds. Cf. § 9.4 **Liste complète des types de fichiers.**

⁷ Le terme « fichier unique » indique que les doublons (cf. § 4.4) sont décomptés avant d'établir ces pourcentages. Plus précisément, ces pourcentages incluent les fichiers uniques trouvés dans l'arborescence des fichiers du fonds et au sein des archives ZIP de premier niveau (les zip inclus dans des zip n'ont pas été explorés).

⁸ ODA : Open Design Alliance : <https://www.opendesign.com/>

Éditeur	Logiciel producteur dans le contexte du fonds	Extensions	Ouvert/fermé	Libre/propriétaire	Pourcentage des fichiers uniques ⁷ du fonds
HP	AutoCAD / Arc+	.plt, .plz	Ouvert (HPGL, HPGL/2 ⁹)	Propriétaire	2,19%

Au total, les fichiers techniques et leurs impressions virtuelles (celles-ci constituées des fichiers PLT et PLZ) constituent **88,37 %** du fonds.

3.2 FORMATS ET SORT FINAL DES FICHIERS

Le fonds, sur le serveur du Centre d'archives, est organisé par objets correspondant aux projets de l'architecte (cf. § 2.4). Sans accès aux archives papier, seuls les fichiers informatiques peuvent servir d'orientation dans l'exploration et l'identification des fichiers à conserver. Différents types de fichiers existent dans les dossiers.

3.2.1 Repérage des formats présents dans le fonds

Le tableau ci-dessous mentionne les formats les plus représentés dans le fonds (ayant plus de 500 fichiers uniques), ainsi que certains types de fichiers essentiels à l'exploration des fichiers techniques, avec une proposition concernant le sort à leur réserver lors du traitement. Plusieurs autres types existent, mais ne concernent pas des données produites par l'agence (des fichiers de système d'exploration archivés ou exécutables de logiciels divers), et sont ignorés dans la présente recherche.

Extension	Description	Nombre de fichiers uniques	Dont fichiers uniques contenant le caractère @ ¹⁰	Dont fichiers uniques contenant le caractère \$	Sort des fichiers ¹¹
ddd	Chaque document Arc+ est représenté par 2 fichiers de type ddd et iii.	33876	4021	1525	À convertir pour les fichiers contenant un @ ou un \$
iii	Chaque document Arc+ est représenté par 2 fichiers de type ddd et iii.	33369	3960	1503	À convertir pour les fichiers contenant un @ ou un \$
icn	Certains documents Arc+ sont représentés par des icônes de type icn	13655	2082	24	À supprimer

⁹ Hewlett Packard Graphic Language, plus connu sous le nom de HPGL ou HP-GL, est un protocole d'impression informatique développé par HP, qui est devenu un standard de l'industrie (http://fr.wikipedia.org/wiki/Hewlett_Packard_Graphic_Language).

¹⁰ Les fichiers contenant @ ou \$ sont les fichiers de présentation d'Arc+.

¹¹ La colonne « Sort des fichiers » décrit une synthèse rapide du traitement à appliquer à chaque fichier.

Extension	Description	Nombre de fichiers uniques	Dont fichiers uniques contenant le caractère @ ¹⁰	Dont fichiers uniques contenant le caractère \$	Sort des fichiers ¹¹
dwg	Un document AutoCAD est représenté par un fichier au format natif du logiciel dwg.	3388	4		À vérifier ¹² / à convertir
plt	Impression dans un fichier destinée à l'envoi sur un traceur ou une imprimante indépendamment du logiciel qui l'a créé	2178		209	À convertir
jpg	Format d'image standard	1735			À vérifier
doc	Document Word	1722			À vérifier/convertir
dxf	Format d'export (initialement ¹³ lié à AutoCAD) permettant l'échange de documents techniques	1244	1	1	À vérifier/convertir
xls	Document Excel	820			À vérifier/convertir
met	Fichier préparatoire d'Arc+ pour la génération d'une impression dans un fichier de type PLT	815	313		À supprimer
zip	Archive de documents pouvant contenir une arborescence de fichiers contenant des documents divers et d'autres archives	693	4		À explorer ¹⁴
rpt	Fichier de rapport textuel généré par Arc+. Le contenu de ces rapports n'est pas toujours structuré de la même façon. Il dépend de la fonction du logiciel qui l'a généré.	510	109		À explorer

¹² « À vérifier » : le format peut être pérenne si le fichier est conforme à sa spécification.

¹³ De plus en plus de logiciels utilisent le format DXF comme format d'export

¹⁴ « À explorer » : signifie que le format contient des informations ou des fichiers qu'il est impossible de déduire du format, une exploration détaillée du contenu sera nécessaire.

Extension	Description	Nombre de fichiers uniques	Dont fichiers uniques contenant le caractère @ ¹⁰	Dont fichiers uniques contenant le caractère \$	Sort des fichiers ¹¹
par	Fichier de définition de la configuration des couleurs et des épaisseurs de traits, de l'échelle des types de lignes et d'autres paramètres spécifiques à Arc+	212	18	11	À interpréter ¹⁵
tif	Format d'image standard	140			À explorer
tga	Format d'image standard	118			À explorer
psd	Format d'image natif du logiciel Photoshop	104			À explorer
ctb	Table des couleurs et des épaisseurs de traits d'AutoCAD	85			À interpréter
rtf	Export normalisé de fichier de traitement de texte	57			À vérifier/convertir
lsp	Programme AutoLisp lié à AutoCAD	24			À explorer
plz	Archive de plusieurs PLT	21			À convertir

3.2.2 Analyse des formats utilisés

La présente recherche avait l'ambition de fournir des informations d'analyse des fichiers techniques sur tous les formats disponibles. À l'expérience, cette ambition n'était pas raisonnable – il faudrait couvrir l'ensemble des logiciels existant depuis plus de 20 ans, dont certains ont disparu ou ont évolué à plusieurs reprises. Ainsi, le traitement de chaque fonds devrait commencer par une recherche préliminaire déterminant les formats et le périmètre des outils nécessaires à la conversion pour une pérennisation optimale.

À travers le fonds Fainsilber, et grâce aux entretiens avec le responsable informatique de l'agence à l'époque du don des archives, on peut comprendre l'évolution de l'outillage de l'agence, très marqué par le travail en 2D et l'utilisation de l'outil informatique comme « planche à dessin informatisée ». Passer d'Arc+ (sous-utilisé) à AutoCAD semblait être l'évolution naturelle pour optimiser la production de l'agence et rendre plus efficaces les échanges avec ses partenaires.

Les fichiers Arc+ sont caractérisés par la complexité des liens entre eux, et par la dépendance à certains fichiers de configuration et à des fichiers de bibliothèque¹⁶ attachés parfois à la version du logi-

¹⁵ « À interpréter » : le fichier doit subir une analyse manuelle et une conversion spécifique pour être pérennisé ou pour restituer une information utile pour d'autres fichiers de travail.

ciel (cf. §4.3.2). Comme précisé au §3.1, nous allons nous limiter à l'analyse et la pérennisation des formats de fichiers techniques trouvés dans le fonds.

3.3 FORMATS DE PÉRENNISATION : EXAMEN DES FORMATS CANDIDATS

Ce paragraphe présente les formats candidats pour la pérennisation des données techniques DAO présentes dans le fonds Adrien Fainsilber.

Pour chaque format natif, il est nécessaire d'identifier s'il est ouvert (spécification publiée), et valide (respecte sa spécification). Il est aussi utile de privilégier les formats libres (n'imposant pas de droits à payer pour acquérir la spécification) quand cela est possible.

Pour le Centre d'archives, le choix du format dépendra notamment des outils effectivement disponibles pour réaliser les conversions souhaitées à moindre coût, et de la disponibilité d'outils de vérification fiables.

Sont distingués des formats « techniques » et des formats « conteneurs ». Ces derniers peuvent « contenir » des données techniques (2D, 3D), graphiques (illustrations et images) et textuelles assemblés dans une composition. Les logiciels de PAO¹⁷ ou d'illustration sont typiquement des logiciels producteurs de ce type de formats, par opposition aux logiciels de DAO producteurs principalement des formats techniques.

Le choix d'un format conteneur complexe (type PDF) a été privilégié, plutôt qu'une « simple » image (JPEG ou TIFF) correspondant à une impression numérique ou aplat graphique, pour les raisons suivantes :

- Un format image (*raster*) peut être intéressant uniquement si l'opérateur connaît sa résolution, et respecte cette résolution lors de l'impression pour aboutir à un document papier correspondant au document technique numérique d'origine à la bonne échelle définie à l'origine.
- Ce format est plus adapté aux numérisations à partir des sources papier.
- La contrainte de la résolution fige l'information technique vectorielle dans un état obligeant l'impression à une résolution fixe définie arbitrairement en fonction des conditions matérielles de restitution à un moment donné. Un format vectoriel permet le respect de l'échelle définie à l'origine en maintenant l'information vectorielle accessible à n'importe quelle autre échelle nécessaire, sans dégradation de l'information ni surcharge du poids du fichier, et en profitant pleinement des évolutions du matériel de restitution pour une restitution optimale.
- Sur un format raster, il est difficile de réaliser des mesures relatives précises, d'appliquer des opérations d'exploration (agrandissements / zooms) sans dégradation de l'affichage (interpolation / pixellisation). Afin d'éviter cet effet, la sortie image est réalisée, en général, à très haute résolution, occupant ainsi un espace de stockage trop important par rapport à son équivalent en format vectoriel.

¹⁶ Les fichiers de bibliothèque correspondent à des dessins techniques partiels utilisés souvent dans différents projets. Ces fichiers peuvent être fournis par les logiciels de DAO et/ou par des fournisseurs tiers et complétés par l'agence.

¹⁷ Publication assistée par ordinateur.

3.3.1 Formats techniques

DXF¹⁸, *Drawing eXchange Format*, est un format créé par la société Autodesk servant à échanger des fichiers DAO ou CAO entre systèmes CAO n'utilisant pas le même format de fichier natif. Autodesk publie systématiquement les spécifications de ce format au fur et à mesure de son évolution (dernier en date : DXF 2013). Les fichiers DXF peuvent être de format ASCII ou binaire. [Volumineux]

DWG¹⁹, *DraWinG*, est un format binaire natif des fichiers issus du logiciel AutoCAD de la société Autodesk à l'origine. De par sa diffusion, il est devenu *de facto* le standard de l'industrie CAO et DAO. L'ODA : Open Design Alliance publie les évolutions des spécifications du format de façon régulière. [Utilisé par la majorité des logiciels de CAO/DAO comme format d'export/import ou comme format natif.]

Ces formats préservent toute l'information technique vectorielle (organisation des calques, types de lignes, hachures, etc.).

3.3.2 Formats complexes (conteneurs)

PDF/A²⁰, *Portable Document Format/Archive*, est une standardisation (ISO 19005-1à3) du format PDF, à l'origine propriétaire de la société Adobe. Le PDF/A est ainsi spécialisé pour l'archivage numérique de documents électroniques. Le SIAF et le CINES ont produit un guide méthodologique²¹ et une étude sur les outils de conversion du format PDF²² (cf. § 7 webographie).

Découpage du standard ISO 19005 :

Partie	Nom	Nom initial	Date de sortie	Standard	Basé sur la version du PDF
19005-1	PDF/A-1	Utilisation de PDF 1.4 (PDF/A-1)	2005	ISO 19005-1	PDF 1.4 (Adobe Systems, PDF Reference third edition, 2001)
19005-2	PDF/A-2	Utilisation de l'ISO 32000-1 (PDF/A-2)	2011	ISO 19005-2	PDF 1.7 (ISO 32000-1:2008)
19005-3	PDF/A-3	Utilisation de l'ISO 32000-1 avec le support de fichiers encapsulés (PDF/A-3)	2012	ISO 19005-3	PDF 1.7 (ISO 32000-1:2008)

¹⁸ <http://fr.wikipedia.org/wiki/DXF>,
<http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454&linkID=10809853>

¹⁹ <http://fr.wikipedia.org/wiki/DWG>

²⁰ <http://en.wikipedia.org/wiki/PDF/A>

²¹ NUMEN-SIAF-CINES-GM-PDF-1.0, août 2012, <http://www.archivesdefrance.culture.gouv.fr/static/6189>

²² NUMEN-SIAF-HUMANUM-CINES-GM-OCPDF-1.0, janvier 2014,
<http://www.archivesdefrance.culture.gouv.fr/static/7248>

PDF/E²³, *PDF/Engineering*, « Gestion de documents – Format de documents d'ingénierie utilisant le PDF – Partie 1: Utilisation du PDF 1.6 (PDF/E-1) » : Cette norme (ISO 24517-1:2008) vise à définir un format de fichier dédié à l'échange de documents d'ingénierie basé sur le format PDF.

Elle vise également à spécifier l'utilisation correcte de PDF pour l'affichage et l'impression de documents d'ingénierie, et a vocation à favoriser l'échange de documents et la collaboration dans le cadre des workflows des projets d'ingénierie.

L'évolution vers PDF/E-2 vise la spécification des configurations nécessaires pour permettre l'échange de la documentation 2D/3D dans les domaines du bâtiment, de l'industrie et géo-spatial. Cette nouvelle norme inclut le support des formats 3D PRC (ISO 14739-1) et U3D (ECMA-363), l'impression précise des documents techniques (fichiers vectoriels), la sécurisation des échanges de documents, la gestion des annotations et des commentaires, la signature électronique, la stabilité du format, les calques, et l'inclusion de fichiers multimédias, y compris des animations. Ce format tend vers la préservation long-terme des données techniques permettant à la fois leur présentation et leur utilisation (cf.3.4, cf.5.2.2) au sein d'un même fichier au format PDF.

Nom	Nom initial	Date sortie	Standard	Basé sur la version du PDF
PDF/E-1	Utilisation du PDF 1.6 (PDF/E-1)	2008	ISO 24517-1:2008	PDF 1.6
PDF/E-2	Utilisation de l'ISO 32000-2 incluant le support de la préservation long-terme (PDF/E-2)	2013	Draft International Standard (DIS) ISO 24517-2	PDF 2.0 (ISO 32000-2) en cours de développement

SVG²⁴, *Scalable Vector Graphics* (recommandation W3C SVG version 1.1) : c'est un format de fichiers XML pour la représentation de l'illustration graphique. Ce format peut contenir des liens vers des fichiers externes ou incorporer ces fichiers au sein du document SVG par encodage spécifique. Il permet aussi l'inclusion de scripts permettant de rendre dynamique et interactive l'exploration d'une illustration sous forme d'animation par exemple.

Il est apparu possible d'exporter les fichiers techniques dans ce format après plusieurs manipulations. Le résultat est proche d'un fichier d'illustration lourd et difficile à manipuler avec les navigateurs supportant ce format de nos jours. (Cf.3.4.1)

3.3.3 Grille d'analyse

La grille d'analyse de ces formats et des formats techniques trouvés dans les archives du fonds, inspirée de celle ayant servi pour le Guide méthodologique sur les formats des données orales et visuelles²⁵, est développée en annexe **9.5 « Tableau de grilles d'analyse »** en reprenant les descriptions des formats. Cette grille permet d'identifier de façon objective le ou les formats à retenir pour la pérennisation, ainsi que les outils de conversion et de validation disponibles. Elle couvre les for-

²³ <http://fr.wikipedia.org/wiki/PDF/E>

²⁴ <http://www.w3.org/TR/SVG11/>

²⁵ Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles, ADONIS/SIAF/CINES-GM-0.5, 2011, <http://www.archivesdefrance.culture.gouv.fr/static/4923>

mats techniques existants dans le fonds Adrien Fainsilber et les formats candidats pour la pérennisation.

La grille d'analyse comprend les champs suivants :

Identifiants PRONOM²⁶ s'il y a lieu

Nom du format

Date / affiliation / norme ...

Description sommaire inspiré des sources libres sur internet

Catégorie du format (technique, conteneur généraliste, format 3D, etc.)

Destination du format (natif logiciel, échange, impression, archivage, etc.)

Caractère ouvert ou fermé

Libre ou propriétaire

Standard ou de facto ou inconnu

Textuel ou binaire

Compressé ou non

Support de métadonnées

Outils de vérification du format par rapport à sa spécification

Logiciels d'exploitation/conversion (nom, éditeur, URL, fonctions supportés : lecture, écriture, annotation, formats d'export, etc.)

3.4 FORMATS DE PÉRENNISATION RETENUS

Pour un document numérique à pérenniser, deux formes de conservation sont définies (voir § 5.2.2).

- L'une des formes de conservation est destinée à la **présentation**, afin de restituer à l'écran le fichier dans la forme de sa destination originelle (l'impression, dans notre cas).
- L'autre forme est destinée à l'**utilisation** (ou la **réutilisation** a posteriori), afin de restituer à l'écran le fichier tel qu'utilisé par son créateur sur le poste d'origine avec un logiciel de substitution (logiciel DAO ou logiciel de visualisation).

On parlera donc de format de présentation (destiné à la consultation courante du fonds d'archives) et de format d'utilisation (destiné à l'exploration technique).

On peut aussi parler respectivement de format de conservation pour l'archivage et de format de conservation pour le partage public de données (plus communément appelé Open Data). En effet, les formats sources sont souvent propriétaires et sont intimement liés aux logiciels qui les gèrent. Pour pouvoir pérenniser ces fichiers sources, divers compromis ont été établis en fonction de critères objectifs et réalistes prenant en compte les moyens matériels, logiciels et financiers du Centre d'archives.

3.4.1 Format de présentation : PDF/A ou PDF/E vs SVG

Les outils disponibles aujourd'hui fournissent des imprimantes virtuelles permettant l'impression dans un fichier. Auparavant, ce sont les formats .prn, .plt ou .eps qui étaient privilégiés. Aujourd'hui,

²⁶ PRONOM : <http://apps.nationalarchives.gov.uk/PRONOM>

les pilotes d'impression virtuelle permettent essentiellement la production de documents PDF. Le but est dans tous les cas de garder la fidélité au document papier imprimé sous forme numérique. La majorité de la production numérique de l'agence Alain Fainsilber était destinée à l'impression, et donc essentiellement à une production papier. Il faut donc privilégier un format de présentation qui, à l'écran, s'approche au plus près du rendu des documents destinés à être imprimés, en s'assurant de la disponibilité de normes, d'outils de visualisation adaptés et des outils de validation.

Les formats PDF et SVG sont les formats candidats.

Format textuel, le SVG (XML contrôlé par un schéma) se prête bien à la pérennisation des données.

C'est cependant le format PDF qui est ici préconisé, pour les raisons suivantes :

- Il dispose de normes pour l'archivage (PDF/A-1 à A-3, et d'autres en cours de finalisation comme PDF/E-2, spécialement conçu pour la préservation des données techniques à long terme) ;
- Le fichier SVG produit est verbeux puisque textuel (c'est le cas de tous les formats basés sur XML), donc lourd, et difficile à ouvrir et à vérifier avec les outils disponibles actuellement, même s'il existe une version compressée (SVGZ) : un même fichier converti en SVG et PDF prend 2 minutes pour s'ouvrir en SVG et 3 secondes pour s'ouvrir en PDF sous un navigateur comme Firefox. Une fois ouverts, les manipulations de zoom et panoramique sont saccadées pour le premier et fluides pour le second.
- La traduction d'un fichier technique vectoriel dans un fichier au format SVG n'est, pour l'instant, pas optimisée dans les logiciels qui le permettent.

3.4.2 Format d'utilisation : DXF vs DWG

À l'heure actuelle, les données techniques ne sont pas exploitables directement à partir des conversions vers le format PDF. Les outils de conversion utilisés ne fournissent pas une compatibilité et un archivage des données techniques originelles, mais, dans la plupart des cas, produisent un fichier PDF dit de présentation permettant d'obtenir une visualisation vectorielle de la sortie du document destiné à l'impression. En effet, le document produit contient des entités graphiques vectorielles pour décrire la forme finale du document technique traduisant ainsi toutes les configurations techniques (épaisseurs des traits, types et échelles des types de lignes, polices de caractères, cotations, etc.) en entités graphiques.

C'est sans doute une lacune temporaire (PDF/A-3 ou PDF/E-2 promettent de belles avancées sur ce sujet, cf. § 4.7.3), mais pour le moment il a fallu chercher des alternatives existantes et fiables sur le long terme. Deux formats voisins semblent s'imposer pour les documents techniques 2D : Le DWG et le DXF.

Si le DWG est actuellement très répandu, sa spécification est toujours propriétaire et fermée.

Également propriétaire, le DXF a l'avantage de disposer d'une forme basique transcrite en ascii décrivant les objets à créer. Sa spécification est publique et sert essentiellement pour les échanges de fichiers techniques. C'est donc le format DXF qui est finalement retenu ici.

Il faut signaler que ces formats sont tous deux propriétés de l'éditeur Autodesk. Chaque changement de format interne induit par une nouvelle version d'AutoCAD affecte le format natif DWG, et la spéci-

fication évoluée pour la production du DXF correspondant. Actuellement, ces évolutions ont un cycle de 3 ans (dernières en date : 2007, 2010, 2013).

3.5 CONCLUSION

En dehors du format PDF/A-1 ou A-2, aujourd'hui, aucun format n'est spécialement conçu pour l'archivage, tous les autres étant essentiellement des formats d'échange ou des normes en cours de stabilisation. De ce fait, pour la pérennisation de documents techniques, en 2014, en attendant la sortie du PDF/E-2, le format PDF/A semble être la meilleure solution de présentation, sauf s'il existe un module d'export pour le format PDF/E-1. À notre connaissance, seul AutoCAD permet un export PDF 1.6 proche de la norme PDF/E-1.

Pour la pérennisation en vue d'une utilisation ultérieure, c'est aujourd'hui le DXF 2013 qui est à préconiser. Il permet à la fois de préserver une grande partie des données techniques même s'il n'est pas compatible à 100% avec l'ensemble des logiciels techniques du marché.

4. TRAITEMENT DU FONDS

La mission inclut le traitement d'un ensemble de fichiers dans les conditions du Centre d'archives, afin de valider la pertinence des préconisations effectuées, et d'évaluer le temps de traitement de ce nouveau type d'archives.

4.1 CHOIX DE PROJETS POUR L'EXPÉRIMENTATION

Il était prévu au départ que l'étude porterait sur un sous-ensemble du fonds d'archives composé des fichiers numériques relatifs aux trois plus gros projets du fonds, représentant ensemble environ 30% du volume du fonds (plus ou moins 10% chacun) :

- HPGO (hôpital pédiatrique et gynéco-obstétrical), Bron,
- Usine de traitement des eaux, Valentigney,
- Hôtel du département, Colmar.

La proportion de ces projets à traiter pouvait être redéfinie lors de réunions de suivi, mais en principe l'étude devait au minimum inclure l'intégralité de l'un des trois projets.

En fait, il s'est vite avéré nécessaire de raisonner au contraire sur des projets de taille beaucoup plus restreinte afin de préciser les procédures de conversion et de description d'un projet complet au fur et à mesure de son traitement. Une autre liste de projets a ainsi été proposée par Sonia Gaubert, constituée des projets les plus légers en nombre de fichiers ou en poids sur le serveur :

- 1988-1998. Musée d'art moderne et contemporain, Strasbourg (Bas-Rhin),
- 1991-1995. Hôpital pour enfants, Centre hospitalier universitaire de Purpan, Toulouse,
- 1992-1999. Institut mutualiste Montsouris, bd Jourdan, Paris 14^e,
- 1994-2001. Palais de justice, bd Limbert, rue Jacques-Tati, Avignon,
- 1995-2001. Îlot, ZAC Paris Rive Gauche, Paris 13^e,
- 1998-2003. UFR de médecine et de pharmacie, rue Ambroise-Paré, ZAC Les Hauts de Chazal, Besançon (Doubs).

Au cours de ces premières explorations, les projets les plus particulièrement explorés ont été ceux de l'hôtel de département de Colmar et du HPGO de Bron (parmi les trois plus gros), et ceux du Musée d'art moderne et contemporain de Strasbourg et de l'hôpital pour enfants, Centre hospitalier universitaire de Purpan de Toulouse. C'est celui de Strasbourg qui a fait finalement l'objet d'un traitement détaillé.

4.2 TYPES DE DONNÉES



Le projet du Musée d'art moderne et contemporain de Strasbourg s'avère représentatif de la complexité du traitement des données numériques quand nous l'effectuons sans autre repère que les fichiers eux-mêmes et leur organisation dans le système de fichiers. L'exploration des différents pro-

jets cités a permis d'avancer sur la précision des procédures de conversion, ainsi que le choix des outils de traitement nécessaires pour chaque type de données :

- Les fichiers Arc+ sont majoritaires, et redondants. On trouve par exemple une série de fichiers dans un répertoire associé à une phase dite « PEO : Plans d'exécution des ouvrages », et d'autres, identiques, dans un répertoire associé à une phase dite « DOE : Dossier des ouvrages exécutés ». Certains doublons doivent être conservés puisque l'arborescence des fichiers doit être respectée afin de minimiser les interventions manuelles, comme dans cet exemple. D'autres doivent être fusionnés, s'ils se trouvent dans des arborescences identiques parallèles telles les bibliothèques trouvées en plusieurs versions dans le fonds.
- Les fichiers PLT représentent les sorties préparées par l'agence à l'époque du projet. Ils sont la seule trace fiable dans ces archives, de l'identité exacte des documents tels que produits (dessinés, paramétrés et imprimés) par l'agence. Ils peuvent donc être convertis en PDF sans perte.
- Les fichiers DXF sont à vérifier et à conserver.

Une synthèse des fichiers uniques par type de format pour le Musée d'art moderne et contemporain de Strasbourg est présentée ci-dessous :

Légende des couleurs :

-  Documents à traiter en priorité.
-  Documents à conserver après vérification du format.

Format	Analyse	Nombre de fichiers uniques	Dont fichiers uniques contenant @	Dont fichiers contenant \$
.ddd	Étudier uniquement les documents de présentation contenant @ ou \$	627	37	92
.iii	Étudier uniquement les documents de présentation contenant @ ou \$	608	37	92
.icn	Ignorer ou supprimer	71	21	
.plt	Convertir en PDF/A	30	10	
.met	Ignorer ou supprimer	20		
.dxf	Conserver, vérifier et si nécessaire convertir en PDF/1.6 à partir d'AutoCAD	14		1
.par	Analyser, saisir et intégrer dans un document Arc+ (§ 4.4.3)	5		
.psd	Explorer	5		
.rpt	Fichier Arc+ textuel de rapport généré	5		

.plz	Fichier	1		
.zip	Explorer	1		

4.3 STRUCTURE D'UN DOCUMENT TECHNIQUE FINALISÉ (ARBRE DE FICHIERS LIÉS)

Un document de DAO, dans sa structure la plus basique, a la capacité de stocker des entités graphiques (primitives 2D / 3D, conteneurs), des entités non graphiques (textes, annotations, cartouche, etc.) et des références externes : liens vers d'autres documents techniques de DAO, ou graphiques d'illustration et d'images. Ces références externes sont visuellement incorporées dans le document²⁷.

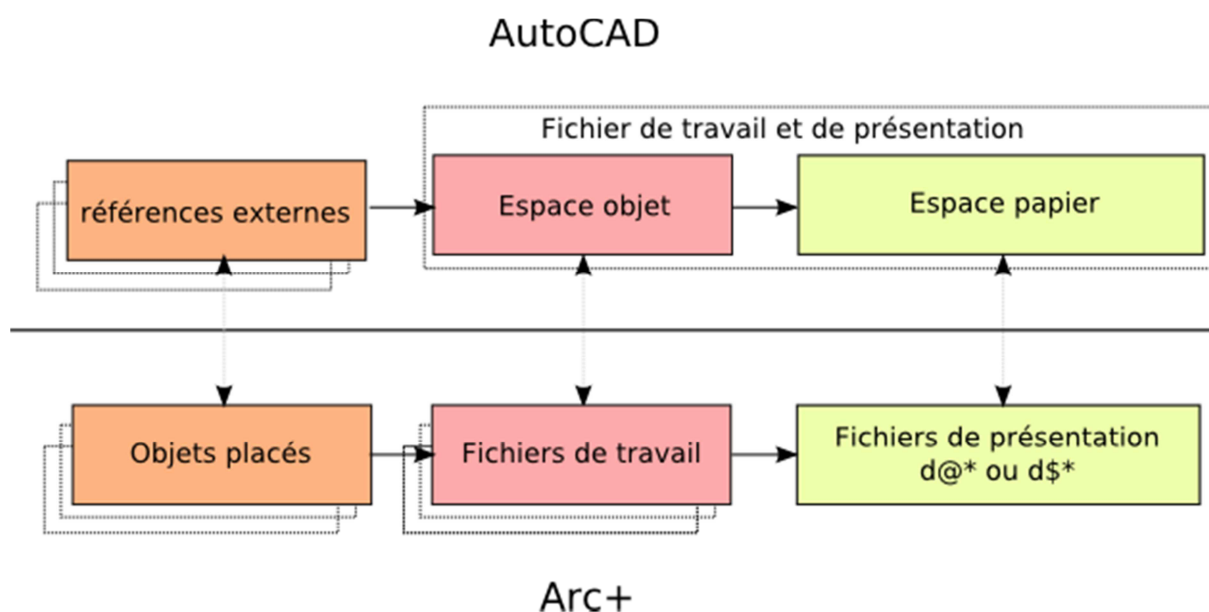


Figure 1. Parallèle entre AutoCAD et Arc+

Chaque entité, graphique ou non graphique, est décrite par un ensemble de propriétés définissant sa morphologie, les transformations appliquées, son appartenance à des catégories, groupes, composants ou calques, etc.

En fonction du logiciel utilisé, des fichiers de configuration ou des éléments de bibliothèques peuvent être associés à une version du logiciel et pas spécifiquement à un projet. Or ces fichiers de configuration et de bibliothèque peuvent influencer l'affichage du document, ses paramètres de mise en page, la table d'affectation des épaisseurs de traits, les couleurs, les polices de caractères, les types de lignes, etc.

Les documents DAO se trouvent ainsi liés à des fichiers de configuration et de bibliothèque ne se trouvant pas forcément dans le même répertoire de travail lié au projet mais dans des répertoires spécifiques au logiciel utilisé.

Sont ici présentés plus en détail deux aspects précis du logiciel Arc+ : les liens en cascade et les fichiers de bibliothèque.

²⁷ Ou, dans le cas du logiciel AutoCAD, d'un onglet à l'intérieur d'un document.

4.3.1 Cascades de liens

Les entités graphiques peuvent être constituées d'éléments géométriques simples et de conteneurs d'éléments. Ces conteneurs sont à leur tour, constitués d'éléments géométriques simples et de conteneurs d'éléments, et ainsi de suite.

Certains logiciels peuvent gérer ces conteneurs au sein d'un document DAO (AutoCAD permet la création de Blocs) ; d'autres établissent des liens à partir d'une collection de conteneurs disponibles dans l'arborescence sous forme de fichiers accessibles au logiciel (Arc+ utilise des fichiers externes pour les objets placés, AutoCAD utilise des fichiers externes pour les Références externes). Arc+ privilégie les liens externes (objets placés, cf. § 4.5.2), tandis qu'AutoCAD permet les deux types (blocs et références externes).

Un document DAO dit de présentation peut ainsi être constitué d'une collection d'entités graphiques géométriques et de données non graphiques textuelles, ainsi que de liens vers des documents DAO de travail (plan, façade, coupe, etc.). Ceux-ci, à leur tour, peuvent faire appel à d'autres documents DAO de travail et à des documents DAO provenant d'une bibliothèque. Voici un schéma théorique de ce principe d'organisation (à gauche) et un exemple concret (à droite).

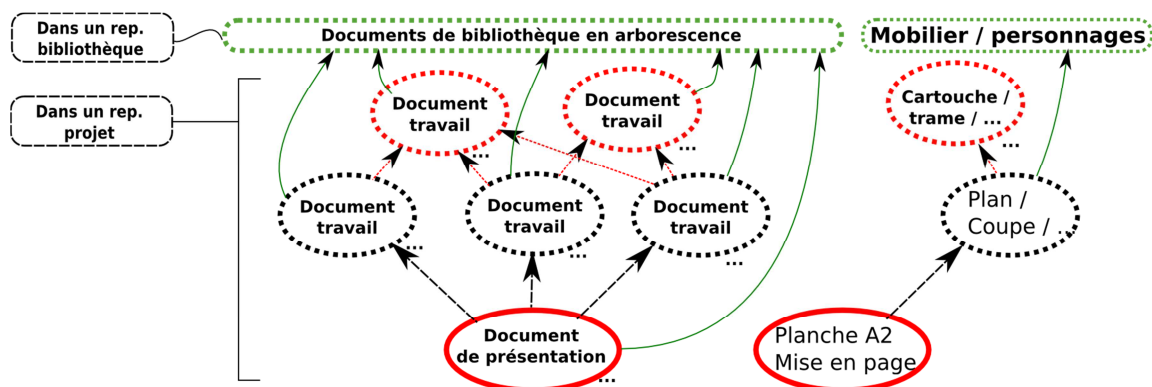


Figure 2 Structure complexe pour la restitution d'un document de présentation

Le cas du fonds Fainsilber illustre bien cette cascade caractérisée par la méthode de travail choisie au sein de l'agence ainsi que par les contraintes liées aux fonctionnalités du logiciel Arc+.

Cette cascade de relations inter-documents optimise certes la productivité, mais influe négativement sur la pérennité et la consistance des documents à archiver. Il suffit qu'un document lié soit manquant pour qu'un ensemble de documents devienne incomplet, partiel et non représentatif des documents communiqués par l'architecte à ses maîtres d'ouvrage ou à ses partenaires.

4.3.2 Bibliothèques d'objets

Dans ses versions utilisées à l'époque de la production des projets dans l'agence Fainsilber, Arc+ positionne la bibliothèque dans un dossier éloigné des fichiers de projet. Les postes de travail n'étaient pas connectés entre eux pour partager la même bibliothèque. Chaque poste disposait donc d'une copie de la bibliothèque commune, copie qui n'a pas été mise à jour systématiquement sur l'ensemble des postes. Les documents DAO de bibliothèque se trouvent ainsi perturbés par le changement de version du logiciel, la suppression, ou même le simple déplacement de fichiers dans l'arborescence de la bibliothèque, ou encore les changements de poste de travail.

Des opérations de fusion entre les différentes bibliothèques ont permis, lors de la présente intervention, d'établir un référentiel commun de bibliothèque accessible à l'ensemble du fonds (cf. § 4.4.4).

4.4 SPÉCIFICITÉS D'ARC+

4.4.1 La bibliothèque ARCALIB

L'étude de 2011 avait développé (cf. § 5.2) un programme Java Archives (.jar) qui permettait, à partir des informations contenues dans les fichiers textuels de rapports RPT générés par la commande \olist, de recopier les fichiers liés (cf. § 4.3.1) de leurs emplacements d'origine vers le répertoire de classement définitif.

La présente recherche a permis de constater que ce programme JAR ne garantit pas que l'on puisse restituer l'intégrité initiale des documents, du moins de façon simple et automatique : en effet, à l'ouverture d'un fichier, Arc+ recherche les fichiers liés dans la bibliothèque spécifique au logiciel (« ARCALIB ») et dans le répertoire du fichier ouvert. Or, étant donné que la structure organisationnelle des répertoires a été modifiée lors du transfert des archives, et notamment que ces bibliothèques étaient enregistrées par défaut sur le disque dur de l'utilisateur du logiciel et ont été enregistrées et transmises avec les archives dans un dossier à part, Arc+ ne peut retrouver les fichiers qu'elles contenaient. Il se trouve que celles-ci (et il en existe autant que de versions du logiciel utilisées dans l'agence) ont été heureusement sauvegardées – sans qu'on sache bien comment²⁸ – dans un répertoire spécifique présent dans le versement.

Il s'agit donc avant tout d'intervenir dans le paramétrage d'Arc+ afin de lui indiquer le chemin actuel des fichiers de bibliothèques, afin qu'il puisse retrouver les fichiers liés. Cela s'est avéré faisable.

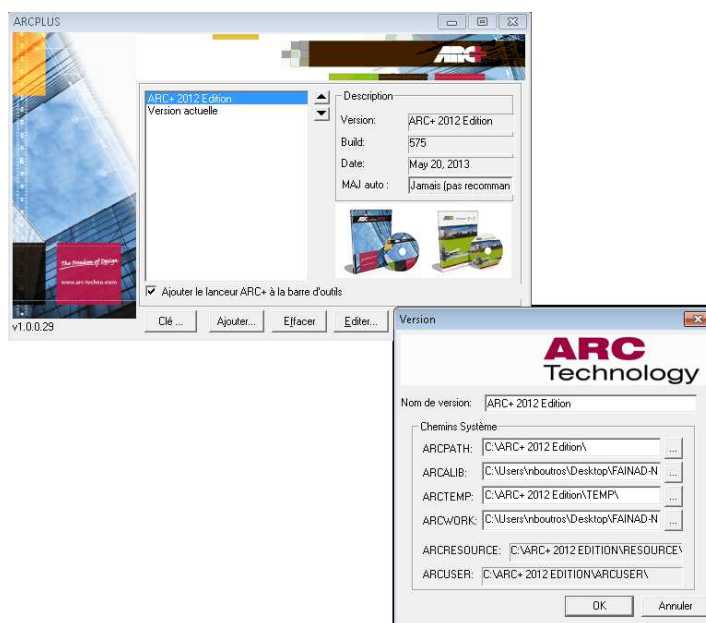


Figure 3 Configuration du chemin de la bibliothèque dite ARCALIB

²⁸ On se demande en effet selon quelle méthode, puisqu'il y a des chances pour qu'il ait existé (pour chaque version du logiciel !) autant de bibliothèques que d'utilisateurs, chacun pouvant « personnaliser » ses outils sur son propre poste.

Mais on ne peut paramétrer dans Arc+ qu'un seul chemin de bibliothèque « ARCALIB », alors que le fonds Fainsilber contenait plusieurs « ARCALIB » distincts, correspondant aux différentes versions du logiciel utilisées successivement dans l'agence (nommées A11, A13, A14, SOFT_DOS). Il fallait ainsi tenter de fusionner toutes ces bibliothèques, ce qui impliquait une étude approfondie des phénomènes de doublons qui ne manquent pas de surgir (vrais doublons ou simples homonymes²⁹). La fusion des répertoires ARCALIB permet d'établir une référence unique aux fichiers de bibliothèque.

En l'occurrence, il s'est avéré que les répertoires des bibliothèques A13 et A14 et leurs contenus étaient strictement identiques, et qu'il suffisait de prendre dans les bibliothèques A11 et SOFT_DOS les fichiers qui n'existaient pas dans A13 et A14 pour arriver à un arbre unique. Il subsistait des fichiers homonymes ayant des empreintes numériques différentes dans A11 et dans SOFT_DOS, que l'on pouvait fusionner en les suffixant (avec par exemple les identifiants : « _A11. » et « _SOFT_DOS. »). L'opération de fusion a permis d'identifier uniquement 6 fichiers homonymes. Étant donné le nombre réduit de ces homonymes, ils ont été explorés un à un pour identifier les différences et ainsi aboutir un arbre unique d'éléments de bibliothèques valable pour l'ensemble du fonds.

Divers cas peuvent impliquer des recherches manuelles de fichiers : outre les bibliothèques de référence mentionnées ci-dessus (choix du fichier lié en cas de doublon), d'autres fichiers non retrouvés dans les répertoires des fichiers maîtres l'impliquent aussi. De plus, les fichiers RPT générés par la commande \olist n'indiquent pas nécessairement le chemin des fichiers liés manquants (mais seulement le nom des fichiers), tandis que des informations complémentaires peuvent figurer à l'ouverture du fichier maître...

Le programme .jar ne peut décidément pas assurer automatiquement la restitution de l'intégrité des documents, et il a été abandonné.

4.4.2 Les fichiers liés et la structure d'une présentation

Comme expliqué plus haut (cf. § 4.3), il est difficile d'isoler ou déplacer un document Arc+ en dehors de son arborescence. En effet, chaque document est lié d'une part à une série d'objets placés (fichiers disponibles dans le même répertoire, cf. § 5.2.1) et des objets de bibliothèques (fichiers disponibles dans un répertoire lié au logiciel Arc+). Dès qu'un lien est manquant, le logiciel génère un rapport signalant les manques introuvables. Les fichiers de présentation (ou de contrôle) permettent d'inclure dans une mise en page finale une composition des éléments graphiques travaillés séparément (fichiers de travail). Ces fichiers de présentation ne peuvent pas identifier les manques éventuels dans les fichiers de travail composant la mise en page finale. Des vérifications systématiques sont nécessaires pour s'assurer de la complétude de chaque mise en page ou fichier de présentation.

4.4.3 Les fichiers de configuration (.par)

Les fichiers PAR enregistrent notamment la codification de la configuration de l'impression (épaisseurs de lignes, échelle des types de ligne, couleurs) du logiciel Arc+.

²⁹ Voir en annexe les notes résultant de l'étude approfondie sur ce point de Sonia Gaubert : « § 9.7 FAI-NAD-A-2 : Analyse » et « § 9.8 Bibliothèque ARC+ ».

Les logiciels (Arc+ et AutoCAD) permettent d'utiliser des couleurs pour identifier visuellement la catégorie de l'élément graphique dessiné (mur, cloison, hachure, espaces verts, ouvertures, mobilier, etc.). Ces couleurs servent aussi au moment de l'impression. Grâce aux fichiers de configuration (PAR pour Arc+ et CTB pour AutoCAD) ces couleurs sont convertis en expression graphique (épaisseur de traits, couleur à l'impression, type de ligne, ...etc).

L'opérateur Arc+ ou AutoCAD, définit ou utilise une définition précise de la codification des couleurs du document de travail. Cette codification (faisant partie en général d'une charte graphique technique), permet d'associer à une couleur visible à l'écran une couleur de sortie et une épaisseur de trait. Arc+ ajoute aussi le type de ligne dans cette codification.

Dans les anciennes versions d'Arc+, les fichiers PAR n'étaient pas liés aux documents graphiques d'Arc+ et servaient à plusieurs dessins. Les documents graphiques d'Arc+ ne gardent aucune référence du ou des fichier(s) de configuration qui auraient servi aux différents tirages.

Certains fichiers .par ne contiennent que la définition des épaisseurs de traits, et les 15 couleurs par défaut n'avaient pas été personnalisées. On peut faire l'hypothèse que les traceurs fonctionnaient avec des plumes à cette époque et que c'était l'opérateur du traceur qui décidait pour l'épaisseur et de la couleur des plumes physiques par rapport à la configuration. N'ayant pas d'autres informations pour identifier les couleurs utilisées à la sortie, on a pris le pari que ces fichiers étaient tracés en noir avec les épaisseurs définis dans le fichier PAR.

L'analyse des 5 fichiers uniques PAR de l'objet Strasbourg (FAINAD-E-1988-2) permet d'identifier une méthode d'exploration permettant la conversion en PDF.

Le contenu des fichiers présente une zone d'impression différente. Cette zone est en général liée à la taille de la mise en page du document de présentation ; en fonction de la taille de la mise en page, nous choisirons pour chaque fichier de travail, la configuration la plus adaptée :

-area=811,1135 -area=1135,811 -area=811,1135 -area=366,564 -area=1200,860

Les tables des épaisseurs de traits liés aux 15 couleurs de base d'Arc+ (versions 11 à 14) présentent plusieurs similitudes :

-color=1,0.2,0.3	-color=1,0.2,0.3	-color=1,0.45,0.6	-color=1,0.25,0.6	-color=1,0.25,0.6
-color=2,0.13,0.3	-color=2,0.13,0.3	-color=2,0.13,0.3	-color=2,0.13,0.3	-color=2,0.13,0.3
-color=3,0.13,0.3	-color=3,0.13,0.3	-color=3,0.13,0.3	-color=3,0.13,0.3	-color=3,0.13,0.3
-color=4,0.13,0.3	-color=4,0.13,0.3	-color=4,0.13,0.3	-color=4,0.13,0.3	-color=4,0.13,0.3
-color=5,0.13,0.3	-color=5,0.13,0.3	-color=5,0.13,0.3	-color=5,0.13,0.3	-color=5,.25,0.3
-color=6,0.13,0.3	-color=6,0.13,0.3	-color=6,0.13,0.3	-color=6,0.13,0.3	-color=6,0.13,0.3
-color=7,0.2,0.3	-color=7,0.2,0.3	-color=7,0.3,0.3	-color=7,0.18,0.3	-color=7,0.18,0.3
-color=8,0.13,0.3	-color=8,0.13,0.3	-color=8,0.13,0.3	-color=8,0.13,0.3	-color=8,0.13,0.3
-color=9,0.45,0.3	-color=9,0.35,0.3	-color=9,0.3,0.3	-color=9,0.3,0.3	-color=9,0.3,0.3
-color=10,0.13,0.3	-color=10,0.13,0.3	-color=10,0.13,0.3	-color=10,0.13,0.4	-color=10,0.13,0.4
-color=11,0.13,0.3	-color=11,0.13,0.3	-color=11,0.5,0.3	-color=11,0.3,0.3	-color=11,0.3,0.3
-color=12,0.13,0.3	-color=12,0.13,0.3	-color=12,0.13,0.3	-color=12,0.13,0.3	-color=12,0.13,0.3
-color=13,0.18,0.3	-color=13,0.18,0.3	-color=13,0.13,0.3	-color=13,0.13,0.3	-color=13,0.13,0.3
-color=14,0.3,0.3	-color=14,0.25,0.3	-color=14,0.13,0.3	-color=14,0.13,0.3	-color=14,0.13,0.3
-color=15,0.13,0.3	-color=15,0.13,0.3	-color=15,0.13,0.3	-color=15,0.13,0.3	-color=15,0.13,0.3

À l'ouverture d'un fichier Arc+, il est impossible d'identifier quel fichier « .par » correspond au dessin. Dans les versions récentes d'Arc+, la configuration des couleurs et des épaisseurs de lignes et types de ligne est intégrée au document Arc+.

Le centre d'archives dispose de la version 2012 d'Arc+, qui semble être la plus récente lors du démarrage de l'étude. Cette version d'Arc+ montre que le logiciel ne permet plus l'utilisation de ces fichiers PAR et définit la configuration des épaisseurs, des couleurs et des types de lignes au sein même du document (un bon enseignement de la pratique des utilisateurs ?). Il est possible d'inclure plusieurs configurations dans un même document Arc+. Or nous ne disposons que de cette version 2012 d'Arc+ et l'accès à une ancienne version d'Arc+ semble compromis pour cette étude. Il a donc fallu retranscrire les différents fichiers .par de l'objet dans un document Arc+ vierge. Lors de l'ouverture d'un document Arc+ du fonds, il est possible de charger l'ensemble des configurations dans ce document afin de tester à loisir les différentes configurations en fonction de la taille de la mise en page.

Dans un fichier vierge d'Arc+, l'ensemble des fichiers « .par » de l'objet Strasbourg a été converti, afin de servir de référence pour le traitement de l'ensemble des fichiers DAO de l'objet.

Assembler l'ensemble des configurations dans un seul fichier Arc+ permet de le charger facilement dans chaque fichier à convertir et ainsi tester le rendu avec les configurations adaptées à la taille de la sortie (format A4, A3, A1, A0, etc.).

Quelques écrans :

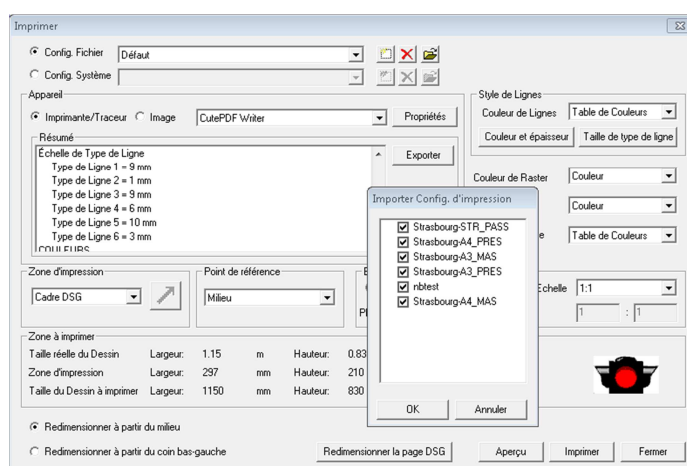


Figure 4. Chargement du fichier contenant l'ensemble des configurations du projet

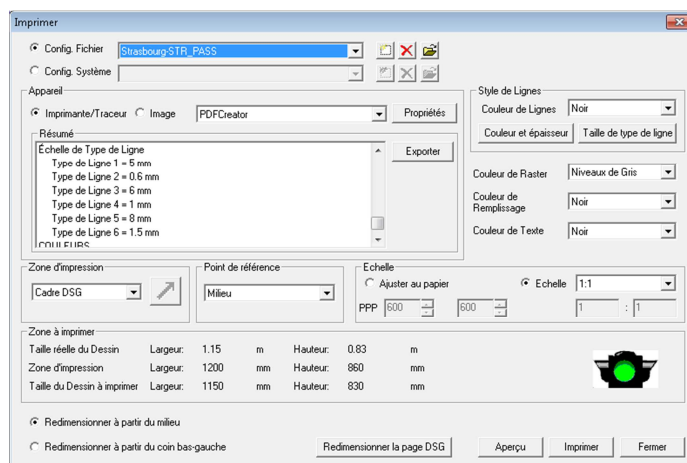


Figure 5. Choix de la configuration adaptée à la taille de la mise en page

Couleur originale	Couleur Épaisseur 1 à imprimer	Épaisseur 1 en mm	Couleur Épaisseur 2 à imprimer	Épaisseur 2 en mm
1 Arc+	Noir	0.25	Noir	0.60
2 Arc+	Noir	0.13	Noir	0.30
3 Arc+	Noir	0.13	Noir	0.30
4 Arc+	Noir	0.13	Noir	0.30
5 Arc+	Noir	0.25	Noir	0.30
6 Arc+	Noir	0.13	Noir	0.30
7 Arc+	Noir	0.18	Noir	0.30
8 Arc+	Noir	0.13	Noir	0.30
9 Arc+	Noir	0.30	Noir	0.30
10 Arc+	Noir	0.13	Noir	0.40
11 Arc+	Noir	0.30	Noir	0.30
12 Arc+	Noir	0.13	Noir	0.30
13 Arc+	Noir	0.13	Noir	0.30
14 Arc+	Noir	0.13	Noir	0.30
15 Arc+	Noir	0.13	Noir	0.30
16 Arc+	Noir	0.10	Noir	0.30
17 Arc+	Noir	0.10	Noir	0.30
18 Arc+	Noir	0.10	Noir	0.30
19 Arc+	Noir	0.10	Noir	0.30

Figure 6. Table des épaisseurs de traits et des couleurs

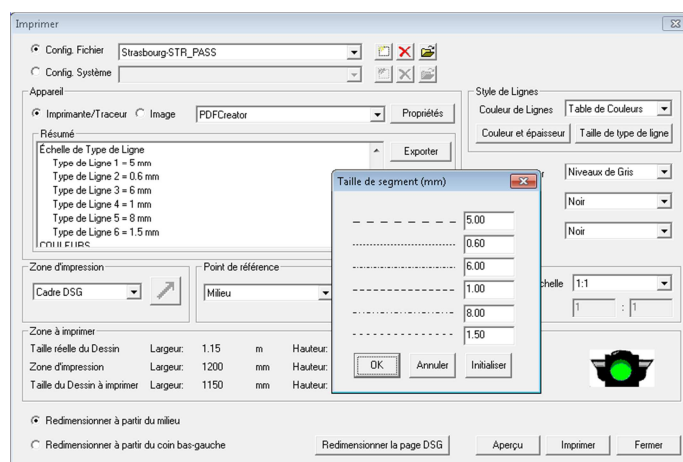


Figure 7. Taille des types de ligne

4.4.4 Fonction d'export DWG/DXF

Arc+ permet d'exporter le dessin technique en DWG ou DXF. Les paramètres d'export permettent de choisir les paramètres les plus adéquates avec l'objectif d'export et ainsi optimiser le fichier exporté afin de retrouver un fichier exporté exploitable et intègre. Par exemple, choisir de convertir les objets placés en blocs au lieu de les convertir en références externes permet de favoriser l'intégrité des fichiers exportés et réduit leur dépendance à d'autres fichiers.

Dans les paramètres d'export, il est important de bien cibler l'échelle des unités de travail pour obtenir un dessin à l'échelle dans AutoCAD, vérifier le type de conversion des objets Arc+ en entités DWG/DXF, vérifier l'adéquation des polices de caractères et vérifier la table des épaisseurs de traits en fonction de la couleur associée aux entités exportées.

On a opté ici pour l'export en DWG afin d'utiliser certains paramètres de configuration qui sont disponibles uniquement pour celui-ci. En effet, le paramètre le plus important concerne l'export de l'ensemble des fichiers liés et fichiers de bibliothèque sous forme de blocs au sein du fichier converti. Cette opération n'est possible que pour l'export DWG. Pour une conversion DXF, il faudra ainsi réaliser une seconde conversion, de DWG vers DXF, avec un logiciel le permettant efficacement.

Ont été privilégiés les paramètres suivants :

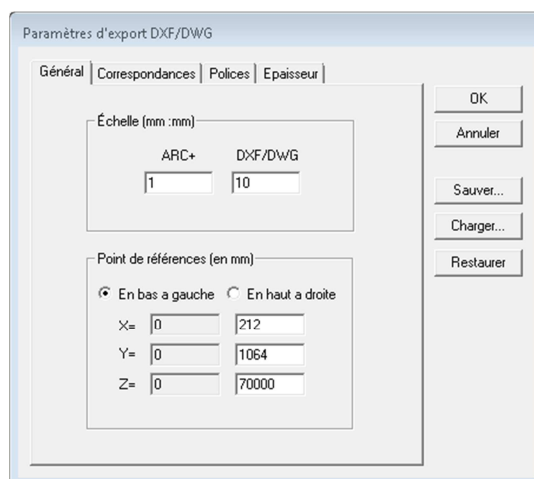


Figure 8 Onglet « général »

Afin de disposer des unités correspondantes aux dimensions affichées, pour les fichiers de type présentation (DSG), il a fallu mettre pour une unité Arc+ un équivalent de 10 unités DXF/DWG.

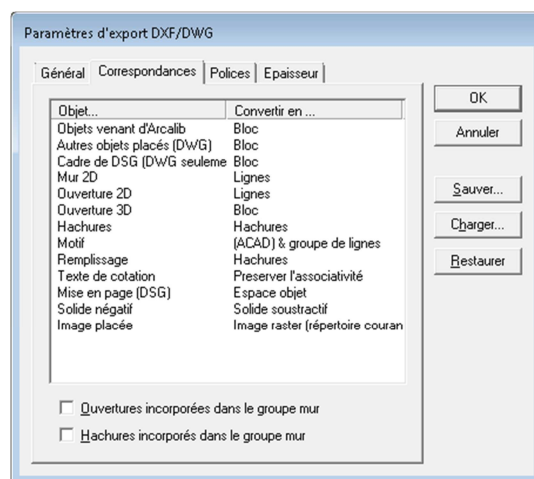


Figure 9 Onglet « Correspondances »

Afin de disposer d'un seul fichier DWG (certains paramètres ne s'appliquent pas pour le DXF) incluant l'ensemble des objets venant de la bibliothèque ARCALIB et l'ensemble des autres objets placés, il a été utile de convertir l'ensemble de ces objets en blocs inclus dans le fichier exporté, et non en références externes, pour éviter la génération d'autant de fichiers à l'export.

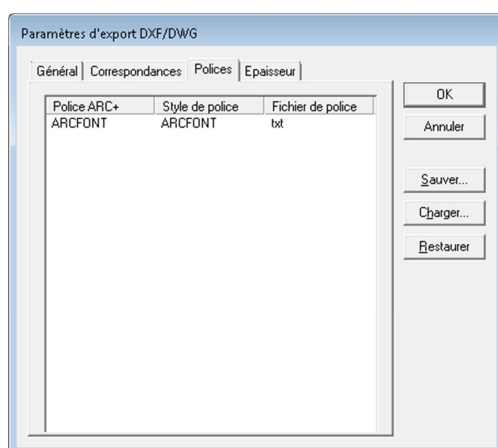


Figure 10 Onglet « Polices »

Les polices internes à Arc+ sont converties vers la police basique d'AutoCAD (txt.shx). Cette police est la police la plus basique et ainsi la plus proche de la police basique d'Arc+ (ARCFONT), qui est celle utilisée dans le fonds d'archives considéré.

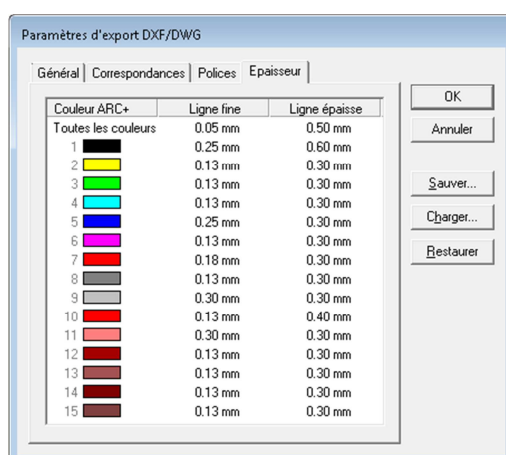


Figure 11 Onglet « Épaisseur »

Cette table d'épaisseurs de traits en fonction de la couleur peut être constituée à partir du fichier PAR adopté pour l'export PDF. Il est aussi possible de ne pas se préoccuper de cet onglet si l'on privilégie la création de fichiers de configuration AutoCAD (fichiers CTB) reprenant la configuration de chaque fichier PAR.

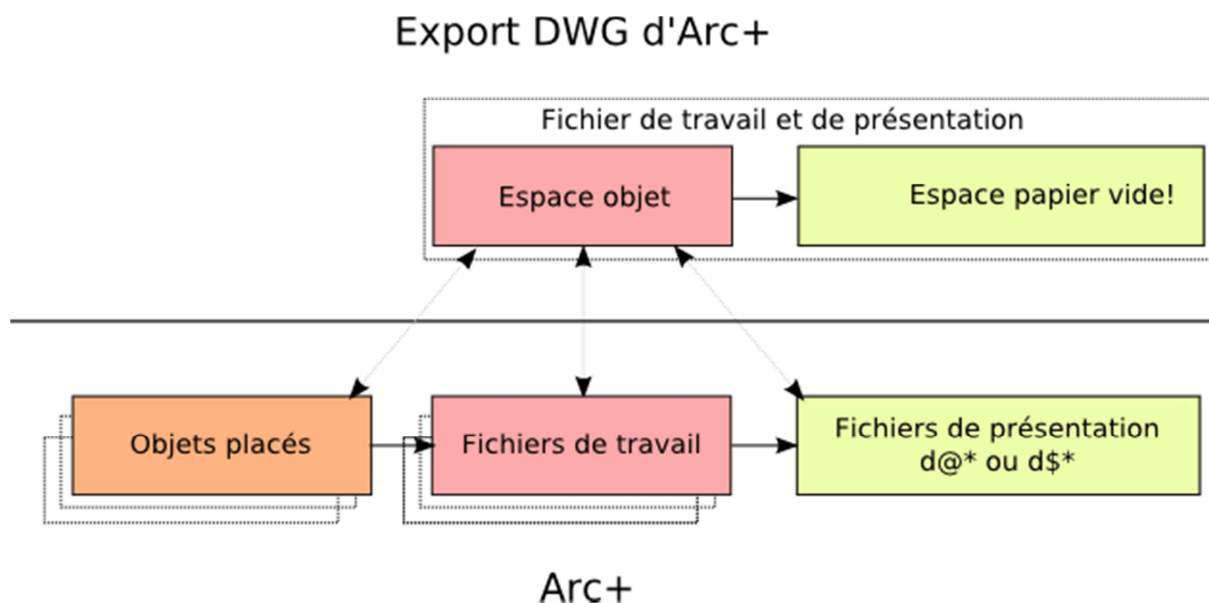


Figure 12. Résultat de l'export DWG d'Arc+ avec les options choisies

Comme on l'a vu, le résultat issu de cet ensemble d'options permet d'inclure l'ensemble des fichiers collaborant à la vue d'un fichier de présentation Arc+ dans un seul fichier AutoCAD au format DWG. Le seul effet de bord identifié dans le schéma ci-dessus est situé au niveau du moteur de conversion qui n'arrive pas à séparer les entités destinées à la mise en page (Espace Papier) des entités destinées à l'espace de travail (Espace Objet) d'AutoCAD.

4.4.5 Synthèse et nettoyage

À partir des fichiers du fonds, il s'est avéré impossible d'exporter dans le format DXF directement. En effet l'export génère un DXF qui à sa lecture au sein d'AutoCAD, n'affiche aucune donnée.

Comme il est détaillé plus bas, l'export au format DWG n'est pas au point non plus, mais il génère un fichier contenant les données du dessin.

Dans les deux cas, le logiciel génère un rapport d'export qui révèle certaines erreurs dans la lecture de certains objets placés d'Arc+.

Certains DWG exportés ne sont pas valides au moment de leur lecture sous AutoCAD. Toute tentative de modification s'est soldée par un échec lors de la sauvegarde. Toute tentative d'export depuis AutoCAD à partir d'un fichier converti en DWG vers le format DXF s'est terminée par un échec de sauvegarde à cause d'entités non valides dans le fichier d'origine. Les tests ont été réalisés sur plusieurs fichiers pour s'assurer que le problème ne provient pas d'un fichier de dessin particulier mais d'un problème de fond lié aux versions d'origine du logiciel (base des fichiers du fonds) et à la version d'Arc+ utilisée (utilisée pour la restitution des fichiers du fonds), la version 2012.

L'utilisation expérimentale du logiciel DraftSight³⁰ permet de résoudre le problème de lecture et de sauvegarde ; et permet ainsi de réaliser la conversion en DXF. Cependant certains problèmes dans le DWG exporté nécessitent un traitement manuel plus approfondi.

Les fichiers exportés d'Arc+ en DWG souffrent des problèmes identifiés suivants :

- L'adéquation des échelles des types de ligne : il sera systématiquement nécessaire de multiplier l'échelle globale des types de ligne par un facteur en liaison avec l'échelle de sortie (par exemple 100 pour une échelle 1/100^e).
- Certains textes se trouvent également réduits par un facteur de 100, sans qu'aucune raison n'ait été identifiée. Il sera nécessaire de multiplier la taille de ces textes par 100 par un traitement manuel car tous les textes n'ont pas la même taille (encore une fois, on peut supposer un lien avec l'échelle du fichier de présentation Arc+).
- La mise en page se trouve dans l'espace objet d'AutoCAD³¹ : si l'on choisissait d'exporter la mise en page dans l'espace papier, c'est tout le document qui se trouverait dans l'espace papier et aucune donnée dans l'espace objet. Ceci est-il une conséquence du choix de l'intégration des fichiers liés sous forme de blocs ? ou un dysfonctionnement de l'export des anciens fichiers Arc+ ? À la fin de cette étude, ces questions restent entières et un avis technique d'expert Arc+ sera nécessaire si une suite est donnée dans cette direction.
- La mise en page se trouve tronquée par un effet de masque inadapté dans AutoCAD. Le fait de décomposer le niveau le plus haut des blocs permet de retrouver l'ensemble des données concernant la mise en page en question, et parfois bien plus :
 - En effet, le masque de mise en page d'Arc+ a été transmis tel quel dans AutoCAD sans user des fonctionnalités de fenêtrage entre espace objet et espace papier.
 - Ce masque est malheureusement mal positionné à l'arrivée et oblige à le supprimer pour retrouver l'ensemble des données du dessin.

4.5 LES DOUBLONS

L'un des principaux points qu'a révélés l'étude, par rapport à celle de 2011, est l'extrême redondance des fichiers, sous des formes souvent difficiles à identifier : nombre d'entre eux possèdent un ou plusieurs doublons, portant le même nom ou non, situés dans des répertoires distincts, voire (avec des noms différents) dans le même répertoire. Il est apparu que leur identification et leur gestion peut devenir une étape indispensable – et longue – du traitement des archives numériques.

Un gros effort a donc porté, lors de la présente étude, sur la définition de méthodes pour identifier et ensuite pour traiter ces fichiers redondants. Heureusement dans le cas de ce fonds, les fichiers de présentation étaient bien identifiables par un préfixe distinctif (d\$*. * ou d@*. *) et limitaient ainsi la charge de travail.

³⁰ DraftSight, logiciel gratuit de DAO conçu par Dassault Systems pour la création et le travail sur des dessins techniques 2D. Il utilise le format DWG comme format natif et permet l'enregistrement au format DXF 2013 (ASCII) . La version utilisée a été la version 1.5. <http://www.3ds.com/fr/produits-et-services/draftsight/>

³¹ Se reporter aux annexes 1 et 2 du rapport de 2011 présentant succinctement et schématiquement le fonctionnement des fichiers dans Arc+ et dans AutoCAD.

En effet, pour le créateur des archives (l'architecte en cours d'exercice), il est très facile de garder de multiples copies des mêmes fichiers dans des répertoires différents et dans des « archives » compressées différentes afin de ne rien perdre. Cette habitude génère un nombre important de fichiers en double ou multiple exemplaires. Le fonds Fainsilber contient de nombreux zips contenant eux-mêmes souvent d'autres zips contenant des sous-ensembles entiers d'objets du fonds. Pour le traitement des fichiers dans l'objectif d'une pérennisation, il serait utile de ne traiter le fichier qu'une seule fois. Un travail préparatoire est nécessaire pour identifier les doublons ayant le même nom de fichier, la même extension et le même hachage³².

4.5.1 Repérage des doublons par hachage ou par comparaison de métadonnées

La société PASS Technologie³³ a utilisé une technique de hachage SHA2 afin de générer une empreinte unique par fichier, concaténée à sa taille et à son extension. Une mini-base de données SQLite, intégrant l'ensemble des empreintes ainsi générées, a été créée, permettant éventuellement d'exporter le résultat dans le format le plus adéquat pour inclure ces informations dans la base de données ArchiVecture. Mais il s'est avéré problématique d'intégrer ces données à la base alors que les niveaux hiérarchiques permettant de les lier aux Objets correspondant n'avaient pas été renseignés ; cela contredisait la logique de traitement et l'esprit de l'outil envisagé jusque-là (import des métadonnées au fur et à mesure du traitement, objet par objet, via un script intégré à la base de données et lancé directement au moment du traitement). Or le respect de ce processus est important car il garantit d'une part une indépendance fonctionnelle (outil intégré) et d'autre part l'actualisation des données (l'import se faisant à la date même du traitement).

Il n'a malheureusement pas été possible, dans le cadre de cette étude, de mettre en place un script en Visual Basic for Applications (VBA) capable de générer cette empreinte numérique³⁴.

Il y a lieu de revenir sur cette question, quitte, s'il s'avère impossible de générer le SHA2 via VBA avec les moyens actuels du centre d'archives, à repenser aux moyens de récupérer les empreintes générées de façon externe à la base.

Le seul protocole d'identification des doublons à partir de la base est donc actuellement la comparaison de la taille et de l'extension des fichiers, qui est bien sûr une indication incomplète.

4.5.2 Analyse des doublons à partir du hachage

Une petite base de données Access indépendante, créée³⁵ à partir de l'export fourni par PASS Technologie mentionné plus haut, a cependant permis une analyse des doublons dans l'ensemble du fonds, notamment ceux contenus dans les fichiers de bibliothèque.

³² Somme de contrôle obtenue par combinaison de plusieurs paramètres du fichier pour être unique pour chaque fichier. Techniques : Md5, SHA1, SHA2, ...

³³ Sur ce point, Pierre Dittgen et Julien Bellion (PASS Technologie) ont apporté leur contribution suite à une spécification précise avec Nader Boutros.

³⁴ Recherches de Sonia Gaubert avec l'appoint de Julien Bellion et de Nader Boutros.

³⁵ Par Sonia Gaubert, octobre 2013.

Il en ressort que le fonds comprend 126199 empreintes en doublon pour un total de 182223 fichiers (auxquels s'ajoutent environ 59000 dossiers qui se trouvent dans les fichiers zippés et qui n'ont pas été analysés à cette occasion) ; on pourrait en déduire que le fonds contient environ 120000 fichiers distincts (le total moins la moitié du nombre de fichiers ayant des empreintes en doublon), mais il existe parfois plus de deux doublons d'un même fichier, ce qui rend le calcul très approximatif.

Sur ces 126199 doublons, 6004 proviennent du répertoire FAINAD-A-2 (répertoire des bibliothèques, bibliothèque agence, contenant en tout 8928 fichiers), tous les autres sont dans des répertoires de projets.

On remarque au passage l'existence de 100 fichiers qui ont une taille nulle (0 octet), certainement à supprimer, ainsi que de nombreuses occurrences qui n'ont pas de nom, seulement désignées par une extension, comme par exemple :

RefSign	dossier	fichier	taille
32675	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A13_PRM\ARC13\ARCALIB\Voitures	.ddd	2645
32675	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A14_PRM\arc14\ARCALIB\Voitures	.ddd	2645
56166	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A14_PRM\arc14\ARCALIB\Voitures	.icn	4102
56166	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A13_PRM\ARC13\ARCALIB\Voitures	.icn	4102

On remarque aussi que les équivalences (sous-répertoires/noms de fichiers) dans les différentes versions des bibliothèques sont nombreuses mais non systématiques (par exemple, deux fichiers identiques peuvent ne pas porter le même nom).

Exemples :

RefSign	dossier	fichier	taille
709	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A11_PRM\ARCPLUS\ARCALIB\SANITARY	SINK6.DDD	1025
709	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A13_PRM\ARC13\ARCALIB\SANITARY	SINK6.DDD	1025
709	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\A14_PRM\arc14\ARCALIB\SANITARY	SINK6.DDD	1025

RefSign	dossier	fichier	taille
2	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\SOFT_DOS\CADREN	CADDRVS.BAT	1
2	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\SOFT_DOS\CADREN	INSTGS.BAT	1

Certains fichiers présents dans les bibliothèques se retrouvent également dans les répertoires projets, parfois avec des noms différents.

Exemples :

RefSign	dossier	fichier	taille
507	Z:\FAINAD\FAINAD-A-2\CD Agence bibliothèque ARC+ versions 11, 13\SOFT_DOS\ARC.V73\ARCA\USERDEF	POP5.POP	1016
507	Z:\FAINAD\FAINAD-B-1992-2\CD IMM 11-09-1998\IMM	POP5.POP	1016
507	Z:\FAINAD\FAINAD-B-1992-2\CD IMM 11-09-1998\IMM	IMM5.POP	1016
507	Z:\FAINAD\FAINAD-B-1992-2\CD IMM 11-09-1998\IMM_COUP	POP5.POP	1016
507	Z:\FAINAD\FAINAD-B-1992-2\CD IMM DOE Avril 2000\BDD	POP5.POP	1016
507	Z:\FAINAD\FAINAD-B-1992-2\CD IMM DOE Avril 2000\BDD	IMM5.POP	1016

Mais ces doublons ne sont pas nécessairement des homonymes, et les homonymes ne sont pas nécessairement des doublons au sens numérique.

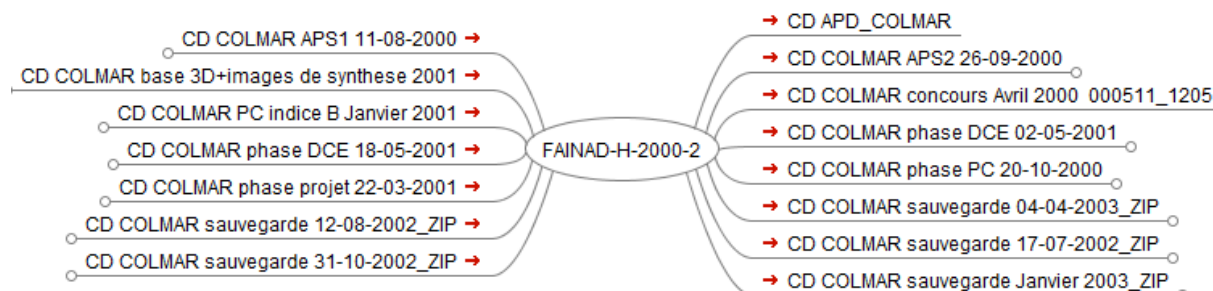
Le fonds contient en effet 151494 fichiers homonymes d'un autre, qui se trouvent parfois dans des répertoires projets différents, avec ou non la même signature numérique et la même taille.

Cette première exploration suggère de se débarrasser des doublons de fichiers de travail et des fichiers de présentation (contenant @ ou \$), mais il est impossible, à moins de l'ouvrir, de distinguer un fichier de travail d'un fichier lié (à l'exception de certains fichiers caractérisés par leur extension, de type configuration des plumes ou autre), et il serait difficilement envisageable de supprimer arbitrairement tel ou tel fichier présent dans un sous-répertoire sous prétexte qu'il double un autre enregistré ailleurs.

La seule démarche possible en amont de l'étude détaillée d'un projet est donc de repérer les redondances de fichiers de l'arborescence principale et de celles des fichiers archives rencontrés de type ZIP.

4.5.3 Un exemple : les doublons des sauvegardes (zip) du projet de Colmar

Les archives numériques concernant l'hôtel du département du Haut-Rhin, à Colmar, ont été remises par l'agence sous la forme de 11 CD (voir *infra* la liste des CD). Cinq d'entre eux contenaient chacun un zip unique qui représente une sauvegarde des fichiers du projet à un moment donné.



Un export³⁶ réalisé à partir d'un script Python a permis d'analyser ces 5 CD, intitulés comme suit :

- CD COLMAR sauvegarde 17-07-2002_ZIP
- CD COLMAR sauvegarde 12-08-2002_ZIP
- CD COLMAR sauvegarde 31-10-2002_ZIP
- CD COLMAR sauvegarde Janvier 2003_ZIP
- CD COLMAR sauvegarde 04-04-2003_ZIP

Il s'agissait de parcourir leur contenu et de les comparer pour identifier les différentes versions de chaque fichier dans les 5 CD. En effet, l'arborescence des fichiers au sein de chaque zip étant très proche, nous pouvons conclure qu'il s'agissait de sauvegardes périodiques de l'ensemble du projet.

Le résultat de cette analyse est présenté en annexe « Cf. § 9.6 **Analyse des 5 CD de Colmar** » sous forme d'un tableau fusionnant l'arborescence des 5 CD et présentant les variations de la signature numérique de chaque fichier présent dans chaque CD.

Cette démarche donne une piste pour le traitement de ces 5 CD en les fusionnant temporairement avec l'arbre principal de l'objet afin d'optimiser le temps de traitement de l'objet.

4.5.4 Les doublons dans les fichiers de bibliothèque (ARCALIB)

Les fichiers de bibliothèque (cf. § 4.4.1) sont présents sous la forme de plusieurs arborescences similaires dans le fonds d'archives. La plupart de ces fichiers étaient fournis par le logiciel Arc+, mais, petit à petit, l'agence a intégré à cette bibliothèque des éléments répétitifs qui lui étaient spécifiques (trame, portions de plans réutilisables, cartouches, etc.). Pour chaque version successive du logiciel, elle a centralisé une version de la bibliothèque d'Arc+ incluant ces compléments spécifiques. Aucun lien spécifique n'a été repéré entre un projet donné et une de ces bibliothèques.

Afin de pouvoir identifier une seule bibliothèque de référence, il a fallu procéder à la fusion de ces bibliothèques, ce qui passait par l'analyse des homonymes ayant des signatures numériques différentes. À la fin de cette procédure, on conserve les doublons quand ils sont dans des branches différentes de l'arborescence, mais on fusionne ceux d'un même répertoire. Seuls 6 fichiers avaient des signatures numériques différentes pour des noms de fichier identiques. Après exploration de leur contenu, la version la plus récente a été privilégiée (aucune différence graphique n'a été détectée). Pierre Dittgen a procédé à la création d'un script respectant cette procédure et permettant ainsi d'aboutir à la fusion des bibliothèques.

4.5.5 Conclusion : principes de gestion des doublons

Ces analyses amènent à **distinguer**, parmi les divers cas de fichiers identiques ou analogues, « **vrais** » **doublons** (noms identiques ou non, empreintes numériques identiques) et « **faux** » **doublons** (noms identiques, empreintes numériques distinctes).

Le principe de la gestion des doublons consiste à garder un exemplaire de chaque version avérée comme distincte des autres, qu'il s'agisse de « faux doublons » ou de vrais doublons appartenant à des « phases » du projet différentes (ceci afin de conserver la structure d'origine des dossiers et de ne pas les amputer d'éléments contribuant à leur compréhension générale : on ne supprimera pas un

³⁶ Constitué par Pierre Dittgen en concertation avec Nader Boutros.

plan faisant partie du dossier de permis de construire sous prétexte qu'il est identique à un autre dans le dossier d'appel d'offres...).

On partira de la dernière copie de l'ensemble des fichiers concernant un projet donné (celle qui n'est pas zippée, donc a priori la copie la plus récente). Le système devra repérer, pour chaque fichier, les « vrais » et les « faux » doublons. Le chemin (racine) où se trouve enregistré le fichier sera un élément important pour en déterminer le traitement :

Les « vrais » doublons, s'ils font partie d'un ZIP (ou d'une copie de sauvegarde) et de chemins identiques, seront ignorés ; s'ils font partie d'une phase distincte, seront recopiés dans les deux phases ;

Les « faux » doublons (mêmes noms, empreintes différentes) seront tous conservés et classés par phases.

4.6 SÉLECTION DE FORMATS DE PÉRENNISATION

Dans le cas du fonds Fainsilber et des fichiers majoritairement liés au logiciel Arc+, et au vu des outils identifiés pour la conversion (détaillés ci-dessous), il est proposé de convertir chaque fichier de présentation (nom de fichier Arc+ contenant un @ ou un \$) dans deux formats distincts (cf. § 3.4.1 et 3.4.2) :

- DXF (qu'on appellera « format d'utilisation ») (plutôt que DWG)
- PDF/A (qu'on appellera « format de présentation »)

Dans le fonds et dans la base de données, pour un document Arc+ de présentation (déjà défini par au moins 2 fichiers « .iii » et « .ddd »), il y aura au moins deux, voire trois, fichiers liés à la conversion : un fichier PDF, un fichier DXF et éventuellement (s'il apparaît utile de le conserver) un fichier RPT de rapport de conversion généré par Arc+.

4.7 MODALITÉS DES EXPORTS VERS PDF ET DXF

Il est nécessaire d'évaluer les performances des processus de conversion des fichiers d'origine vers les formats cibles. Il faut d'emblée souligner qu'Arc+ ne génère pas directement d'exports au format PDF.

De même, pour l'export DXF, il faudra passer en premier par un export au format DWG puis réaliser un minimum d'opérations de nettoyage de ce dernier pour enfin le sauvegarder en DXF.

4.7.1 Exports vers PDF

En reprenant les catégories de l'étude sur les outils de conversion PDF (CINES/SIAF)³⁷, on peut identifier quatre types, dont trois se retrouvent utilisés dans la présente recherche, tandis qu'un cinquième type, absent de l'étude du CINES, semble se dégager :

Type 1 : Logiciel → export PDF → PDF

Dans le logiciel AutoCAD, à partir d'un fichier au format DWG, il est possible de demander un export en PDF. Aujourd'hui, le PDF généré est en version 1.6, qui est proche du PDF/E-1 et intermédiaire

³⁷ http://www.huma-num.fr/sites/default/files/ressourcesdoc/guide_methodologique_formatpdf_partie2.pdf.

entre PDF/A-1 et PDF/A-2. AutoCAD exporte la liste des calques, ce qui permet leur activation/désactivation à partir d'un lecteur PDF compatible.

Type 2 : Logiciel → convertisseur → PDF

Cas des convertisseurs de fichiers DWG indépendamment du logiciel AutoCAD. Les convertisseurs testés par le CINES sont :

- DWG to PDF Converter (<http://anydwg.com/dxf-to-pdf-ex.html>)
- Convert DXF to PDF <http://www.coolutils.com/dxf-to-pdf>
- AutoDWG <http://www.autodwg.com/dxf-to-pdf.htm>
- DWG to PDF Converter MX <http://www.dwgtool.com/dwg-to-pdf.htm>

Ces convertisseurs nécessitent l'achat de licences, ce qui n'a pas été envisagé dans cette recherche. Quelques tests non concluants ont été réalisés avec des versions de démonstration.

Type 3 : Logiciel → impression → pilote d'impression → PDF

Exemple : PDF Creator (<http://www.pdfforge.org/pdfcreator>) — pilote d'impression Windows qui génère du PDF. [Gratuit.]

Attention, lors de cette impression virtuelle en PDF de documents techniques DAO, il est important de respecter toutes les procédures d'impression (choix des couleurs, épaisseurs des traits, échelle des types de lignes, échelle d'impression, zone d'impression et format de papier, entre autres).

Dans cette étude, PDF Creator a été utilisé comme imprimante virtuelle d'Arc+. Lors de l'enregistrement, le profil PDF/A-2b a été systématiquement choisi.

« Type 5 » : Logiciel → Fichier d'impression .plt (HPGL/2) → visionneuse/convertisseur → PDF

Cas du logiciel Arc+ à partir d'un fichier .iii tracé dans un fichier de traçage au format PLT. Ces fichiers PLT, s'ils sont disponibles dans les archives, assurent une fidélité totale aux rendus de l'agence. Ils incluent toutes les informations utiles pour l'impression et ne dépendent d'aucun autre fichier. Ils étaient créés en vue de faire imprimer des plans sur un traceur à l'extérieur de l'agence, et n'existent donc en principe que jusqu'à un certain moment de l'histoire de l'agence et de son équipement.

D'autres logiciels de visualisation et de conversion des fichiers PLT existent. Après quelques expérimentations, ViewCompanion³⁸ se dégage par la simplicité et la fluidité de son interface et le respect des standards lors de l'export (profil PDF/A-2b).

4.7.2 Exports vers DXF

Pour la conversion en DXF 2013, la solution la plus efficace consiste en quelques étapes systématiques :

- export DWG depuis Arc+,
- ouverture du DWG avec DraftSight,
- nettoyage du DWG (cf. § 4.5.5),
- sauvegarde en DWG et DXF 2013 (ASCII)

³⁸ <http://www.softwarecompanions.com/viewcomp.html>

4.7.3 Futurs formats PDF

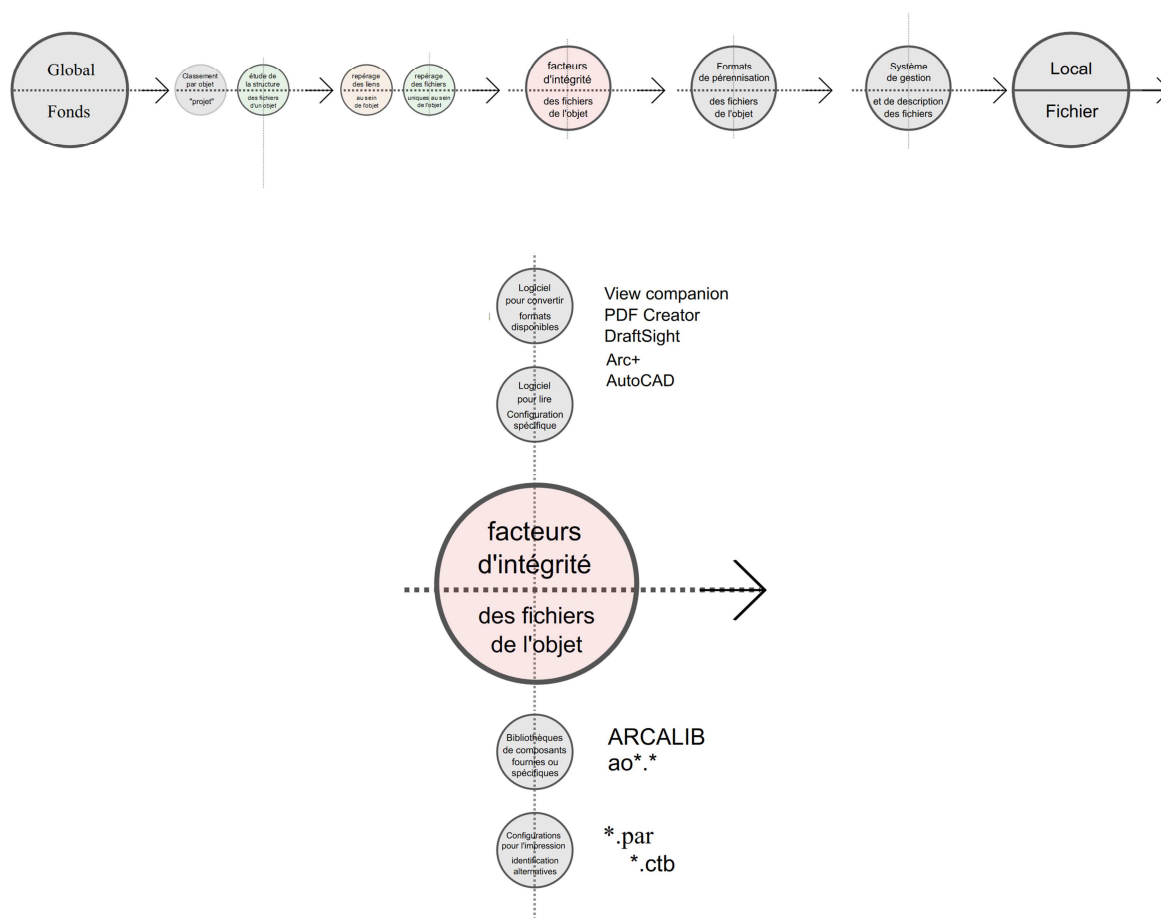
Dès que des outils d'assemblage PDF/A-3 seront disponibles, si d'autres formats comme le DXF s'avèrent importants pour garantir la pérennité des données, ils pourront, d'après la spécification, être inclus dans le PDF/A-3. Les deux formats DXF et PDF peuvent ainsi être présents dans un même conteneur PDF. Les outils de vérifications devront être capables d'analyser le conteneur et chaque contenu. Le risque existe cependant de trouver au sein des fichiers PDF/A-3 des formats non pérennisés.

Le PDF/E-2 est aussi un autre candidat intéressant pour la pérennisation, conservant ainsi dans un format standard à la fois les données pour la réutilisation et pour la présentation des documents. La partie technique est plus orientée vers la préservation de données 3D.

Avec des outils de conversion et de vérification par rapport à sa spécification, le format PDF/E-2 pourra être privilégié à l'avenir avec la généralisation des maquettes numériques du bâtiment.

4.8 PROCÉDURE COMPLÈTE DE CONVERSION

Une synthèse récapitulant l'ensemble de la procédure de conversion des fichiers Arc+ (DDD et III), AutoCAD (DWG) et PLT est présentée en annexe (§ 9.9 Procédure de conversion des fichiers Arc+ et PLT). Cette annexe reprend les différents points concernant la conversion en fichiers pérennes des fichiers natifs ou d'impression virtuelle issus d'Arc+ et les DWG issus d'export Arc+ ou natifs.



5. SYSTÈME DE GESTION ET DE DESCRIPTION DES FICHIERS

5.1 LA BASE DE DONNÉES DU CENTRE D'ARCHIVES

ArchiVecture est la base de données qui gère l'intégralité des activités du Centre d'archives (inventaires³⁹, bibliothèque, consultations, prêts, logistique, etc.). Développée sous Access, en interne (par Sonia Gaubert), depuis 1997, elle est également la base source qui alimente l'application en ligne ArchiWebture (base de données des inventaires : <http://archiwebture.citechaillot.fr/>).

Comme indiqué plus haut (§ 2.4), l'organisation des fichiers numériques (documents numérisés ou, plus rarement, archives numériques natives) sur le serveur est directement liée à la structuration des inventaires en base de données, dont elle reprend les niveaux hiérarchiques utiles (Fonds / Objets / Phases / Dossiers) ; le système de référencement utilisé est également basé sur celui des inventaires (voir aussi en annexe § 9.10 **Gestion et description des archives numériques : rappels**⁴⁰).

Lors de la présente mission, un module spécifique a été développé dans la base de données, dédié au traitement des archives informatiques natives, qui utilise largement ces concordances dans son processus.

5.2 ORGANISATION DES FICHIERS NUMÉRIQUES SELON L'ARBORESCENCE DE LA BASE DE DONNÉES

5.2.1 Rappel et mise en cause des principes définis en 2011

L'étude de 2011 prévoyait de déplacer tous les fichiers numériques sur le serveur de fichiers du Centre d'archives (dédié aux fichiers images), dans des répertoires organisés selon une arborescence correspondant à celle des inventaires eux-mêmes. On créait ainsi une correspondance entre ces répertoires et les niveaux de la base de données, les noms des répertoires utilisant d'ailleurs les identifiants spécifiques aux éléments désignés. La hiérarchie était la suivante :

- Répertoire Fonds (identifié par l'abréviation du fonds, ici *FAINAD*)
 - Répertoire Objet (identifié par la référence objet : abréviation – lettre désignant le chapitre de l'inventaire – date de début du projet – numéro d'ordre par rapport à l'année. Exemple : *FAINAD-C-1992-3*),
 - Répertoire Phase (identifié par le terme « Phase » suivi du numéro d'ordre de la phase. Exemple : *Phase-2*)
 - Répertoire Dossier (identifié par la cote dossier : numéro d'enregistrement du fonds suivi de la cote physique de l'article. Exemple : *401-lfa-532*)

³⁹ Sur le module des Inventaires de cette base, sa genèse et ses grands principes, voir « Principes de classement et instruments de recherches au Centre d'archives de l'ifa », *La Gazette des archives* (Association des archivistes français), n° 190-191, 3^e et 4^e trimestres 2000, p. 255-272 (http://portaildocumentaire.citechaillot.fr/userfiles/file/Centre-archives/4-principesdeclassification_a9135.pdf).

⁴⁰ Cette note n'intègre actuellement pas les développements réalisés notamment en 2014.

Par ailleurs, comme tous les fichiers de CAO-DAO, **les fichiers Arc+ sont répartis en fichiers de travail et fichiers liés (objets placés)**. Les informations sur les relations liant ces fichiers sont visibles à l'ouverture des fichiers dans Arc+, et peuvent être exportées (commande \olist) sous forme d'un rapport (fichier texte) portant l'extension .rpt.

Le processus de traitement préconisé impliquait les étapes suivantes :

- ouverture de chaque fichier « de présentation » (nom en d@ ou d\$) pour identification du document (prise de note pour description), génération du fichier RPT identifiant les fichiers liés, génération du fichier de pérennisation (au format PDF ou autre),
- déplacement de ces fichiers (fichier original, RPT et fichier au format pérenne) vers le répertoire de classement définitif (c'est-à-dire dans un sous-répertoire Dossier du sous-répertoire Phase concerné),
- lancement d'un programme Java autonome (.jar), développé dans le cadre de cette mission, ayant pour fonction, à partir des informations contenues dans les fichiers RPT, de recopier les fichiers liés de leur(s) emplacement(s) d'origine vers le répertoire de classement définitif.

Enfin, à l'issue du traitement de tous les fichiers d'un projet donné, un « nettoyage » (suppression des doublons) devait être entrepris.

Ces préconisations ont été totalement remises en question lors de l'étude menée en 2013, pour plusieurs raisons : outre l'analyse plus fine des doublons qui a pu être faite alors (voir ci-dessus § 4.5), on a mieux compris les conséquences de la structure des bibliothèques, qui rendent le programme JAR moins opérant que prévu (cf. § 4.4.1).

5.2.2 Une conversion partielle, uniquement pour les fichiers de présentation...

Le rapport de juillet 2011 mettait déjà en évidence que certains fichiers ou types de fichiers – au-delà des doublons proprement dits – n'ont pas à être pérennisés, notamment dans le logiciel Arc+ qui crée trois fichiers numériques (avec des extensions distinctes) par document. En 2013, les indications et précisions sur la nomenclature apportées par Yves Chauvin, l'architecte informaticien de l'agence Fainsilber⁴¹, ont permis de préciser que **seuls les fichiers de présentation nécessitent une conversion** – ce qui a des conséquences importantes sur le volume des archives à traiter. Cette conversion permet de fusionner, dans des formats pérennes, l'ensemble des fichiers liés à ces fichiers de présentation.

Les fichiers de présentation, dans le fonds Fainsilber, sont reconnaissables par leur dénomination car leurs noms commencent tous par les identifiants d@ ou d\$. Ils contiennent donc la plupart des fichiers liés. Selon la définition d'Yves Chauvin, **ces fichiers liés se composent d'« objets placés »** (documents de bibliothèque et/ou de travail au niveau du projet dont les noms commencent par le caractère « o » : « o*.ddd, o*.iii ») **et de « fichiers de dessins »** (morceaux de plans dont les noms commencent par les caractères aos, aop, aoc, akh).

En conséquence de cette conversion partielle qui part des fichiers de présentation, si des « fichiers de dessin » (noms commençant par « a.. ») ne sont intégrés (liés) à aucun fichier de présentation, ils ne seront pas convertis. Il n'y a aujourd'hui aucun moyen de les identifier, à moins d'un très lourd travail de pointage manuel (sans objet au regard de l'enjeu et du coût), mais il semble que l'abandon

⁴¹ Contacts entre Yves Chauvin et Nader Boutros début octobre 2013, mail d'Yves Chauvin du 8 octobre 2013.

de ces fichiers non intégrés à une présentation soit sans conséquences sur la compréhension et sur le caractère complet des archives.

5.2.3 ...Mais une double conversion

Seule l'émulation de l'environnement de travail d'origine garantirait la parfaite adéquation des contenus avec ceux des fichiers d'origine. Mais la mise en place d'un tel environnement ne peut se concevoir que si l'architecte a fait l'effort de conserver les informations nécessaires à l'émulation de son environnement de travail, ce qui n'est sans doute jamais le cas.

Face à ce manque d'informations, il faut identifier les perspectives d'usage prévisibles pour les archives numériques. Deux usages possibles s'avèrent importants et complémentaires :

1. L'exploration des documents *finis, exportés* ou *communiqués* à l'extérieur de l'agence (qui devrait suffire pour la recherche en histoire de l'architecture ou dans d'autres disciplines),
2. La réutilisation (à des fins d'intervention sur l'existant par exemple).

Les fichiers retenus pour la pérennisation seront donc convertis dans deux formats distincts : l'un en PDF pour la consultation (« format de présentation », § 3.4.1) et l'autre en DXF pour une éventuelle réutilisation (« format d'utilisation », cf. § 3.4.2).

En outre, les fichiers d'origine sont conservés dans leur format originel.

5.2.4 Une réorganisation seulement partielle des fichiers

Au lieu de tenter de réorganiser entièrement les fichiers d'archives natifs par Phases/dossiers spécifiques, il apparaît préférable de **préserver l'intégrité organisationnelle des archives numériques originelles (es fichiers sources)**, en les regroupant, pour chaque projet, dans un seul dossier, toujours coté⁴² 10000. Ainsi, quel que soit le projet, tous les dossiers contenant les fichiers sources seront enregistrés, à l'intérieur du répertoire objet correspondant à ce projet, dans un sous-répertoire dossier intitulé « [Numéro du fonds]-10000 », lui-même placé dans un sous-répertoire « Phase-0 » (correspondant à la phase intitulée dans la base « Archives numériques : fichiers sources »).

(Ainsi, le chemin pour les fichiers sources du projet qui reçoit la référence objet FAINAD-C-1992-3 est I: / FAINAD / FAINAD-C-1992-3 / Phase-0 / 401-lfa-10000)

Au-delà de la simplification de gestion apportée par ce principe, cela allégera considérablement le poids des archives elles-mêmes. Ne seront, en effet, reclassés dans des répertoires phases (phases du projet : Conception, Exécution, etc.) et, à l'intérieur de ceux-ci, dans des dossiers spécifiques (cotés 10001, 10002 et ainsi de suite) que les fichiers résultant de la conversion, c'est-à-dire ceux qui seront directement consultables et réutilisables.

Les fichiers de bibliothèques, d'usage transversal par définition, seront quant à eux conservés dans une phase attachée à l'objet FAINAD-A-2 (Documentation professionnelle de l'agence). Il y aura lieu d'opérer la fusion de ces bibliothèques dans une autre phase du même objet (vers laquelle Arc+

⁴² Afin de bien distinguer ce type spécifique d'archives, les dossiers d'archives numériques reçoivent, à l'intérieur de chaque fonds, des cotes commençant au numéro 10000. Comme l'indique la suite du paragraphe, ce numéro est réservé aux fichiers sources, les numéros à partir de 10001 constituent les cotes des dossiers contenant les fichiers issus de la conversion.

pointera, au moins le temps du traitement) pour conserver, pour ces fichiers-là aussi, un état originel de l'ensemble.

NB : PASS Technologie a réalisé la fusion des fichiers de bibliothèques et a placé le dossier ARCALIB résultant à la racine de l'objet FAINAD-A-2 en attendant le sort de cet objet dans la base de données.

5.3 MODULE DE TRAITEMENT DES ARCHIVES NUMÉRIQUES DANS LA BASE DE DONNÉES

Cf. plus loin § 5.3.6, qui récapitule les étapes du traitement d'un dossier, et, en documentation externe, « **Manuel ArchiVecture : Guide de saisie** » (décrit en § 9.11), qui contient une procédure détaillée illustrée de nombreuses copies d'écran.

Aux éléments mentionnés jusqu'ici s'ajoute la nécessité d'importer et de gérer les métadonnées des fichiers numériques. Pour toutes ces fonctions, un module spécifique destiné au traitement des fichiers numériques a été créé dans la base de données ArchiVecture.

The screenshot shows the 'Métadonnées des fichiers liés à l'objet' interface in Microsoft Access. The main title is '1988-1998. Musée d'art moderne et contemporain, Strasbourg (Bas-Rhin). FAINAD-E-1988-2'. The interface includes a navigation pane on the left, a menu bar at the top, and a main data entry area. The data area contains a table of files with the following columns: Nom, Taille, Format, Date, Date Modif., Date copie, Dossier source, Dossier destin., Statut, Dossier, Format conv., and Nom conv. The table lists several files, including 'DS1ER1', 'DS1ER2', 'DS1ER200', and 'DS1ER200'. Below the table, there is a 'DOUBLONS' section for duplicate entries, showing a file named 'IER' with a size of 290169. The interface also includes a search bar and a 'Rechercher' button.

5.3.1 Import des métadonnées dans le formulaire général

Le formulaire de gestion des métadonnées est accessible depuis toute fiche Objet.

Il rappelle les dates, titres et référence de l'Objet et comporte six boutons :

- Le bouton « Créer Phase/Dossier + Import métadonnées »,
- Trois boutons permettant d'ouvrir les répertoires Objet, Phase-0, et Dossier source (Num du fonds-10000),
- Un bouton permettant l'accès au formulaire Phases,
- Le bouton « Déplacer fichiers convertis », à actionner en fin de traitement (cf. 5.3.5).

Pour chaque objet traité successivement, le placement des fichiers sources dans le sous-répertoire NumFonds-10000 (lui-même situé dans le sous-répertoire Phase-0) se fait manuellement : cette étape est en effet difficile à automatiser étant donné que ces dossiers portent des noms variables (et que des fichiers et des dossiers peuvent préexister dans les répertoires objets).

Le bouton « Créer Phase/Dossier + Import métadonnées » vérifie l'existence du répertoire Objet dans lequel on se situe (avec possibilité de le créer automatiquement s'il n'existe pas), puis de répertoires préexistants dans le répertoire Objet ; il vérifie aussi l'existence de leurs équivalents dans la base (niveaux Phases et Dossiers) et propose de les créer s'ils n'existent pas (répertoires et enregistrements Phases et Dossiers). Il est conseillé d'utiliser ce bouton pour créer les répertoires initiaux, afin de garantir une écriture correcte.

L'importation des métadonnées pour un objet donné ne se fera donc que si toutes ces occurrences existent, et, a contrario, seules seront importées les métadonnées des fichiers n'existant pas déjà dans la table (pour faire ce contrôle, le système utilise le champ contenant le chemin, nom du fichier lui-même compris).

5.3.2 Sous-formulaire des métadonnées

La table des métadonnées comprend les champs suivants :

Légende :

* : Les noms suivis d'un astérisque sont ceux des champs dont les valeurs sont récupérées automatiquement par le programme depuis les feuilles de propriété des fichiers

Italiques : champs non apparents dans le sous-formulaire

Soulignés : champs à informer obligatoirement en cas de conversion

RefMetad : numéro automatique (clé primaire),

IdFichier* : texte (50 car.), nom du fichier,

RefObjLie : numérique (entier long avec doublons) lien avec la table InvObjets,

RefDosLie (étiquette : « Dos. source ») : numérique (entier long avec doublons), référence du dossier source - lien avec la table InvDossiers,

CheminFich* : texte (255 car.), chemin du fichier (y compris nom du fichier),

DateCopie* : date de la copie (format abrégé : jj/mm/aaaa),

DateDernModif* : date de la dernière modification (format et source : idem),

FormatFich* : texte (35 car.), format du fichier (intègre automatiquement le terme « Fichier » suivi de l'extension => resterait à supprimer le terme « fichier » un peu superflu),

TailleFich* : numérique (entier long), taille du fichier,

RefDosDest (étiquette « Dos. destinat. ») : numérique (entier long avec doublons), permettant d'indiquer la cote du dossier où sera enregistré le fichier converti (contient la liste des références numériques des cotes dossiers déjà créés pour l'objet en cours : la liste se met à jour automatiquement),

Statut : numérique (entier), zone de liste avec valeurs « Fichier source » et « Fichier converti »,

FormatConv : texte (6 car.) champ de liste précisant le format de conversion contenant actuellement quatre valeurs correspondant aux extensions des formats les plus courants (.pdf, .tif, .txt, .dwg). Le contenu est limité à la liste (pourra être enrichie...) et sera utilisé par le programme pour identifier le fichier à déplacer après conversion,

NouvidFichier (étiquette : « Nom conv. ») : texte (20 car.), nom du fichier converti. Doublet interdit. Champ obligatoire en cas de conversion (utilisé par le script qui déplace les fichiers convertis). La nomenclature utilise la date et l'heure du moment de la conversion (AAAA-MM-JJ-HHMM)⁴³. Le bouton portant l'icône crayon à droite du champ remplit automatiquement le champ.

SignalDouble : case à cocher, mise à jour automatiquement lors de l'import si le programme repère des fichiers ayant les mêmes noms et tailles que ceux préexistants dans la table des métadonnées, mais dont le chemin est différent. Cela peut correspondre à deux situations distinctes :

- 1) un état initial dans les archives (même fichier recopié dans des sous-répertoires distincts),
- 2) une modification accidentelle du contenu des répertoires et sous-répertoires sources par rapport à un premier import : ceci ne devrait pas se produire dans la base (des sécurités existent à différents niveaux – niveaux Objets, Phases, Dossiers... – pour que l'on ne supprime pas malencontreusement des enregistrements auxquels seraient liées des métadonnées), mais il est quasiment impossible d'empêcher les manipulations directes et malencontreuses sur le serveur. On pourrait concevoir, dans ce cas, de supprimer automatiquement les métadonnées plus anciennes dont les chemins (nom du fichier compris) ne correspondraient plus à aucun fichier sur le serveur, mais, outre que la modification du chemin peut avoir été accidentelle, on risquerait de perdre les informations qui auraient été apportées à ces enregistrements avant le second import (description ajoutée en commentaire). D'où le choix d'ajouter les enregistrements correspondant au nouveau chemin en les « marquant » (champ SignalDouble coché et avertissement explicite ajouté au champ Commentaire) et de marquer de la même façon leurs équivalents préexistants. La suppression éventuelle de l'un des deux est laissée au jugement de l'opérateur.

Commentaire : champ texte (mémo), permettant la description sommaire du fichier (qui servira pour la synthèse descriptive qui sera faite au niveau Dossier).

DatImport : date de l'import (format abrégé)

⁴³ Le choix de cette nomenclature non signifiante et indépendante du nom d'origine du fichier se justifie à deux titres : d'une part, l'univocité du nom du fichier doit être garantie (ces documents sont en effet susceptibles d'être décrits dans le fichier « Documents » où l'univocité du nom est obligatoire) ; d'autre part, l'accès aux fichiers convertis se faisant directement depuis la base (table des métadonnées), le nom d'origine du document lui-même, de même que sa description, sont accessibles aux consultants.

EmpreinteNum : texte (255 car.), empreinte numérique, à ce jour inutilisé car il ne s'est pas avéré possible de la générer automatiquement (développement à confier à un prestataire extérieur ?)

Outre les champs de la table, le sous-formulaire Métadonnées dispose des fonctionnalités suivantes :

- un champ de liste, en haut du sous-formulaire, qui permet d'atteindre un enregistrement (liste associée),
- un bouton qui permet de filtrer les enregistrements dont les noms de fichiers (format .iii) contiennent « @ » ou « \$ » (ces caractères identifient les fichiers de présentation Arc + dans le fonds Fainsilber⁴⁴),
- le bouton (icône crayon) à droite du champ « NouvIdFichier » (cf. plus haut), qui sert à le remplir automatiquement en créant le nom (non signifiant) du fichier converti,
- un bouton (icône Dossier), à droite du champ « Chemin », qui ouvre le sous-répertoire contenant le fichier désigné en cours (et pointe directement ce dernier),
- par ailleurs, un double clic sur le champ refDosLie (étiquette : « Dossier source ») permet d'ouvrir la fiche Dossier désignée et un bouton à droite de ce champ permet d'ouvrir le répertoire correspondant ; de même pour le champ refDosDest (étiquette : « Dossier de destination »).

5.3.3 Sous-formulaire Doublons

Placé sous le sous-formulaire des métadonnées, il affiche les doublons de l'enregistrement actif dans le sous-formulaire des métadonnées. Il est basé sur une requête utilisant le champ « Taille » : cela n'assure pas que les fichiers signalés soient de « vrais » doublons (dans l'idéal, il devrait utiliser l'empreinte numérique : cf. § 4.5.1).

Les champs et fonctionnalités de ce sous-formulaire sont identiques à ceux du sous-formulaire des métadonnées, excepté les fonctionnalités de gestion (création de nouvelles phases et dossiers) : les données sont verrouillées par sécurité, à l'exception du champ Commentaire. On peut cependant supprimer un enregistrement depuis ce sous-formulaire.

En cas de doublons avérés, appartenant à des phases différentes, il est probable que, si l'on peut éviter de convertir deux fois le fichier, on fera une copie du fichier dans les deux sous-répertoires Phases correspondants pour ne pas créer de dossiers organiquement incomplets⁴⁵.

5.3.4 Attribution du dossier de destination

Le fait de saisir une nouvelle cote de dossier dans le champ « cote de destination » (événement « not in list ») provoque un message proposant d'ouvrir le formulaire des phases, d'où l'on peut, soit créer une nouvelle phase (puis un nouveau dossier), soit créer un nouveau dossier dans une phase pré-existante. On peut également accéder directement au formulaire Phases via le bouton « Form. Phase » en haut à droite du formulaire.

⁴⁴ Il y aura lieu de créer ultérieurement un champ permettant de filtrer les enregistrements en fonction des caractéristiques de dénomination des fichiers, dans la mesure où chaque fonds a ses propres règles de nommage.

⁴⁵ En termes de méthode, il est suggéré de traiter le premier fichier (conversion, information des champs...) et de ne se préoccuper de la copie que lorsque l'on tombera sur le doublon (le fait qu'existe déjà un fichier converti apparaîtra automatiquement dans le sous-formulaire Doublons, et il suffira de faire un copier-coller).

Une fois dans le dossier, un nouveau bouton (avec une croix) permet de revenir au formulaire des métadonnées : la nouvelle cote dossier est alors automatiquement implémentée dans le champ « cote de destination ».

5.3.5 Finalisation du traitement d'un objet

En fin de traitement de tous les fichiers d'un objet donné, le bouton « Déplacer fichiers convertis » permet :

- de créer automatiquement tous les sous-répertoires phases et dossiers correspondant aux enregistrements créés dans la base lors du parcours des fichiers,
- d'y déplacer automatiquement tous les fichiers convertis (enregistrés au même endroit que les fichiers originaux) vers leurs répertoires de destination.

5.3.6 Récapitulatif des opérations de traitement des fichiers numériques pour un objet donné

L'ensemble des opérations à mener pour chaque objet, détaillées ci-dessus, peut se résumer ainsi :

1. Pour l'ensemble des *fichiers sources* de l'objet considéré, créer automatiquement une phase (numérotée 0) dans la base de données, et, à l'intérieur du répertoire Objet correspondant sur le serveur, le répertoire source (numéroté 10000) ;
2. Copier manuellement l'ensemble des fichiers sources dans ce répertoire puis extraire (par import automatique) les métadonnées qui leur sont liées ;
3. Parmi celles-ci, sélectionner tous les *fichiers sources de présentation* (dont les noms d'origine comportent les signes @ ou \$) ;
4. Ouvrir et décrire chaque fichier source de présentation (champ Description de l'enregistrement métadonnée) ;
5. Puis copier (automatiquement) l'enregistrement décrivant le fichier source pour créer l'enregistrement qui décrira le *fichier converti* ; compléter, pour ce dernier enregistrement, les champs propres aux fichiers convertis (format de conversion, nom du fichier converti, répertoire de destination) ;
6. Créer, dans la base, au fur et à mesure de l'exploration des fichiers, les phases et les dossiers nécessaires pour y organiser logiquement les fichiers ;
7. Convertir, toujours au fur et à mesure de leur exploration, les fichiers sources en DXF et en PDF, en les laissant d'abord dans le répertoire des fichiers sources (s'il existe un doublon déjà converti, on peut se contenter de le recopier) ;
8. Lancer le script créant automatiquement, sur le serveur, les répertoires Phases et Dossiers définis dans la base de données et y transférant les fichiers convertis (depuis le répertoire source) ;
9. Phase par phase, rédiger une description synthétique dans le champ Description de la fiche dossier ;
10. Tout à la fin du traitement d'un objet, rédiger la description du répertoire 10000 (fichiers sources).

6. CONCLUSIONS

La présente étude devait déboucher sur l'évaluation du temps et du coût de traitement du fonds Fainsilber. Cela ne s'est pas avéré possible à partir des étapes parcourues, très exploratoires et qui ont requis la mise en place de méthodologies et de modules de description. Néanmoins, ce chapitre se conclut sur une estimation hypothétique à partir des tests et des expérimentations.

Pour évaluer la charge de traitement d'un fonds (cf. § 6.4.2), il est important de commencer par une étude détaillée de celui-ci afin d'identifier son organisation et les types (formats) de fichiers qu'il contient.

Le présent chapitre réunit quelques conclusions générales sur le traitement d'un fonds d'archives numérique, des questions, et des indications sur la production des archives au sein d'une agence aujourd'hui.

6.1 OUTILLAGE SPÉCIFIQUE POUR LE TRAITEMENT D'UN FONDS D'ARCHIVES NUMÉRIQUE

6.1.1 Outil d'identification des formats

La base PRONOM⁴⁶ fournit le descriptif d'un grand nombre de logiciels et de formats. Si le format recherché n'y figure pas (ce qui est le cas des formats issus du logiciel Arc+), essayer de repérer dans l'arborescence du fonds une information décrivant le ou les logiciels utilisés et se documenter sur les formats liés à ces logiciels.

6.1.2 Outils de conversion pour atteindre des formats de pérennisation

Pour chaque format à pérenniser, il convient de se poser quelques questions :

- Ce format est-il lisible avec d'autres outils que le logiciel qui l'a créé ?
- Ce format est-il ouvert, documenté ? A-t-il fait l'objet d'une normalisation ?

Si la réponse à l'ensemble de ces questions est négative, il est nécessaire de trouver le moyen de convertir ces fichiers, même imparfaitement, afin de préserver leur contenu dans le temps.

En fonction de la catégorie des fichiers, des outils de conversion existent. Dans cette étude, nous avons exploré les outils suivants :

- PDF Creator 1.7.3 : pilote d'impression servant comme outil de conversion vers le format PDF/A 2b suivant le profil supporté par l'outil à partir d'Arc+ 2012.
- ViewCompanion : logiciel de lecture des fichiers PLT et de conversion de ces fichiers vers PDF/A 2b suivant le profil supporté par l'outil. Il permet de même le traitement par lot.
- L'export intégré à Arc+ 2012 du format DDD/III vers le format DWG
- L'export intégré à AutoCAD 2012 du format DWG vers PDF/1.6
- L'export intégré à DraftSight 1.5 du format DWG vers le format DXF 2013

⁴⁶ <http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx>

6.1.3 Outils de vérification des formats de pérennisation

Les outils de vérification permettent de valider la conformité d'un format à sa spécification.

Les outils Facile / Jhove ont été testés pour valider les fichiers PDF produits.

À ce stade de l'étude, nous n'avons pas identifié d'outils de validation des fichiers produits au format DXF 2013, en dehors de leur ouverture sous AutoCAD ou DraftSight sans signalement d'erreurs.

6.2 COMPÉTENCES ET CONNAISSANCES NÉCESSAIRES

La personne chargée d'un traitement d'archives numériques d'architectes doit être un archiviste formé aux outils de dessin assisté par ordinateur fréquemment utilisés (AutoCAD, ArchiCAD, Arc+, Rhinoceros, 3DStudio, Allplan, SketchUp, Revit etc.). Autrement dit, il est difficile de faire l'économie d'une véritable formation d'architecte : il sera préférable de rechercher un profil d'architecte ayant acquis des compétences en traitement des archives.

Quand le fonds est bien identifié et les documents à traiter bien énumérés et sans difficulté d'interprétation, un traitement en masse est possible. Soit les logiciels disponibles permettent de faire ce traitement en interne, soit il faudra le déléguer à une société spécialisée. L'évaluation budgétaire de la société spécialisée dépendra de la qualité de l'étude préliminaire.

6.3 ARCHIVAGE PRÉLIMINAIRE DANS L'AGENCE

Dans l'organisation type d'une agence d'architecture, un ou plusieurs architectes sont à l'origine des flux d'informations générateurs de documents numériques susceptibles d'être présents dans le fonds d'archives numériques de l'agence. Ces architectes ont parfois des rôles bien définis et représentent les interlocuteurs privilégiés de la maîtrise d'ouvrage. Ils ont en général besoin de supports de présentation électroniques ou sur papier afin de valider leur idées et de communiquer sur leur projet. Avec les entreprises, les documents électroniques transmis ont valeur soit de documents contractuels finis, soit de documents de travail sur lesquels chaque entreprise apporte sa contribution. En fonction de la destination et de la qualification de cette valeur (parfois subjective), les formats d'échanges diffèrent⁴⁷.

Pour produire ces documents, l'architecte a le choix de travailler seul ou avec un collectif de dessinateurs et projeteurs. En fonction de la taille de la structure, elle pourra comprendre en outre des ingénieurs informaticiens (*CAD managers*) garants de la cohérence documentaire de l'agence⁴⁸. Ceux-ci assurent :

- une politique d'identification technique des composants sources de chaque document produit et diffusé par l'agence,

⁴⁷ Aujourd'hui, on constate que les formats PDF ou PLT deviennent les formats des documents contractuels et le format DWG le format de travail et d'échange avec les entreprises. À ce jour, les maquettes numériques (de type BIM – Building information model) ne dominent pas encore la production des documents. Vu l'évolution de l'adoption de ces maquettes par d'autres pays (Angleterre, États-Unis), une réflexion sur les modes d'archivage de ces maquettes numériques sera nécessaire dans un avenir très proche.

⁴⁸ Et, de plus en plus, des « BIM managers » (voir note précédente), garants de la cohérence inter-entreprises des échanges d'informations et de collaboration sur le projet.

- un classement et un archivage plus ou moins productif (réutilisable ou permettant la communication sur les références de l'agence),
- et la rédaction d'une charte graphique basée sur les bonnes pratiques de classification interne et de codification des données techniques stockées dans les documents électroniques sources.

Le contexte informatique général (évolution rapide du matériel et des logiciels) impose une traçabilité permanente de la chaîne permettant la production de ces documents : Poste de travail → Système d'exploitation → Logiciels utilisés → documents sources et exports associés. Si cette chaîne est interrompue (et c'est presque toujours le cas pour les archives numériques antérieures au ^{xxi}^e siècle), il sera parfois impossible de reconstituer après coup un environnement permettant de restaurer à l'identique les fichiers sources archivés.

La majorité des agences ne se soucient pas de la réutilisabilité de leurs données. Elles stockent ces données sur des CD ou des DVD qu'elles rangent dans un placard. Les plus avancées ont une infrastructure réseau et archivent ces données sur des serveurs de fichiers redondants. Ces derniers augmentent leur productivité et leur permettent d'accéder ainsi à une mémoire d'entreprise enrichie par chaque projet.

Malheureusement, même dans ce dernier cas, une perte de données par écrasement, effacement ou défaillance matérielle n'est jamais exclue. Les actions de restauration à partir de copies de sauvegarde sont en général possibles mais peu pratiquées. La cause principale est le rythme de production, qui ne laisse pas le temps nécessaire pour l'organisation de l'archivage préliminaire au sein des agences, sa classification, et, en conséquence, la restitution rapide en cas de besoin.

Lors d'un dépôt, malgré toute la rigueur de l'architecte, certaines informations sur tous ces éléments sont omises ; l'archiviste applique ses compétences à faire des choix sur ce qui peut être archivé et dans quel format.

En résumé, un archivage préliminaire au sein de l'agence permettrait :

- L'émulation de l'environnement de travail de l'architecte (poste → système d'exploitation → version des logiciels utilisés) ou la collecte d'informations précises sur cette chaîne ;
- Une classification chronologique des projets et un découpage de chaque projet par phase de projet ;
- L'identification des fichiers sources à conserver et l'élimination de tous les fichiers pouvant apporter une confusion dans l'interprétation par autrui.

6.4 BILAN DU TRAITEMENT DES ARCHIVES FAINSILBER

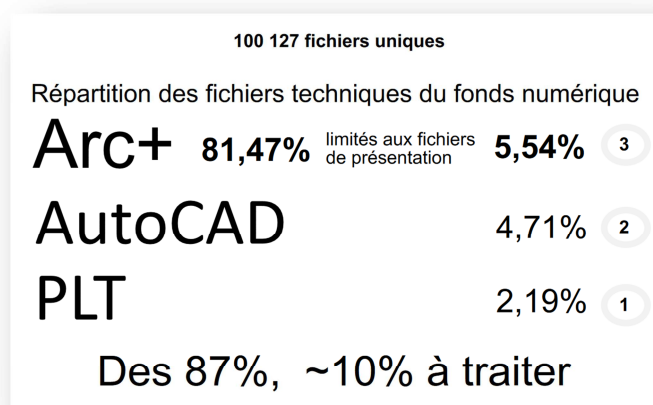
6.4.1 Comparaison entre le traitement des archives numériques et des archives sur papier

Le traitement des fichiers numériques est de diverses façons marqué par les habitudes de traitement des archives papier. L'usage de la base de données ArchiVecture contribue sans doute dans une certaine mesure à cette analogie, qui paraît de toute façon nécessaire et logique compte tenu des analogies existantes dans la conception même des documents (malgré les importantes différences).

Il a cependant été nécessaire d'adapter la base ArchiVecture pour qu'elle puisse gérer la conservation parallèle de plusieurs formats de pérennisation (deux en l'occurrence).

La conservation des fichiers sources, qui a été retenue pour le fonds Fainsilber et pour le moment, peut être remise en cause dans l'avenir : dans le cas de ce fonds en particulier, il est probable que ces fichiers ne pourront pas être ouverts indéfiniment (maintenance d'Arc+ et non-compatibilité de ses futures versions).

6.4.2 Priorités de traitement pour les fichiers techniques du fonds Fainsilber

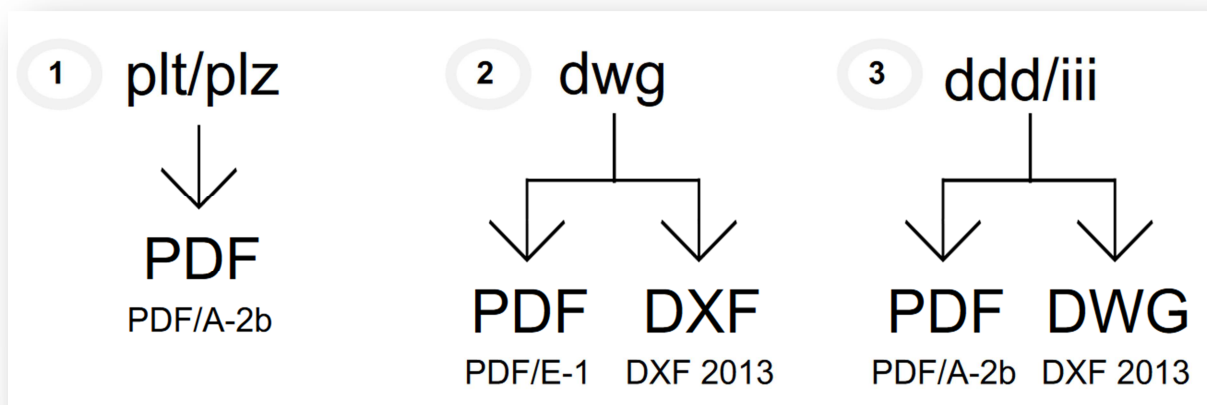


Une fois identifiés les fichiers uniques du fonds (60% du fonds), l'objectif est de repérer les fichiers techniques (87% des fichiers uniques), puis, à l'intérieur de ceux-ci :

- d'identifier les fichiers d'impression (PLT/PLZ) (2,19% des fichiers uniques), à traiter en priorité (1) ;
- puis, en visitant les fichiers natifs ou d'échange en DWG ou DXF, d'identifier et de traiter (2) ceux qui ne sont pas déjà traités dans l'étape précédente (moins de 4,71% des fichiers uniques) ;
- enfin, d'identifier les fichiers Arc+ uniques de présentation et de traiter (3) ceux qui ne sont pas traités dans les deux étapes précédentes (moins de 5,54% des fichiers uniques).

On réduit ainsi la marge d'erreur d'interprétation et on limite le temps de traitement au strict nécessaire (moins de 10% des fichiers uniques) afin de disposer d'une vue d'ensemble de l'ensemble du fonds.

6.4.3 Formats de destination pour les fichiers techniques du fonds Fainsilber



Les fichiers PLT sont identiques aux productions papier de l'agence. Leur conversion en PDF sera de facto la représentation la plus fiable du fonds.

Les fichiers DWG accompagnés des fichiers CTB de configuration des plumes sont identiques aux fichiers de travail de l'agence fournis aux entreprises. Réalisés par l'agence à partir de fichiers Arc+, ils permettent d'identifier visuellement la méthode de conversion d'Arc+ vers le format DWG d'AutoCAD. Leur conversion en PDF et DXF permettra la pérennisation de ces fichiers à la fois pour une illustration visuelle et une consultation technique détaillée. L'export PDF (1.6) d'AutoCAD se rapproche du standard PDF/E-1.

Les fichiers ddd/iii, fichiers natifs d'Arc+, ne seront pris en compte que si aucun fichier PLT ou DWG ne couvre leur contenu visuel et technique. L'impression virtuelle en PDF avec un pilote d'impression adapté permet la production d'un PDF standardisé avec une hypothèse d'interprétation des fichiers PAR. L'export DWG d'Arc+ permettra l'adaptation manuelle, qui à son tour permettra l'export DXF à partir d'outils comme DraftSight.

6.4.4 Quelques éléments d'évaluation (temps et coût) du traitement du fonds Fainsilber

L'évaluation repose sur les questions suivantes :

- Pour quel usage considère-t-on que le fonds est archivé ? (Présentation/impression, réutilisation...)
- Quel est le temps initial de mise en place du traitement pour un fonds (détection des doublons, classement des objets, description de chaque objet, identification des documents à traiter) ?
- Les formats des données techniques sont-ils exploitables avec des outils abordables et disponibles lors du traitement ?
- Quel est le temps unitaire de traitement de chaque document en fonction de son format et des outils disponibles ?

Pour le fonds Fainsilber, on peut faire les remarques suivantes :

- Le choix de conserver deux formats parallèlement (de présentation/impression, PDF, et de réutilisation, DXF) génère une charge de travail supplémentaire qui a multiplié par 6 le temps de traitement dans les tests effectués (passant de l'ordre de 1 à 6 minutes par document) pour les fichiers natifs d'Arc+.
- Le temps initial doit prendre en compte la qualité du fonds d'archives fourni par l'architecte. Dans le cas considéré, le manque d'homogénéité entre les projets induit un travail initial important pour l'ensemble du fonds et un travail spécifique pour chaque projet avant de travailler sur chaque document :
 - Au niveau fonds, l'harmonisation de l'arborescence des fichiers de bibliothèque (ARCALIB) ;
 - Au niveau objet (projet), le travail sur les tables de couleurs, épaisseurs de traits, types de ligne et format de sortie.

Les outils permettant la conversion testés ici peuvent être qualifiés ainsi :

- Pour la conversion PLT → PDF, ViewCompanion est **efficace**. Un traitement par lots peut être configuré, de même que pour l'export DWG → PDF 1.6 (AutoCAD) ;
- L'impression virtuelle en PDF des formats Arc+ **nécessite une action manuelle** et une part d'interprétation (utilisation d'un pilote d'impression PDFCreator). Le traitement nécessite le choix d'une table de couleurs/épaisseurs de trait/types de lignes dans une liste préétablie afin de s'approcher au mieux des choix de l'architecte.
- L'export du format Arc+ vers DWG/DXF (Arc+ version 2012) est **défectueux**, générant des traitements supplémentaires manuels par document à conserver (le logiciel DraftSight a été utilisé pour le nettoyage, la correction de certaines entités graphiques ou textuelles et pour permettre l'export valide en DXF 2013).

On aboutit finalement aux temps de traitement indicatifs suivants :

- Sur la base des documents Arc+ traités, il faut prévoir un temps unitaire de 6 à 10 minutes par document unique à convertir (si les fichiers PAR correspondants sont bien identifiés) pour aboutir à l'ensemble des fichiers de pérennisation, à la description des documents dans la table des métadonnées correspondante et à l'analyse des doublons liés.
- Sur la base des documents PLT traités, il faut prévoir moins d'une minute par document ou lot de documents (dans un même répertoire) pour la conversion avec ViewCompanion en PDF/A, puis 1 à 2 minutes de description par document.
- Sur la base des documents AutoCAD à convertir, il faut prévoir moins d'une minute pour la conversion en PDF/1.6 (si les fichiers CTB correspondants sont bien identifiés) et l'export en DXF, puis 1 à 2 minutes de description par document.

Suite aux différents éléments de bilan établis ci-dessus, on peut proposer l'hypothèse suivante sur le temps nécessaire pour le traitement des fichiers techniques du fonds :

- Conversion des fichiers PLT/PLZ en PDF/A-2b :

- PLT : $2178 \times (1+2) = 6534$ minutes = 108,9 heures = 15,6 jours⁴⁹
- PLZ (qui sont des fichiers contenant chacun plusieurs fichiers PLT) : $21 \times (1+(\sim 10 \times 2)) = 441$ minutes = 7,35 heures = ~ 1 jour
- Conversion des DWG / DXF en PDF (v1.6) et DXF 2013 :
 - Identification CTB : $85 \times (\sim 5) = 425$ minutes = 7,1 heures = ~ 1 jour
 - DWG : $3388 \times (1+2) = 10164$ minutes = 169,4 heures = 24,2 jours
 - DXF : $1244 \times (1+2) = 3732$ minutes = 62,2 heures = 8,9 jours
- Conversion des fichiers III/DDD en PDF/A-2b :
 - Identification PAR : $212 \times (\sim 10) = 2120$ minutes = 35,3 heures = ~ 5 jours
 - DDD/III : 33876 dont 4021 contenant @ et 1525 contenant \$ à traiter en priorité : $5546 \times (10) = 55460$ minutes = 924,3 heures = 132 jours
- Total du temps nécessaire pour le traitement des fichiers techniques du fonds et la saisie de leurs descriptions : $15,6 + 1 + 1 + 24,2 + 8,9 + 5 + 132 = 187,7$ jours.

Au total donc 187 jours, soit 9 mois et demi, uniquement pour le temps de traitement matériel du sous-ensemble très restreint de fichiers considérés comme prioritaires (et en gros suffisants) d'un fonds d'archives... C'est un temps très comparable à celui que prendrait le traitement des archives papier du même fonds, et qui, dans les classements d'archives à venir, s'y ajoutera.

Un budget prévisionnel devrait prévoir un coût jour entre 350 et 500 euros/jour pour une entreprise spécialisée, d'où un budget de $187,7 \times (350 \text{ à } 500) = 65\,695 \text{ € à } 93\,850 \text{ €}$. Ce budget ne prend pas en compte le facteur de redondance entre les différents formats et les traitements par lots possibles. Ce facteur impliquerait une réduction de la charge de travail qui peut être estimée entre 30% et 50%.

6.4.5 Ce qui peut être généralisé

Les choix faits pour le fonds Fainsilber semblent partiellement généralisables, impliquant cependant un important travail de réflexion initiale avant tout traitement. Plus précisément, le choix des formats de présentation (PDF/A) et de réutilisation (DXF) est généralisable pour la majorité de la production des années 1980 et 1990. Dès les années 2000 en revanche, la 3D et la maquette numérique du bâtiment prennent une part plus ou moins importante dans la production des agences et nécessiteront une étude préliminaire approfondie sur les formats associés.

L'optimisation de l'outil de traitement n'a pu être complète, avec en particulier le problème non résolu de la génération d'empreintes numériques à partir d'ArchiVecture⁵⁰. Plus largement, conceptuellement, il serait encore prématuré de considérer que la méthodologie concernant les doublons est généralisable, même si les grandes lignes en sont bien tracées dans cette étude.

⁴⁹ 1 jour = 7 heures de travail sur la base de 35 heures hebdomadaires.

⁵⁰ Toutes les analyses sur les doublons sont faites à partir de la constitution d'une base SQLite contenant l'empreinte SHA2 de chaque fichier dans le fonds ainsi que des fichiers contenus dans les zip de premier niveau.

6.4.6 Précautions ou recommandations

Précautions techniques :

- La durée de vie d'un fichier au format propriétaire dépend de l'existence du logiciel associé et du support du format (rarement correctement exploitable au-delà de 10 ans).
- Le producteur et son environnement de travail (matériel, système d'exploitation, logiciels, configurations) sont rarement bien identifiés.
- Les conversions de sécurité (en préservant les originaux) doivent être réalisées vers des formats normalisés.

Précautions métier pour les documents graphiques :

- Identifier la valeur d'un document (ce qui implique de faire des choix),
- Assurer la restitution graphique du contenu.

Remarques d'ordre archivistique :

- Le document est virtuel, est constitué d'une collection de fichiers sources et une collection de fichiers convertis.
- Le document peut être dépendant d'une technologie, d'un logiciel, d'une plateforme ou d'un système d'exploitation.
- Le document peut être dupliqué facilement, archivé à plusieurs reprises, partagé dans différents projets.
- La quantité de documents numériques est exponentielle par rapport aux archives physiques traditionnelles.

7. WEBOGRAPHIE

7.1 CITÉ DE L'ARCHITECTURE ET DU PATRIMOINE

- Archiwebture : <http://archiwebture.citechaillot.fr>
- Fonds Fainsilber : http://archiwebture.citechaillot.fr/fonds/FRAPN02_FAINAD
- Coordination de l'action « Archives d'architecture » du programme européen Gaudi, 2002-2008 ; colloque « Architecture et archives numériques », Paris, Cité de l'architecture et du patrimoine et INHA, décembre 2007.
- Architecture et archives numériques. L'architecture à l'ère numérique : un enjeu de mémoire / Architecture and digital archives. Architecture in the digital age: a question of memory, dir. David Peyceré et Florence Wierre, Gollion : InFolio, coll. Archigraphy, 2008
- (actes du colloque mentionné ci-dessus).

7.2 SIAF

Guides méthodologiques SIAF :

- Le format de fichier PDF :
<http://www.archivesdefrance.culture.gouv.fr/static/6189>
- Choix de formats numériques pérennes dans un contexte de données orales et visuelles :
<http://www.archivesdefrance.culture.gouv.fr/static/4923>
- Les outils de conversion vers le format PDF :
<http://www.archivesdefrance.culture.gouv.fr/static/7248>

7.3 CINES

Critères d'éligibilité d'un format à l'archivage :

<https://www.cines.fr/archivage/des-expertises/expertise-formats/criteres-deligibilite-dun-format-a-larchivage/>

Facile, Outil de Validation du Format d'Archivage du CINES par analyse et Expertise :
<http://facile.cines.fr/>

Jhove, projet libre sur lequel est basé l'outil en ligne « Facile » : <http://jhove.sourceforge.net/>

7.4 FORMATS, NORMES ET STANDARDS

Choix de formats numériques pérennes dans un contexte de données graphiques vectorielles 2D :

- SVG : Recommandations W3C : <http://www.w3.org/TR/SVG11/>

- DXF : Spécifications :

<http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=12272454&linkID=10809853>

- DWG : Spécifications de l'Open Design Alliance :

http://opendesign.com/files/guestdownloads/OpenDesign_Specification_for_.dwg_files.pdf

- PDF/A-2 :

ISO 19005-2:2011 60.60 : Gestion de documents -- Format de fichier des documents électroniques pour une conservation à long terme -- Partie 2: Utilisation de l'ISO 32000-1 (PDF/A-2) :

http://www.iso.org/iso/fr/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50655

- PDF/E-1 : ISO 24517-1

http://www.iso.org/iso/catalogue_detail.htm?csnumber=42274

- PDF/E-2 : ISO/DIS 24517-2:2013-08

<http://www.beuth.de/en/draft-standard/iso-dis-24517-2/193015617>

8. BIBLIOGRAPHIE

8.1 CITÉ DE L'ARCHITECTURE ET DU PATRIMOINE

- Coordination de l'action « Archives d'architecture » du programme européen **Gau:di**, 2002-2008 ; colloque « Architecture et archives numériques », Paris, Cité de l'architecture et du patrimoine et INHA, décembre 2007.
- *Architecture et archives numériques. L'architecture à l'ère numérique : un enjeu de mémoire / Architecture and digital archives. Architecture in the digital age: a question of memory*, dir. David Peyceré et Florence Wierre, Gollion : InFolio, coll. Archigraphy, 2008 (actes du colloque mentionné ci-dessus).

8.2 AUTRES

- *Les écrits s'envolent. La problématique de la conservation des archives papier et numériques* / Charles Kecskeméti et Lajos Körmendy, Lausanne : Éditions Favre, 2014. 208 p. ISBN 978-2-8289-1425-7

9. ANNEXES

9.1 ÉTUDE DE FAISABILITÉ DU TRAITEMENT DU FONDS, JUILLET 2011

Cité au § 1.1

Rapport d'étude de faisabilité du traitement du fonds Adrien Fainsilber, Cassandra Éveillard-Kervarrec, Sonia Gaubert, David Peyceré, Centre d'archives d'architecture du xx^e siècle, Cité de l'architecture et du patrimoine, juillet 2011.

Document en annexe externe intitulé : « 9.1-Etude-faisabilite-juillet-2011.pdf »

9.2 CHARTE GRAPHIQUE TRANSMISE PAR L'AGENCE

Cité au § 2.2

Charte graphique de l'agence Adrien Fainsilber et associés datée d'août 2006.

Voir documents en annexe externe intitulés :

« 9.2a-charte-graphique-mode-emploi-agenceAFA.doc »,

« 9.2b-charte-graphique-liste-calques-agenceAFA.xls »

9.3 UTILISATION D'ARC+ À L'AGENCE

Cité au § 2.2

Document explicatif de l'organisation des fichiers des archives numériques de l'agence Adrien Fainsilber.

Voir document en annexe externe intitulé : « 9.3-utilisation-arc+-agenceAFA.pdf »

9.4 LISTE COMPLÈTE DES TYPES DE FICHIERS

Cité au § 3.1

Le classeur en annexe externe intitulé « 9.4-liste-complete-types-fichiers-fondsAFA-PASS-TECH.ods » est constitué de 3 feuilles :

- Tout type unique : liste de l'ensemble des types de fichiers, le nombre de documents uniques par type, le pourcentage de document par type par rapport au nombre total de documents uniques et enfin une proposition pour le sort de chaque type identifié.
- Arobase unique : Sous-ensemble de la première feuille filtrant sur les documents contenant un « @ » dans le nom du fichier.
- Dollar unique : Sous-ensemble de la première feuille filtrant sur les documents contenant un « \$ » dans le nom du fichier.

9.5 TABLEAU DE GRILLES D'ANALYSE

Cité au § 3.3.3

Voir document en annexe externe intitulé : « 9.5-tableau-grilles-analyse-formats-techniques.ods »

9.6 ANALYSE DES 5 CD ZIP DE COLMAR

Cité au § 4.4.3

Voir document en annexe externe intitulé : « 9.6-analyse-5-CD-Colmar.pdf »

9.7 FAINAD-A-2 : ANALYSE

Cité au § 4.5.1

Voir document en annexe externe intitulé : « 9.7-FAINAD-A-2-Analyse.pdf »

9.8 BIBLIOTHÈQUE ARC+

Cité au § 4.5.1

Voir document en annexe externe intitulé : « 9.8-Bibliotheque-ARCPLUS.pdf »

9.9 PROCÉDURE DE CONVERSION DES FICHIERS ARC+ ET PLT

Cité au § 4.8

Voir document en annexe externe intitulé : « 9.9-procedure-de-conversion-arcplus.pdf »

9.10 GESTION ET DESCRIPTION DES ARCHIVES NUMÉRIQUES : RAPPELS

9.11 MANUEL ARCHIVECTURE : GUIDE DE SAISIE

9.12 MATÉRIEL, SYSTÈMES ET LOGICIELS DE L'AGENCE FAINSILBER

Liste (non datée) des postes dédiés à la DAO et l'infographie à l'agence lors du dépôt du fonds (d'après un fichier du fonds d'archives qui recense le matériel informatique, les systèmes d'exploitation utilisés et certains logiciels de traitement). N'ont été retenus ici que les logiciels de DAO et d'infographie et les postes qui en disposaient.

Machine	Système	HD	RAM	Logiciels
JK PII 450	Win 98	10	128	Arc+
YC K6-233	Win 95	6.4	64	Arc+
GH PII 450	Win 98	10	128	Arc+ Photoshop 5

Machine	Système	HD	RAM	Logiciels
PB PII 450	Win 98	10	256	Arc+ Photoshop 5 Art Lantis
K6-233	Win 95	6.4	64	Arc+
PVB PII 200	Win 95	4	96	Arc+
VD(1) PII 233	Win 98 Win 3.11 Dos			Arc+ 7
VD(2) K6-233	Win 95	6.4	64	Arc+
JRG PII 450	Win 98	10	128	Arc+
JF P200S	Win 95	64	4	Photo Look Agfa Photoshop 5
Ychu K6-233	Win 95	6.4	96	Arc+
EP PII 266	Win 95	3.2	96	Arc+