

Améliorer la qualité des données

La fiabilisation des données de scolarité

Pierre-Emmanuel JEUNEHOMME

Université de Nancy 2

UNIVERSITE NANCY 2

- ❖ 4 domaines de formation :
 - ❖ Droit, Economie, Gestion
 - ❖ Civilisations, Lettres et Langues
 - ❖ Sciences de l'Homme et de la Société
 - ❖ Informatique, Sciences Cognitives et Applications
- ❖ 11 DUT, 22 mentions de Licence (+ 10 spécialités en langue), 21 Licences professionnelles et 25 mentions de masters dont 11 cohabilitées
logoNancyU_nancy2.jpg
- ❖ 18 800 étudiants : 70% sont originaires de Lorraine, 13% sont de nationalités étrangères
- ❖ 7 étudiants sur 10 sont en cycle Licence, 2 sur 10 en Master
- ❖ Plus de 2 000 grades de Licence et près de 1 000 grades de Master sont décernés chaque année
- ❖ 6 UMR, 17 équipes d'accueil, 590 doctorants, plus de 60 thèses par an
- ❖ 722 enseignants, 481 personnels administratifs et 20 millions d'euros de budget

LES DONNEES DU DOMAINE FORMATION

- ❖ **Les données de scolarité sont issues :**
 - ❖ de documents papier (dossiers d'inscription, de candidature, enquêtes d'insertion, dossier d'habilitation, conventions de stage, ...)
 - ❖ de questionnaires « internet » (inscription en ligne, évaluation des enseignements, ...)
 - ❖ de fichiers pré-existants (OPI, fiche de service, ...)
 - ❖ d'actes de gestion (résultats d'examens, charges d'enseignement, ...)

LES ACTEURS DU DOMAINE FORMATION

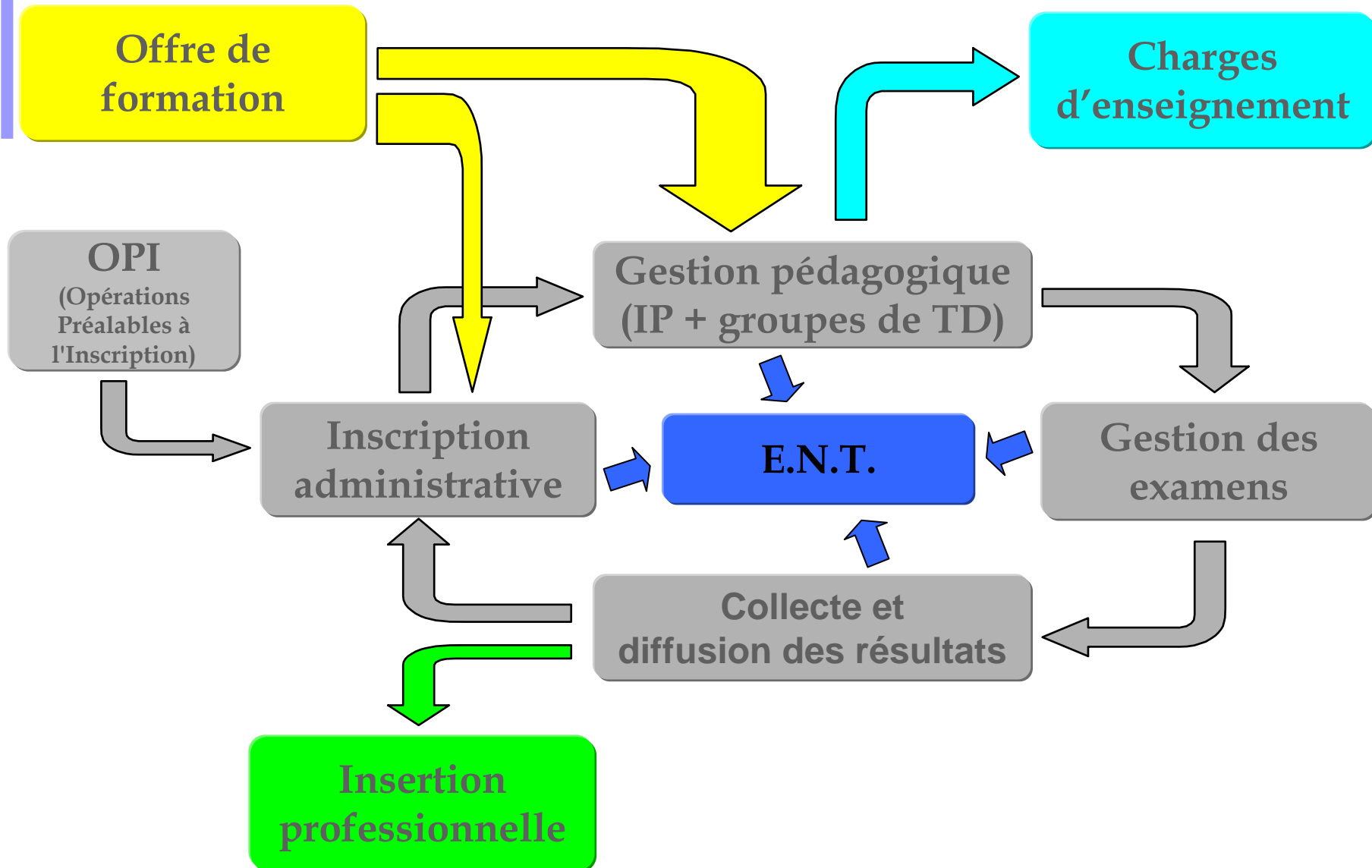
- ❖ **Les données sont produites :**
 - ❖ **par les étudiants (dossier d'inscription, enquête, évaluation, ...)**
 - ❖ **par les enseignants (offre de formation, service d'enseignement, notation, ...)**

- ❖ **Les données sont traitées et exploitées principalement au travers du logiciel APOGEE par le personnels administratif (et enseignant)**

L'UTILISATION DES DONNEES

- ❖ **Les données sont utilisées dans les actes de gestion quotidiens qui permettent de gérer la scolarité des étudiants (Inscriptions, emplois du temps, examens, délivrance de diplôme).**
- ❖ **Les données sont utilisées par l'intermédiaire des indicateurs « formations » pour piloter l'établissement :**
 - ❖ Taux de réussite, taux d'évaporation → amélioration des dispositifs pédagogiques
 - ❖ Evolution des effectifs, taux d'attractivité, taux d'insertion → adaptation et évolution de l'offre de formation
 - ❖ Effectifs et taux d'encadrement → redéploiement de postes

UNE MULTITUDE D'ACTES DE GESTION



POURQUOI FIABILISER LES DONNEES

- ❖ Pour garantir un pilotage fondé sur des indicateurs fiables, précis, vérifiables et cohérents
 - ❖ définir les objectifs / les orientations stratégiques
 - ❖ comment atteindre ces objectifs / identifier les actions à mettre en œuvre
 - ❖ suivre la réalisation des objectifs
- ❖ Pour une meilleure image et une appropriation des procédures par tous les acteurs de l'établissement
 - ❖ Pour un meilleur service rendu à l'étudiant

DES DONNEES A FIABILISER ?

Echelle du risque de détérioration des données :

de 1 à 5 - 1 : détérioration forte de la donnée ; 5 : absence de biais lors de la saisie de la donnée

❖ **Données issues de dossiers papier : 2**

- ❖ Moins : problème d'harmonisation de la saisie par les opérateurs / problème d'interprétation du questionnaire / faible possibilité de contrôle de cohérence dans le document lors de son remplissage.
- ❖ Plus : permet le retraitement de la donnée par les opérateurs de saisie (possibilité d'harmonisation de l'interprétation).

❖ **Données issues de formulaires « internet » : 3**

- ❖ Moins : fiabilité des informations communiquées (niveau de maîtrise de celles-ci par l'étudiant).
- ❖ Plus : possibilité d'intégrer les processus de contrôle dès la saisie (contrôle de cohérence).

❖ **Données issues de fichiers pré-existants : 4**

- ❖ Moins : possibilité d'une mauvaise connaissance de l'origine des données.
- ❖ Plus : les fichiers ayant déjà servi dans le cadre d'un acte de gestion ont déjà été l'objet d'une procédure de contrôle et sont la résultante d'actes de gestion.

❖ **Données issues d'actes de gestion : 5**

- ❖ Plus : la donnée (l'information) étant produite par l'acte de gestion, elle n'est pas soumise à dégradation mais est enrichie par celui-ci.
- ❖ L'université possédant la maîtrise de l'acte de gestion, elle peut s'assurer de la qualité des informations produites.

❖ Les données peuvent être altérées par :

- ❖ une mauvaise connaissance par le questionné de la donnée (année d'entrée dans l'enseignement supérieur, CSP des parents, ...)
- ❖ une erreur d'interprétation ou des difficultés pour transformer la donnée par l'enquêté (dernier établissement fréquenté, CSP des parents, régime de sécurité sociale des parents, ...)
- ❖ une erreur de saisie (année du bac, notes, ...)
- ❖ une duplication de la donnée (étudiants en double, enseignement dupliqué, ...)
- ❖ une redondance de l'information ou une multiplication de l'origine de l'information (offre de formation)

LES CONTRAINTES

❖ Des difficultés rencontrées :

- ❖ pour assurer la rigueur de la saisie des informations dans les bases de données
- ❖ pour faire respecter le calendrier pour la saisie et la remontée des données au niveau central
- ❖ pour maintenir sur la durée la cohérence des données et leur suivi régulier

FIABILISER LES DONNEES

❖ Les actions mises en œuvre pour fiabiliser les données (1) :

❖ Reformuler les questions sur lesquelles les données subissent une forte dégradation

- | | |
|-------------------------|-----------------------|
| ❖ L'analyse des données | L'analyse des erreurs |
| ❖ La correction | Le contrôle |

❖ Mettre en place des contrôles de cohérence au moment de la saisie

- ❖ L'analyse des données et leurs recoupements : *mettre en place des variables de contrôle*
- ❖ Les contrôles de cohérence : *tout ce qui est contrôlable par algorithme doit être contrôlé*

❖ Mettre en place des contrôles de cohérence a posteriori

- ❖ La surveillance : *mise en place d'un processus de suivi*
- ❖ L'amélioration : *acquisition des « bonnes pratiques »*

FIABILISER LES DONNEES

❖ Les actions mises en œuvre pour fiabiliser les données (2) :

❖ Éviter la (re)saisie de données existantes par ailleurs

- ❖ Identifier les origines multiples des données
- ❖ Mettre en relation les différentes sources de données (*fichier OPI*)
- ❖ S'assurer de l'unicité de stockage de l'information (*1 information = 1 source*)
- ❖ Intégrer les différentes bases de données (*Entrepôt de données*)

❖ Saisir au plus près de la source

- ❖ Impliquer les étudiants dans la saisie des données les concernant
- ❖ Faciliter le contrôle des données par les personnes concernées
- ❖ Favoriser la mise à disposition rapide de l'information
- ❖ Solliciter la mise à jour des informations par l'intéressé
- ❖ Faire connaître les enjeux afin de s'assurer de la motivation des étudiants
- ❖ Mettre en place un mécanisme de retour sur la qualité des données

FIABILISER LES DONNEES

- ❖ **Les actions mises en œuvre pour fiabiliser les données (3) :**
 - ❖ **Intégrer la donnée dans les actes de gestion quotidiens des services administratifs**
 - ❖ Vérifier la réalité et l'utilité de la donnée ou de l'information
 - ❖ Faire interagir les actes de gestion de différents domaines d'activité entre eux
 - ❖ **Positionner l'opérateur dans le processus complet de circulation des données**
 - ❖ Permettre aux personnes de se situer dans l'ensemble des actes de gestion et dans le processus de production de l'indicateur
 - ❖ Former aux processus
 - ❖ **Associer les personnes responsables de la saisie à l'utilisation faite de l'information**
 - ❖ Sensibiliser les personnes sur les indicateurs produits à l'aide de leurs actions
 - ❖ Engager la responsabilité des contributeurs

LES LIMITES

❖ Des limites, nécessairement ...

❖ Identifier les données sensibles

- ❖ Définir le processus à mettre en œuvre pour atteindre le niveau de qualité souhaité
- ❖ Vérifier la faisabilité de la mise en œuvre du processus et sa pérennité dans les actes de gestion quotidien

❖ Savoir renoncer à l'acquisition de l'information

- ❖ Vérifier la réalité de la donnée
- ❖ Définir l'utilité de la donnée
- ❖ Vérifier la complexité du processus à mettre en œuvre
- ❖ Connaître le coût du processus

CONCLUSION

❖ **Un projet ?**

❖ **Une méthode ?**

❖ **Un enjeu ?**