

Harmoniser la production des données culturelles

ou

Comment un projet documentaire devient un projet structurant

Katell Briatte

Ministère de la Culture et de la Communication

Séminaire VocabNomen, Poitiers, 04 mai 2010

- **Les origines du projet**
- **Les objectifs**
- **Définition du projet**
- **État d'avancement**
- **Impacts attendus**

Les origines du projet

▪ L'état des lieux

- Des corpus « historiques »
 - ✓ Un volume de données considérable
 - ✓ Une grande variété de ressources documentaires
 - ✓ Des technologies et des modèles anciens
- Une multiplicité d'acteurs
 - ✓ Des « disciplines » nombreuses
 - ✓ Une production de plus en plus externalisée
- Des objectifs de production variés
 - ✓ Connaître le patrimoine culturel
 - ✓ Le préserver
 - ✓ Le faire connaître au plus grand nombre
- Des procédures et des outils divers
 - ✓ Un système d'information calqué sur l'organisation
 - ✓ Un Système d'Information (SI) « en silos »

▪ Des réalisations

- Un cadre stratégique
 - ✓ L'axe « documentaire » : une priorité du schéma directeur du Système d'Information depuis 2005
- Un accès « transverse » aux productions
 - ✓ Le moteur « Collection » de culture.fr (moteur sémantique *SineQua*)
 - ✓ « Généalogie »
 - ✓ « Atlas de l'architecture et du patrimoine »
- Développement de la politique d'accès aux collections et aux corpus numérisés
 - ✓ Contexte européen : « Patrimoine numérique » (MICHAEL, MINERVA, EUROPEANA)
 - ✓ Portail « Histoire des arts »
 - ✓ Emprunt national : axe « Numérisation, diffusion et valorisation des contenus culturels »

▪ Les constats

- Un exemple de recherche sur le moteur « Collection »
 - ✓ La recherche : « palais des Comtes de Poitiers »
 - ✓ Premiers résultats : 258 résultats « triés par ordre de pertinence »
 - Une série d'archives (plans, dossiers de travaux)
 - Des photographies
 - Un tableau représentant les « comtes de Poitiers » situé au château de Versailles
 - En 101ème position, la « notice d'œuvre » (notice de recensement MH)
 - ✓ Tous les mots dans la même phrase sans élargir aux termes proches : 168 résultats
 - En 21ème position, la « notice d'oeuvre »
- Satisfaction du besoin de l'internaute ?

- La diffusion « miroir déformant » de la production
 - ✓ Une information décontextualisée
 - Perte du sens donné par le contexte de production
 - Un implicite « métier » générateur de silence ou d'incohérence
 - ✓ Une information nivelée
 - Les œuvres
 - Les représentations des œuvres
 - Les discours ou les témoignages sur les œuvres ...
 - ✓ L'explosion de la « redondance »
- La diffusion « témoin » des besoins de la production ...
 - ✓ « cycle de vie » des objets culturels
 - Gestion du patrimoine culturel
- et « révélateur » de ses faiblesses
 - ✓ Multiplication des saisies, erreurs et incohérence

Les objectifs du projet

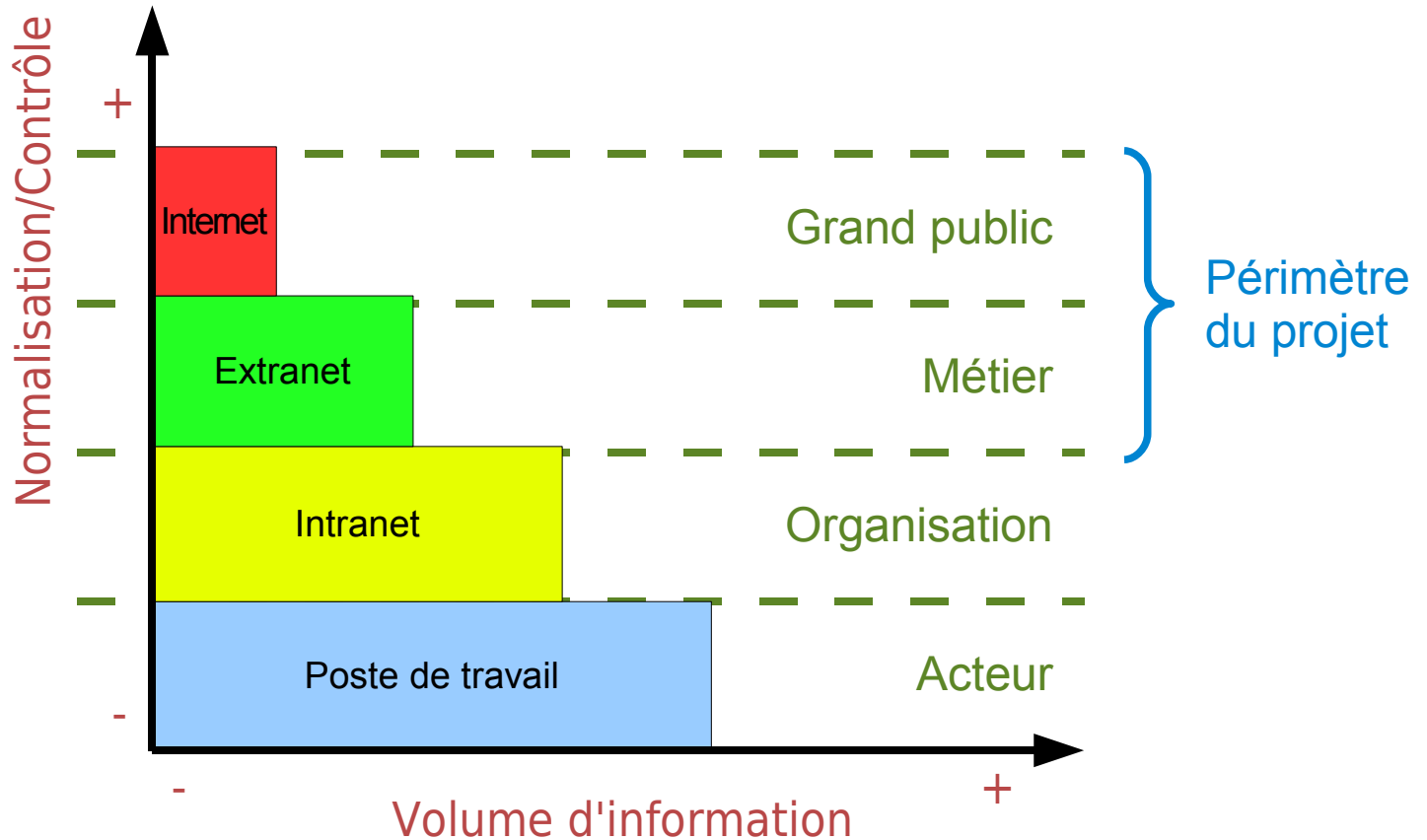
- **Agir sur la production pour**
 - mieux connaître et préserver
 - ✓ augmenter la qualité
 - ✓ augmenter la cohérence
 - ✓ augmenter la fiabilité
 - moderniser la production
 - ✓ rationaliser
 - ✓ automatiser au maximum les traitements et les contrôles
 - ✓ désimbriquer les processus de production et de diffusion
 - préparer la diffusion
 - ✓ expliciter l'implicite
 - ✓ exploiter la structure des données
 - ✓ + intégrer les problématiques d'accessibilité et de réutilisabilité des données publiques

▪ Les leviers

- Tutelle, contrôle scientifique et technique, partenariat
 - ✓ imposer ou préconiser
 - Différents niveaux de préconisation
 - ✓ contrôler
- La normalisation
 - ✓ produire des normes et/ou les préconiser
 - ✓ préconiser de manière cohérente
 - Élaborer un référentiel normatif
- La maîtrise du SI
 - ✓ harmoniser les procédures pour partager les outils et les infrastructures
 - ✓ organiser et planifier les flux
 - ✓ sécuriser le SI
 - ✓ rationaliser et pérenniser les investissements

Définition du projet

Définition du projet

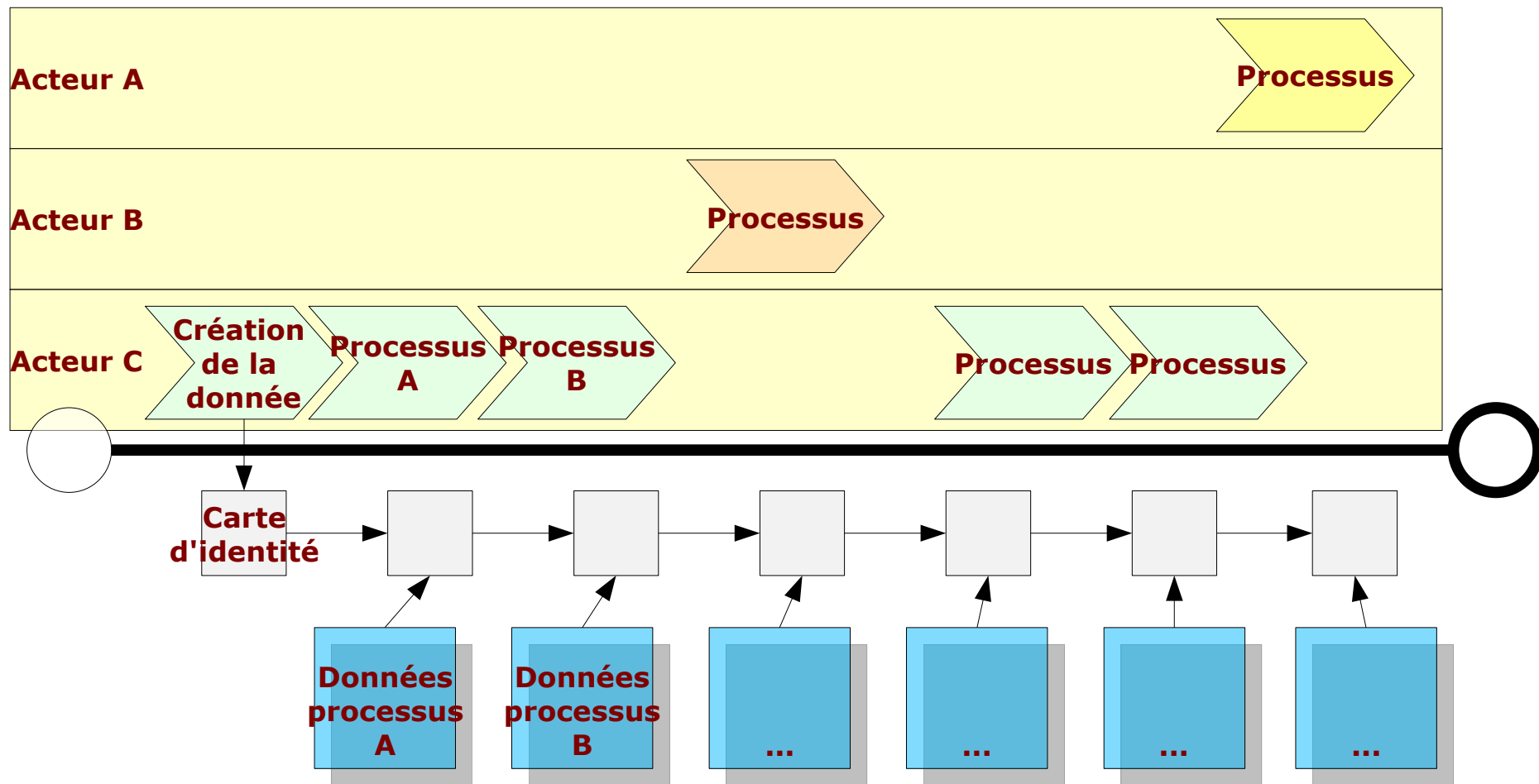


- **Un projet en trois volets :**
 - Les données
 - Les normes
 - Les processus

▪ **Le coeur du projet : la donnée**

- Des données offrant des garanties de qualité, exhaustives, fiables, accessibles et réutilisables
 - ✓ Supprimer les saisies multiples
 - ✓ Réduire la redondance des informations
 - ✓ Réduire les incohérences d'information
 - ✓ Automatiser les contrôles et les traitements
 - ✓ Faciliter les échanges
 - ✓ Faciliter la diffusion
 - ✓ Faciliter la réutilisation

= une problématique de « référentiel »



- **Construire le référentiel des données culturelles**
 - les « cartes d'identité » des œuvres et des objets culturels
 - ✓ Définir collectivement le plus grand commun dénominateur
 - ✓ Définir la structure des données
 - Granularité
 - Explicitation de l'implicite
 - ✓ Définir le contenu
 - Définition partagée des éléments
 - Partage des nomenclatures (listes d'autorité, terminologies destinées à alimenter ces éléments)
 - ✓ Mettre en place les processus d'alimentation et de gestion du référentiel
 - ✓ Mettre en place les règles de gouvernance
 - Certification

- **Au-delà des données d'identification, harmoniser des éléments de contenus**
 - Identifier les dénominateurs communs
 - ✓ Description
 - Description physique
 - Description conceptuelle
 - ✓ État
 - ✓ Historique
 - Événements
 - Dates
 - Acteurs
 - Lieux ...
 - Modéliser des structures de données communes
 - ✓ Vers un « modèle des données communes » du MCC
 - rendre notre SI documentaire intelligible à d'autres systèmes (**interopérabilité**)

- **Une problématique de référentiels partagés**
 - Des référentiels d'objets métiers
 - ✓ Acteurs
 - ✓ Ressources iconographiques, multimédia
 - ✓ Ressources bibliographiques
 - ✓ Référentiel juridique ...
 - Des nomenclatures
 - ✓ Tirer parti de la richesse interne
 - ✓ Mieux la partager et la faire connaître
 - Les actions
 - ✓ Harmoniser
 - ✓ Construire
 - ✓ Externaliser
 - ✓ Gouverner
 - ✓ Partager

▪ **Le constat**

- des productions peu normalisées
 - ✓ normes métier, normes de description
 - ✓ normes techniques
- des informations souvent mal structurées
 - ✓ mélange de données de gestion et de contenu
 - ✓ « strates » successives
- des données peu interopérables
 - ✓ sémantiquement
 - ✓ techniquement

▪ **Les orientations**

- Définir un cadre normatif pour la production ...
 - ✓ généraliser le recours aux normes et standards nationaux et internationaux
- ... tout en préservant la spécificité des métiers

▪ **Panorama normatif**

➤ référentiels généraux de l'administration

✓ Interopérabilité (RGI)

- normes techniques (échange, archivage, encodage, formats d'implémentation ...)
- normes de métadonnées (descriptives, techniques, administratives ...)

✓ Accessibilité (RGAA)

- nouveaux éléments de contenus

✓ Sécurité (RGS)

➤ normes métier

✓ Normaliser ce qui ne l'est pas

✓ Développer l'usage des normes existantes

➤ identifiants normalisés (ARK ?)

▪ **Le modèle des données communes du MCC comme norme de production ?**

État d'avancement

- **Ce qui est réalisé**
- **Ce qui est en cours**
- **Ce qui reste à faire**

- **Inventaire systématique des productions**
 - Fiches de recueil normalisées
 - ✓ Fonds
 - ✓ Acteurs
 - ✓ Processus
 - ✓ Normes
 - ✓ Nomenclatures
 - ✓ Structures de données
 - ✓ Textes juridiques
 - Analyse critique
 - Identification des axes de progrès

- **Définition du référentiel des données culturelles**
 - Structure de données
 - Processus de gestion du référentiel
 - ✓ processus de production
 - ✓ processus de certification
- **La méthode**
 - ✓ Des structures de données existantes vers une structure commune des données d'identification
 - Batterie d'exemples
 - ✓ Élaboration parallèle du dictionnaire de données
 - Des concepts partagés
 - ✓ Identification des nomenclatures candidates
 - ✓ Modélisation
 - Données (diagrammes de classe)
 - Processus (Modèles BPMN)

- **le modèle des données communes du MCC**
 - Définition de structures de données partagées
 - Définition sémantique des éléments
 - Modélisation
 - Règles d'usage
 - Exemples
- **les référentiels transverses**
 - Atelier « acteurs historiques »
 - ✓ Un référentiel partagé des auteurs
- **le cadre de certification**
 - Une méthodologie MCC en cours de définition
 - Un atelier de gouvernance à mettre en place

- **Mettre en œuvre**
 - Élaborer les scénarios de migration
 - Élaborer les documents de préconisations
 - ✓ conduire le changement
 - Outiller la production
 - ✓ outils de gestion, de traitement et de contrôle
 - ✓ services
 - Chantier de la reprise de l'existant
 - ✓ peupler les référentiels

- **Mettre en œuvre la gouvernance**
 - Gouvernance des processus
 - Gouvernance des référentiels
 - ✓ Une multitude de chantiers à ouvrir

Impacts attendus

▪ **pour les producteurs**

➤ Un retour sur investissement

- ✓ une « boîte à outils » technique et méthodologique
- ✓ un accès aisé au référentiel
- ✓ moins de saisie, moins de risques d'erreurs ou d'incohérences
- ✓ une pérennisation de leurs données

➤ Une plus grande adhésion au projet

- ✓ des échanges facilités
 - une connaissance claire et partagée de la structure de données « cible »
- ✓ un label, une vitrine nationale voire européenne pour leur données
 - une garantie d'accessibilité
 - un moteur de recherche performant

- **pour les gestionnaires de fonds**
 - simplification des traitements et des contrôles
 - des missions recentrées sur la qualité des contenus
- **pour les « métiers »**
 - gestion et suivi des œuvres facilités
 - vision globale et cohérente de l'action du MCC
 - consolidation nationale des données facilitée
 - ✓ pilotage et évaluation
- **Pour le grand public**
 - une offre plus cohérente, plus lisible

▪ **Pour le SI**

- des « briques » d'urbanisation réutilisables pour d'autres besoins
- un SI plus évolutif, plus facile à maintenir
- des flux mieux contrôlés, plus sécurisés
 - ✓ Procédures d'alimentation et rôles mieux définis

▪ **Impacts organisationnels**

- de nouvelles tâches et de nouvelles responsabilités liées à la mise en œuvre du référentiel des données culturelles
 - ✓ gouvernance

Katell Briatte

Chef de projet

Ministère de la culture et de la communication

DGP/DSIP

182 rue Saint-Honoré

75033 Paris cedex 01

katell.briatte@culture.gouv.fr

Corpus mentionnés :

- **portail grand public** : <http://www.culture.fr>

- **moteur Collections** :

http://www.culture.fr/fr/sections/collections/moteur_collections

- **Généalogie** : <http://www.culture.fr/fr/sections/collections/genealogie>

- **Histoire des arts** : <http://www.culture.fr/fr/sections/collections/histoire-arts>

- **Patrimoine numérique** : <http://www.numerique.culture.fr>

Harmoniser la production des données culturelles

ou
Comment un projet documentaire devient un projet structurant

Katell Briatte
Ministère de la Culture et de la Communication

Séminaire VocabNomen, Poitiers, 04 mai 2010

- **Les origines du projet**
- **Les objectifs**
- **Définition du projet**
- **État d'avancement**
- **Impacts attendus**

Les origines du projet

▪ L'état des lieux

- Des corpus « historiques »
 - ✓ Un volume de données considérable
 - ✓ Une grande variété de ressources documentaires
 - ✓ Des technologies et des modèles anciens
- Une multiplicité d'acteurs
 - ✓ Des « disciplines » nombreuses
 - ✓ Une production de plus en plus externalisée
- Des objectifs de production variés
 - ✓ Connaître le patrimoine culturel
 - ✓ Le préserver
 - ✓ Le faire connaître au plus grand nombre
- Des procédures et des outils divers
 - ✓ Un système d'information calqué sur l'organisation
 - ✓ Un Système d'Information (SI) « en silos »

▪ **Des réalisations**

- Un cadre stratégique
 - ✓ L'axe « documentaire » : une priorité du schéma directeur du Système d'Information depuis 2005
- Un accès « transverse » aux productions
 - ✓ Le moteur « Collection » de culture.fr (moteur sémantique *SineQua*)
 - ✓ « Généalogie »
 - ✓ « Atlas de l'architecture et du patrimoine »
- Développement de la politique d'accès aux collections et aux corpus numérisés
 - ✓ Contexte européen : « Patrimoine numérique » (MICHAEL, MINERVA, EUROPEANA)
 - ✓ Portail « Histoire des arts »
 - ✓ Emprunt national : axe « Numérisation, diffusion et valorisation des contenus culturels »

▪ Les constats

- Un exemple de recherche sur le moteur « Collection »
 - ✓ La recherche : « palais des Comtes de Poitiers »
 - ✓ Premiers résultats : 258 résultats « triés par ordre de pertinence »
 - Une série d'archives (plans, dossiers de travaux)
 - Des photographies
 - Un tableau représentant les « comtes de Poitiers » situé au château de Versailles
 - En 101ème position, la « notice d'œuvre » (notice de recensement MH)
 - ✓ Tous les mots dans la même phrase sans élargir aux termes proches : 168 résultats
 - En 21ème position, la « notice d'oeuvre »
- Satisfaction du besoin de l'internaute ?

- La diffusion « miroir déformant » de la production
 - ✓ Une information décontextualisée
 - Perte du sens donné par le contexte de production
 - Un implicite « métier » générateur de silence ou d'incohérence
 - ✓ Une information nivelée
 - Les œuvres
 - Les représentations des œuvres
 - Les discours ou les témoignages sur les œuvres ...
 - ✓ L'explosion de la « redondance »
- La diffusion « témoin » des besoins de la production ...
 - ✓ « cycle de vie » des objets culturels
 - Gestion du patrimoine culturel
- et « révélateur » de ses faiblesses
 - ✓ Multiplication des saisies, erreurs et incohérence

Les objectifs du projet

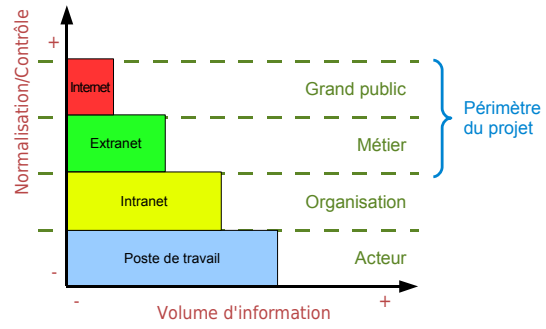
▪ Agir sur la production pour

- mieux connaître et préserver
 - ✓ augmenter la qualité
 - ✓ augmenter la cohérence
 - ✓ augmenter la fiabilité
- moderniser la production
 - ✓ rationaliser
 - ✓ automatiser au maximum les traitements et les contrôles
 - ✓ désimbriquer les processus de production et de diffusion
- préparer la diffusion
 - ✓ expliciter l'implicite
 - ✓ exploiter la structure des données
 - ✓ + intégrer les problématiques d'accessibilité et de réutilisabilité des données publiques

▪ Les leviers

- Tutelle, contrôle scientifique et technique, partenariat
 - ✓ imposer ou préconiser
 - Différents niveaux de préconisation
 - ✓ contrôler
- La normalisation
 - ✓ produire des normes et/ou les préconiser
 - ✓ préconiser de manière cohérente
 - Élaborer un référentiel normatif
- La maîtrise du SI
 - ✓ harmoniser les procédures pour partager les outils et les infrastructures
 - ✓ organiser et planifier les flux
 - ✓ sécuriser le SI
 - ✓ rationaliser et pérenniser les investissements

Définition du projet





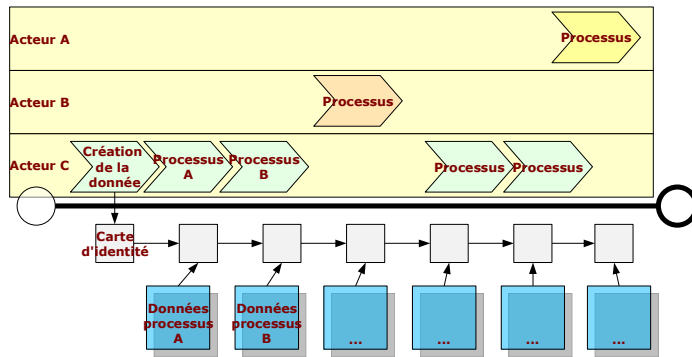
▪ **Un projet en trois volets :**

- Les données
- Les normes
- Les processus

▪ **Le coeur du projet : la donnée**

- Des données offrant des garanties de qualité, exhaustives, fiables, accessibles et réutilisables
 - ✓ Supprimer les saisies multiples
 - ✓ Réduire la redondance des informations
 - ✓ Réduire les incohérences d'information
 - ✓ Automatiser les contrôles et les traitements
 - ✓ Faciliter les échanges
 - ✓ Faciliter la diffusion
 - ✓ Faciliter la réutilisation

= **une problématique de « référentiel »**



- **Construire le référentiel des données culturelles**
 - les « cartes d'identité » des œuvres et des objets culturels
 - ✓ Définir collectivement le plus grand commun dénominateur
 - ✓ Définir la structure des données
 - Granularité
 - Explicitation de l'implicite
 - ✓ Définir le contenu
 - Définition partagée des éléments
 - Partage des nomenclatures (listes d'autorité, terminologies destinées à alimenter ces éléments)
 - ✓ Mettre en place les processus d'alimentation et de gestion du référentiel
 - ✓ Mettre en place les règles de gouvernance
 - Certification

- **Au-delà des données d'identification, harmoniser des éléments de contenus**
 - Identifier les dénominateurs communs
 - ✓ Description
 - Description physique
 - Description conceptuelle
 - ✓ État
 - ✓ Historique
 - Événements
 - Dates
 - Acteurs
 - Lieux ...
 - Modéliser des structures de données communes
 - ✓ Vers un « modèle des données communes » du MCC
 - rendre notre SI documentaire intelligible à d'autres systèmes (**interopérabilité**)

▪ **Une problématique de référentiels partagés**

- Des référentiels d'objets métiers
 - ✓ Acteurs
 - ✓ Ressources iconographiques, multimédia
 - ✓ Ressources bibliographiques
 - ✓ Référentiel juridique ...
- Des nomenclatures
 - ✓ Tirer parti de la richesse interne
 - ✓ Mieux la partager et la faire connaître
- Les actions
 - ✓ Harmoniser
 - ✓ Construire
 - ✓ Externaliser
 - ✓ Gouverner
 - ✓ Partager

▪ **Le constat**

- des productions peu normalisées
 - ✓ normes métier, normes de description
 - ✓ normes techniques
- des informations souvent mal structurées
 - ✓ mélange de données de gestion et de contenu
 - ✓ « strates » successives
- des données peu interopérables
 - ✓ sémantiquement
 - ✓ techniquement

▪ **Les orientations**

- Définir un cadre normatif pour la production ...
 - ✓ généraliser le recours aux normes et standards nationaux et internationaux
- ... tout en préservant la spécificité des métiers

- **Panorama normatif**

- référentiels généraux de l'administration

- ✓ Interopérabilité (RGI)

- normes techniques (échange, archivage, encodage, formats d'implémentation ...)
- normes de métadonnées (descriptives, techniques, administratives ...)

- ✓ Accessibilité (RGAA)

- nouveaux éléments de contenus

- ✓ Sécurité (RGS)

- normes métier

- ✓ Normaliser ce qui ne l'est pas

- ✓ Développer l'usage des normes existantes

- identifiants normalisés (ARK ?)

- **Le modèle des données communes du MCC
comme norme de production ?**

État d'avancement



- **Ce qui est réalisé**
- **Ce qui est en cours**
- **Ce qui reste à faire**

- **Inventaire systématique des productions**

- Fiches de recueil normalisées
 - ✓ Fonds
 - ✓ Acteurs
 - ✓ Processus
 - ✓ Normes
 - ✓ Nomenclatures
 - ✓ Structures de données
 - ✓ Textes juridiques
- Analyse critique
- Identification des axes de progrès

- **Définition du référentiel des données culturelles**
 - Structure de données
 - Processus de gestion du référentiel
 - ✓ processus de production
 - ✓ processus de certification
- **La méthode**
 - ✓ Des structures de données existantes vers une structure commune des données d'identification
 - Batterie d'exemples
 - ✓ Élaboration parallèle du dictionnaire de données
 - Des concepts partagés
 - ✓ Identification des nomenclatures candidates
 - ✓ Modélisation
 - Données (diagrammes de classe)
 - Processus (Modèles BPMN)

- **le modèle des données communes du MCC**
 - Définition de structures de données partagées
 - Définition sémantique des éléments
 - Modélisation
 - Règles d'usage
 - Exemples
- **les référentiels transverses**
 - Atelier « acteurs historiques »
 - ✓ Un référentiel partagé des auteurs
- **le cadre de certification**
 - Une méthodologie MCC en cours de définition
 - Un atelier de gouvernance à mettre en place

- **Mettre en œuvre**
 - Élaborer les scénarios de migration
 - Élaborer les documents de préconisations
 - ✓ conduire le changement
 - Outiller la production
 - ✓ outils de gestion, de traitement et de contrôle
 - ✓ services
 - Chantier de la reprise de l'existant
 - ✓ peupler les référentiels
- **Mettre en œuvre la gouvernance**
 - Gouvernance des processus
 - Gouvernance des référentiels
 - ✓ Une multitude de chantiers à ouvrir

Impacts attendus

▪ pour les producteurs

- Un retour sur investissement
 - ✓ une « boîte à outils » technique et méthodologique
 - ✓ un accès aisé au référentiel
 - ✓ moins de saisie, moins de risques d'erreurs ou d'incohérences
 - ✓ une pérennisation de leurs données
- Une plus grande adhésion au projet
 - ✓ des échanges facilités
 - une connaissance claire et partagée de la structure de données « cible »
 - ✓ un label, une vitrine nationale voire européenne pour leur données
 - une garantie d'accessibilité
 - un moteur de recherche performant

- **pour les gestionnaires de fonds**
 - simplification des traitements et des contrôles
 - des missions recentrées sur la qualité des contenus
- **pour les « métiers »**
 - gestion et suivi des œuvres facilités
 - vision globale et cohérente de l'action du MCC
 - consolidation nationale des données facilitée
 - ✓ pilotage et évaluation
- **Pour le grand public**
 - une offre plus cohérente, plus lisible

- **Pour le SI**

- des « briques » d'urbanisation réutilisables pour d'autres besoins
- un SI plus évolutif, plus facile à maintenir
- des flux mieux contrôlés, plus sécurisés
 - ✓ Procédures d'alimentation et rôles mieux définis

- **Impacts organisationnels**

- de nouvelles tâches et de nouvelles responsabilités liées à la mise en œuvre du référentiel des données culturelles
 - ✓ gouvernance



Katell Briatte
Chef de projet
Ministère de la culture et de la communication
DGP/DSIP
182 rue Saint-Honoré
75033 Paris cedex 01
katell.briatte@culture.gouv.fr

Corpus mentionnés :

- portail grand public : <http://www.culture.fr>
- moteur Collections : http://www.culture.fr/fr/sections/collections/moteur_collections
- Généalogie : <http://www.culture.fr/fr/sections/collections/genealogie>
- Histoire des arts : <http://www.culture.fr/fr/sections/collections/histoire-arts>
- Patrimoine numérique : <http://www.numerique.culture.fr>