

# PyCON 2011 - Rennes

Manipulation de données avec Python

—

18/09/2011

# Sommaire

- Contexte du projet
- Prototype
- Premier script
- Validation du prototype
- Manipulation de données
- Un peu de coloriage
- Bilan de l'application

# Contexte du projet

- Client : DGFIP
- Besoin : Suivre le déploiement de Thunderbird
  - Remplace Lotus Notes et Outlook
- Données :
  - Un tag est activé sur le profil LDAP de l'utilisateur lors de son déploiement
  - Deux filières (gestion publique et fiscale) qui ont fusionné ou vont fusionner pendant la durée de vie de l'application.

# Prototype

- Demandé : une application intranet temps réel
- Proposé :
  - Une application intranet de restitution de données
  - Un ensemble de scripts pour modifier les données
- Contraintes :
  - Matrice technologique, temps de réponse
- Solution :
  - Dotproject (PHP, opensource, proche du besoin)
  - Scripts lancés deux fois par jour (techno non choisie)

cockpit

dotProject.net

FREE SOFTWARE

StructuresProjetsTâchesCalendrierFichiersContactsUtilisateursAdministration

Bienvenue Admin Person

Aide | Mon Compte | Mes tâches | Aujourd'hui | Déconnexion

Structures

nouvelle structure

Toutes les structuresDélégation Inter-régionaleServices CentrauxServices NationauxNon définiCarte de France

classé par:

Nom de la structure

Projets actifs

Type de la structure

	AUTRES	11	Non défini
	BP	8	Services Centraux
	CE	10	Services Centraux
	CL	8	Services Centraux
	Collectivités Outre-Mer	5	Non défini
	DDG Centre Auvergne Antilles	26	Délégation Inter-régionale
	DDG Est	28	Délégation Inter-régionale
	DDG Ile de France	16	Délégation Inter-régionale
	DDG Nord	14	Délégation Inter-régionale
	DDG Ouest	24	Délégation Inter-régionale
	DDG Rhône Alpes Bourgogne	24	Délégation Inter-régionale
	DDG Sud Est	18	Délégation Inter-régionale
	DDG Sud Ouest	24	Délégation Inter-régionale
	DDG Sud Pyrénées	26	Délégation Inter-régionale
	DG - DIRECTION GÉNÉRALE	6	Services Centraux
	DIRECTION DE LA FISCALITÉ	6	Services Centraux
	DIRECTION DE LA GESTION PUBLIQUE	3	Services Centraux
	DIRECTION LÉGISLATION FISCALE	14	Services Centraux
	DIRECTION PILOTAGE RÉSEAU-MOYENS	4	Services Centraux
	FRANCE DOMAINE	3	Services Centraux
	GF	15	Services Centraux
	JF	9	Services Centraux
	RH	13	Services Centraux
	Services Nationaux	16	Services Nationaux
	SI	31	Services Centraux

- Liste (362)
- Statistiques (tableaux)
- Statistiques BALP (%)
- Statistiques BALP (nb)
- Statistiques BALF (%)
- Statistiques BALF (nb)
- Gantt

Classé par:

Structure	Nom du Projet	Début	Fin	reset	
				Réel	Avancement global
SI	CSI CLERMONT-FERRAND	15/01/2009	15/04/2009	15/04/2009	0.0%
Services Nationaux	SCN COPERNIC	01/07/2009	01/10/2009	01/10/2009	90.0%
Services Nationaux	COPAPPUI	01/07/2009	01/10/2009	01/10/2009	88.0%
Services Nationaux	COP3	01/07/2009	01/10/2009	01/10/2009	78.0%
Services Nationaux	COP4	01/07/2009	01/10/2009	01/10/2009	94.0%
Services Nationaux	COP5	01/07/2009	01/10/2009	01/10/2009	82.0%
Services Nationaux	COP6	01/07/2009	01/10/2009	01/10/2009	71.0%
SI	CSI REIMS	15/10/2009	15/01/2010	15/01/2010	1.0%
SI	CSI NEVERS	15/10/2009	15/01/2010	15/01/2010	1.0%
SI	CSI DE LYON	15/10/2009	15/01/2010	15/01/2010	18.0%
SI	CSI STRASBOURG	15/10/2009	15/01/2010	15/01/2010	22.0%
DG - DIRECTION GÉNÉRALE	DIRECTION GENERALE	01/10/2009	31/01/2010	31/01/2010	22.0%
DG - DIRECTION GÉNÉRALE	DÉPARTEMENT STRATÉGIE ET AUDIT	01/10/2009	31/01/2010	31/01/2010	75.0%
DG - DIRECTION GÉNÉRALE	DÉPARTEMENT DE LA COMMUNICATION	01/10/2009	31/01/2010	31/01/2010	88.0%
DG - DIRECTION GÉNÉRALE	CABINET	01/10/2009	31/01/2010	31/01/2010	56.0%
DG - DIRECTION GÉNÉRALE	DDG	01/10/2009	31/01/2010	31/01/2010	25.0%
DG - DIRECTION GÉNÉRALE	MISSION MAÎTRISE DES RISQUES	01/10/2009	31/01/2010	31/01/2010	100.0%
DIRECTION LÉGISLATION FISCALE	DIRECTION LÉGISLATION FISCALE	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF A	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF B	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF B1	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF B2	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF C	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF C1	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF C2	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF D	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF D1	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF D2	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF E	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF E1	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION LÉGISLATION FISCALE	DLF E2	01/10/2009	31/01/2010	31/01/2010	0.0%
FRANCE DOMAINE	FD	01/10/2009	31/01/2010	31/01/2010	0.0%
FRANCE DOMAINE	FD1	01/10/2009	31/01/2010	31/01/2010	0.0%
FRANCE DOMAINE	FD2	01/10/2009	31/01/2010	31/01/2010	4.0%
DIRECTION DE LA FISCALITÉ	DIRECTION DE LA FISCALITE	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION DE LA FISCALITÉ	MISSION SIMPLIFICATIONS	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION DE LA FISCALITÉ	CF	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION DE LA FISCALITÉ	CF1	01/10/2009	31/01/2010	31/01/2010	3.0%
DIRECTION DE LA FISCALITÉ	CF2	01/10/2009	31/01/2010	31/01/2010	0.0%
DIRECTION DE LA FISCALITÉ	CF3	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	AGR	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF1	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF1A	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF1B	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF2	01/10/2009	31/01/2010	31/01/2010	0.0%
JF	JF2A	01/10/2009	31/01/2010	31/01/2010	5.0%

SI	SI1A	01/10/2009	31/01/2010	31/01/2010	67.0%
SI	SI1B	01/10/2009	31/01/2010	31/01/2010	72.0%
SI	SI1C	01/10/2009	31/01/2010	31/01/2010	82.0%
SI	SI1D	01/10/2009	31/01/2010	31/01/2010	86.0%
SI	SI1E	01/10/2009	31/01/2010	31/01/2010	46.0%
SI	SI1F	01/10/2009	31/01/2010	31/01/2010	30.0%
SI	SI2	01/10/2009	31/01/2010	31/01/2010	50.0%
SI	SI2A	01/10/2009	31/01/2010	31/01/2010	78.0%
SI	SI2B	01/10/2009	31/01/2010	31/01/2010	76.0%
SI	SI2C	01/10/2009	31/01/2010	31/01/2010	2.0%
SI	DI3CE	01/10/2009	31/01/2010	31/01/2010	92.0%
DDG Rhône Alpes Bourgogne	Côte d'Or - FISC	01/10/2009	31/01/2010	31/01/2010	29.0%
DDG Rhône Alpes Bourgogne	Côte d'Or - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	55.0%
DDG Est	Marne - FISC	01/10/2009	31/01/2010	31/01/2010	18.0%
DDG Est	Marne - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	40.0%
DDG Est	Meuse - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Est	Meuse - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Est	Vosges - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Est	Vosges - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Ile de France	Essonne - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Ile de France	Essonne - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Ile de France	Hauts de Seine - FISC	01/10/2009	31/01/2010	31/01/2010	11.0%
DDG Ile de France	Hauts de Seine - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	11.0%
DDG Ile de France	Val de Marne - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Ile de France	Val de Marne - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	5.0%
DDG Nord	Seine Maritime - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Nord	Seine Maritime - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Nord	Nord - FISC	01/10/2009	31/01/2010	31/01/2010	4.0%
DDG Nord	Nord - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	2.0%
DDG Nord	Pas de Calais - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Nord	Pas de Calais - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Est	Alpes Maritime - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Est	Alpes Maritime - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	1.0%
DDG Ouest	Manche - FISC	01/10/2009	31/01/2010	31/01/2010	29.0%
DDG Ouest	Manche - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	19.0%
DDG Ouest	Orne - FISC	01/10/2009	31/01/2010	31/01/2010	1.0%
DDG Ouest	Orne - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	1.0%
DDG Sud Ouest	Deux Sèvres - FISC	01/10/2009	31/01/2010	31/01/2010	1.0%
DDG Sud Ouest	Deux Sèvres - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	3.0%
DDG Centre Auvergne Antilles	Indre - FISC	01/10/2009	31/01/2010	31/01/2010	22.0%
DDG Centre Auvergne Antilles	Indre - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	14.0%
DDG Centre Auvergne Antilles	Guyane - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Centre Auvergne Antilles	Guyane - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Pyrénées	Hérault - FISC	01/10/2009	31/01/2010	31/01/2010	4.0%
DDG Sud Pyrénées	Hérault - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	25.0%
DDG Sud Pyrénées	Pyrénées Orientales - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Pyrénées	Pyrénées Orientales - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Pyrénées	Haute Garonne - FISC	01/10/2009	31/01/2010	31/01/2010	0.0%
DDG Sud Pyrénées	Haute Garonne - GEST_PUB	01/10/2009	31/01/2010	31/01/2010	3.0%
SI	CSI LILLE	15/11/2009	15/02/2010	15/02/2010	79.0%
SI	CSI DE ROUEN	15/11/2009	15/02/2010	15/02/2010	41.0%
SI	CSI NEMOURS	15/11/2009	15/02/2010	15/02/2010	5.0%
SI	CSI VERSAILLES	15/11/2009	15/02/2010	15/02/2010	28.0%
SI	CSI AMIENS	15/11/2009	15/02/2010	15/02/2010	7.0%
DDG Rhône Alpes Bourgogne	Nièvre - FISC	01/11/2009	28/02/2010	28/02/2010	0.0%
DDG Rhône Alpes Bourgogne	Nièvre - GEST_PUB	01/11/2009	28/02/2010	28/02/2010	0.0%
DDG Rhône Alpes Bourgogne	Saône et Loire - FISC	01/11/2009	28/02/2010	28/02/2010	0.0%

# Premier script : Enjeux

- Nécessité d'entrer un jeu de données minimal
  - 1 bureau, 2 régions donc trois grandes structures
  - Une petite cinquantaine de projets
  - Autant de tâches
  - => Une centaine de formulaires au total à remplir
  - Script SQL potentiellement complexe, long à écrire
- Nécessité de recommencer avec les vraies données
- Description de la structure dans 2 CSV



# Premier script : Solution

- Utilisation de `csv.DictReader`
  - Un fichier csv devient une liste de dictionnaires dont les clés correspondent à la première ligne.
  - Une colonne numéro dans chaque fichier permettant de mettre en rapport grande structure et département
- Utilisation de SQLAlchemy
  - Trois classes pour les trois tables impactées
  - Création d'une connexion SQL et double boucle sur les grandes structures et départements

# Premier script : Bilan

- Positif :
  - Temps
    - Deux heures pour réaliser le script
    - Soit moins que le rajout manuel des 50 données
  - Ré-utilisabilité
    - Script déjà fonctionnel pour la future application
    - Composants CSV et SQL créé réutilisables indépendamment
- Négatif
  - Python n'est pas dans la matrice technologique

# Validation du prototype

- Évolutions au fil de l'expérimentation
  - Périmètre de l'application
    - Rajout de fonctionnalités
    - Ajustement des comptes et des ACL
  - Nature des données à présenter
    - Champs supprimés
    - Vocabulaire ajusté
- Application = prototype
- Python obtient une dérogation

# Manipulation de données

- Identification d'une personne
  - Sa filiale : un attribut LDAP ou des cas particuliers
  - Son organisation (département ou service central) :
    - Trois premier chiffres de son identifiant sauf Nantes et 93
    - Présence ou non d'attributs spécifiques
    - Croisement avec des données issues d'autres bases
    - Cas particuliers nominatifs gérés par des fichiers CVS
- Conséquences
  - De une requête LDAP par département à une par personne, réutilisation des composants

# Manipulation de données

- Gestion des BALF (BAL fonctionnelles)
  - Identification des groupes LDAP
    - Filiale
    - Organisation
      - Même problématiques
  - Conséquences
    - Fichiers CSV pour corriger l'affectation de BALF
- Détection de doublons
  - Incohérences entre les chiffres de diverses sources
  - Résolution en listant nominativement les problèmes

# Manipulation de données

- Génération de fichiers CSV
  - Besoins
    - Avoir, département par département, la liste des personnes et des statuts pour que les personnes qui suivent le déploiement puissent vérifier les chiffres
    - Faire un différentiel dans le temps, pour voir l'évolution des données
      - Création / suppression de comptes
      - Modification des affectations des comptes
      - Changements de statuts
      - Utilité dans d'autres contextes pour des personnes n'ayant pas LDAP
  - Solution : utilisation de dictWriter

# Manipulation de données

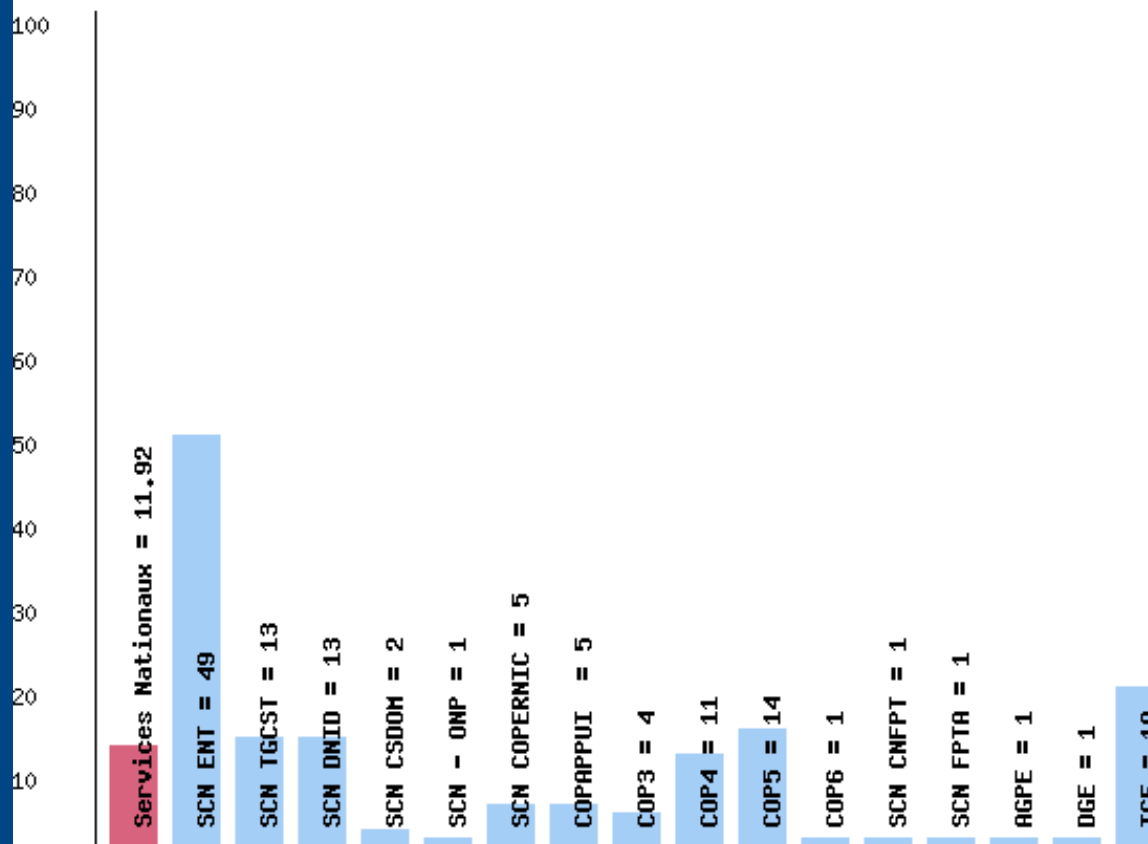
- Historisation des données
  - Besoin
    - Voir sur un graphique la progression du déploiement, département par département
    - Estimer l'avance ou le retard du déploiement
  - Solution
    - Historiser les données, une fois par jour (pickle, dict)
    - Construire un graphique à partir des dates de début et de fin du projet et des données quotidiennes
    - l'historisation et la génération de tous les graphiques après nouveau calcul des données

# Manipulation des données

Avancement du déploiement de : Services Nationaux

Axe des ordonnées : Avancement (%)

DDG ou bureau  
Direction ou service





# Manipulation de données

cockpit dotProject.net  
Structures Projets Tâches Calendrier Fichiers Contacts Utilisateurs Administration

Bienvenue Admin Person Aide | Mon Compte | Mes tâches | Aujourd'hui | Déconnexion

Voir le projet

liste des projets : éditer ce projet : Organiser les tâches

Rechercher: nouvelle tâche Nouvel événement Nouveau fichier effacer le projet

**Doubs - GEST\_PUB**

Détails

Structure: DDG Est  
Nom Court: 025  
Date de début: 01/11/2009  
Date de fin Cible: 28/02/2010  
Date de fin Réelle: 28/02/2010  
Avancement: 0.0%

Avancement

Cliquez sur le graphique pour l'agrandir  
Cliquez ici pour télécharger le graphique

Description

Tâches Diagramme de Gantt Rapports d'activité Evénements Fichiers

Voir: ☐ Seulement les tâches incomplètes

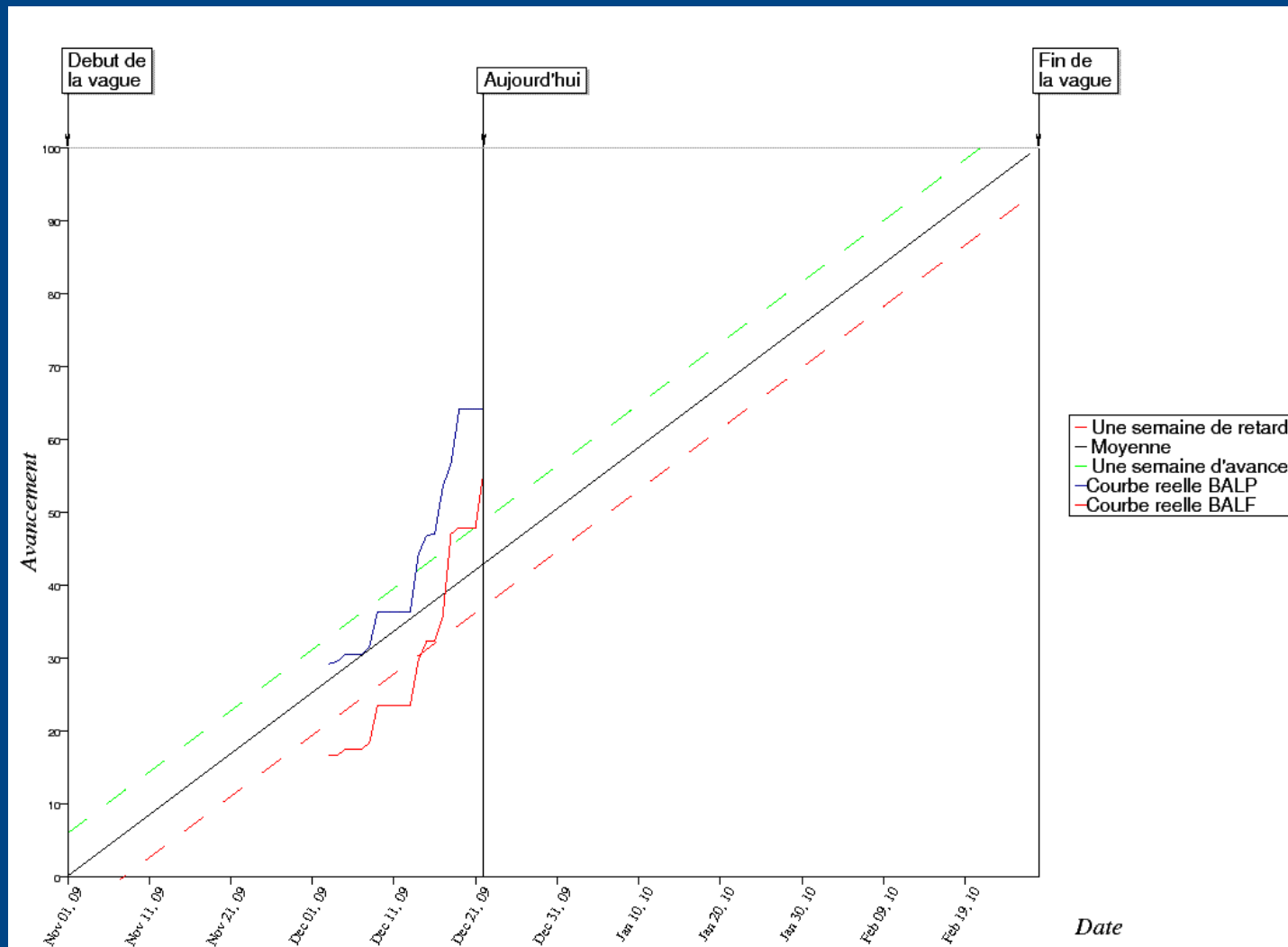
Marque Nouveau rapport Travail

Rapport 0% Doubs - GEST\_PUB

Cli: ☐ Tâche planifiée ☐ Commencé à temps ☐ Aurait du commencer ☐ Hors délais ☐ Fait Open : Close All Tasks

01/11/2009 09:00 67 jours 28/02/2010 17:00

# Manipulation des données



# Un peu de coloriage

- Besoin
  - Visuel pour la hiérarchie : carte de France
- Solution
  - Wikipédia : Téléchargement de fonds de carte
  - InkScape
    - France métropolitaine
    - Île de France
    - 5 départements d'outre-mer
    - Légende
  - Vectoriel = XML > ElementTree

# Un peu de coloriage

cockpit

dotProject.net  
FREE SOFTWARE

Structures Projets Tâches Calendrier Fichiers Contacts Utilisateurs Administration

-- Créer --

Bienvenue Admin Person

Aide | Mon Compte | Mes tâches | Aujourd'hui | Déconnexion



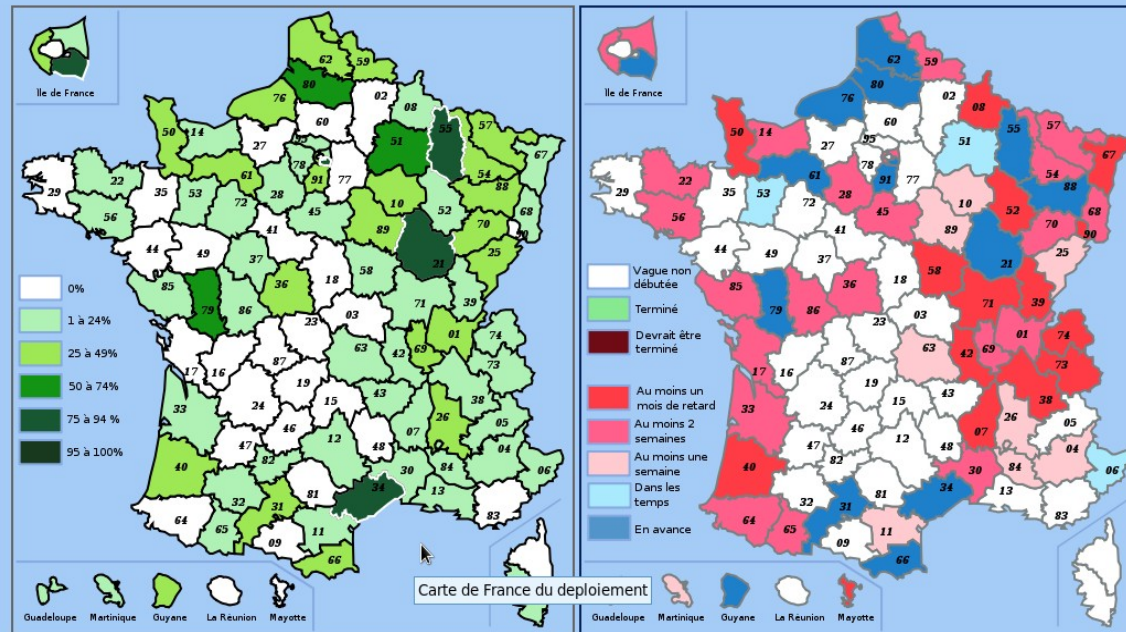
Structures

nouvelle structure

Toutes les structures Délégation Inter-régionale Services Centraux Services Nationaux Non défini Carte de France



Carte de France du déploiement



# Bilan de l'application

- Développement
  - Ultra-rapide
  - Coût très faible
  - Grande complexité dans la manipulation des données, rendue simple par l'utilisation de Python
  - Réalisation de visuels convaincants
    - Carte de France développée dans la journée, sans connaissances préalables
  - Une panne en deux années (due à la limitation du compte LDAP à 500 résultats)

# Bilan de l'application

- Sans Python, point de salut
  - Au début du projet, j'ai refusé de développer une application temps réel.
  - Lors de la création du prototype, j'ai développé en Python par soucis d'aller vite et pour montrer ce que le langage avait dans le ventre
  - Je n'ai pas eu à lutter pour imposer Python, parce que l'application était jetable et donc l'emprise de la matrice technologique limitée
  - Sinon, j'avais dans ma poche une estimation du même travail en PHP et j'aurais refusé de le faire en JAVA

# Bilan de l'application

- Linux magazine HS N°53 sur Python
  - Causes :
    - Article de JPT sur l'excès d'architecture en Java
    - Discussion avec JPT sur le modèle de données Python
  - Idée de base :
    - Travailler avec des données hétérogènes sans excès d'architecture
    - Généraliser les composants créés pour l'application et les partager
    - Expliquer la base pour comprendre les subtilités des composants ( y compris le modèle objet et les types)