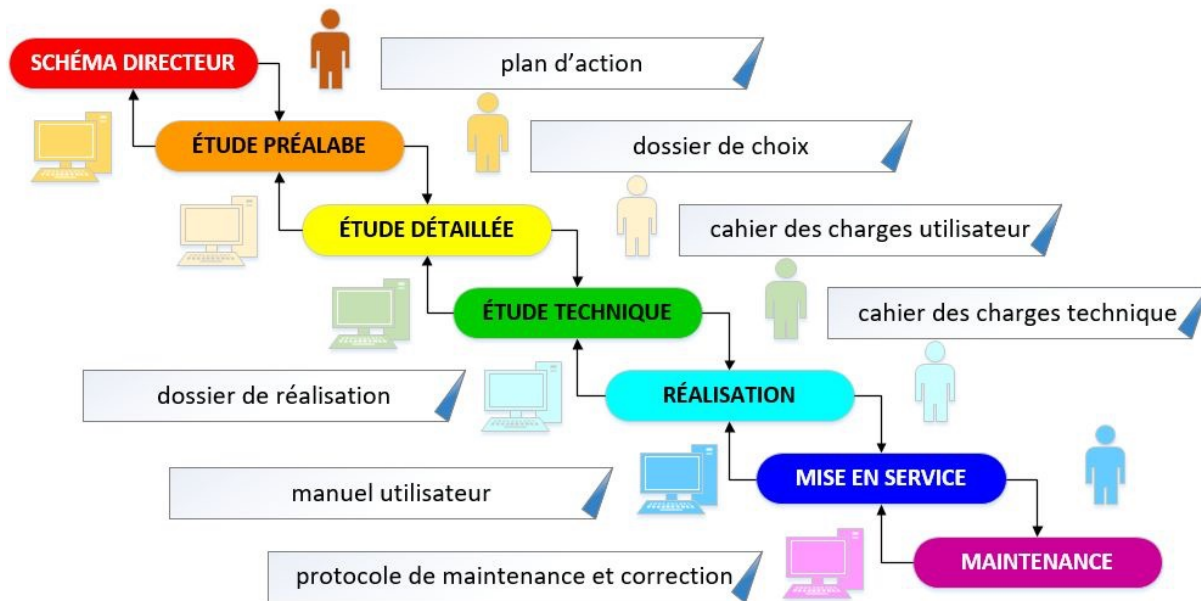


# Approche informatique classique Adaptable aux collection de données

SYNOPTIQUE DE LA MÉTHODE MERISE



Problème avec des corpus conséquents  
(exemple ISTEK, 20.000.000 documents)

Une application informatique est  
nécessaire pour connaître les données.

Assemblage de progiciels  
Méthodologie SCRUM

Pour les collections, archives...  
Systèmes de Gestion de Bibliothèque  
Données non transactionnelles  
Oméka...

**MAIS : Une collection de donnée est finie  
Un système de connaissances est infini**

# Un ouvrage de référence = Une approche spécifique (par partie)

- ▶ Exemple Edition Critique de Léon Gautier (1872)
- ▶ Volume 1 :
  - Introduction historique
    - 201 pages, 20 chapitres (à annoter par des liens sémantiques)
  - Edition critique (toutes les laisses : transcription / traduction)
    - 320 pages => réparties dans les pages laisses du manuscrit d'Oxford
- ▶ Volume 2 :
  - Notes et variantes (organisées par vers)
    - 270 pages distribuées dans les laisses (avec quelques articles)
    - Rédactions à améliorer pour meilleure lisibilité (abréviations...)
  - Glossaire (par mots en français ancien avec graphies et liste de vers)
    - 200 pages, 20 entrées par page, sous 3 formes : facsimilé, texte aéré, par page mot
  - Table des matières (avec les liens vers les pages du volume 2)



# MediaWiki, moteur de Wikipédia moteur de Wicri/Chanson de Roland

- ▶ Plusieurs dizaines de milliers de sites
- ▶ Développés par des contributeurs



# Bilan, quelques chiffres

- ▶ WICRI (10 ans) environ 150 wikis (dont CIDE H2PTM)
  - 100.000 pages dont 30.000 significatives
  - 150 serveurs d'explorations (400.000 documents ISTEEX + PubMed + ...)
- ▶ Wicri/Chanson de Roland (16 mois)
  - 7.383 pages, dont 2.524 significatives, 32.900 items sémantiques
  - Premiers ouvrages (squelette / contenu / présentation) :
    - Chanson de Roland par F. Michel, annotée Paul Meyer (100% / 100% / 50%)
    - Roman de Roncevaux par F. Michel, annotée Paul Meyer (100% / 5% / 5%)
    - Chanson de Roland, édition populaire de Léon Gautier 1895 (100% / 100% / 50%)
    - Manuscrit d'Oxford (100% / 30% / 20%)
    - Manuscrit de Paris (100% / 20% / 10%)
    - Oratorio de Gilles Mathieu (99% / 40% / 30%)
    - Manuscrit de Châteauroux (20% / 5% / 2%)
  - **Tout a été réalisé par des contributions MediaWiki**
    - (pas de développement informatique)

# Comment déployer cette approche

## L'autonomie numérique pour les humanités

### ▶ Nos expériences :

- JD : 7000 pages wikis en un an **sans une ligne de développement**
  - Comme étudiant (praticien débutant) en :
    - philologie, linguistique, musicologie, etc
- Etudiants en SIC/Humanités : 3 mois de projet/stage avec assistance JD
- Les mots de l'Agronomie 2 fois 3 jours + assistance

### ▶ Questions

- Comment créer un réseau de learning centers
  - Avec un noyau informatique restreint mais de très haut niveau
- Comment expérimenter dans les sciences de l'information
  - Exemple : colloques CIDE, expérience Wivri/Maroc

# Wicri réflexion alternative et orthogonale

- ▶ Galaxie Wikipédia,
  - Des services classiques
    - encyclopédie, dictionnaire, bibliothèque
  - Gigantisme pluri disciplinaire
  - Mécanismes de régulation souvent obscurs
- ▶ Quelques illusions du Web sémantique
  - La combinaison des triplestores a-t-elle un sens ?
    - A quoi sert un triplestore généré par voie automatique à 60 de fiabilité ?
  - WikiData référentiel universel et éternel
    - Le même pour les récits historiques et imaginaires ?
    - Le même pour les musulmans, les chrétiens, les athées ?
    - Alimenté par les services secrets russes, extrémistes, etc..

