

# Projet de Référentiel de Métadonnées Statistiques (RMÉS)

Marc Cohen-Solal  
Division Coordination de la  
méta-information  
4 Avril 2012





# Présentation des standards internationaux

**et ...**

**de leur mise en œuvre dans le  
futur référentiel de métadonnées  
statistiques à l'Insee (RMÉS)**



# Qu'est ce qu'une norme ou un standard ?

## ISO (International Organization for Standardization)

- › Au sens ISO, une norme est un document établi par consensus et approuvé par un organisme reconnu qui fournit, pour des usages communs et répétés :
- › un langage commun, des définitions, des règles, des lignes directrices précises, voire des spécifications techniques.



# L'utilité des normes en statistiques (1)

## Essor dans les années 2000 des initiatives de standardisation

1995	2000	2002	2004	2006	2008	2010	2012
<b>SDMX</b>	2000 : Naissance de l'initiative <b>SDMX</b>		2004 : <b>SDMX 1.0</b>		2007 : <b>SDMX 2.0</b>		2011 : <b>SDMX 2.1</b>
<b>DDI</b> 1995 : naissance de l'alliance <b>DDI</b>	2000 : <b>DDI 1.0</b>				2008 : <b>DDI 3.0</b>		
<b>Neuchatel</b>	2000 : <b>Neuchâtel 1.0</b>	2002 : <b>Neuchatel 2.0</b>	2004 : <b>Neuchatel 2.1</b>				
<b>Autres Initiatives</b> 1995 : Publication du <b>Dublin Core</b>			2003 : <b>Dublin Core</b> reconnue par la norme ISO		2009 : METIS publie le <b>CMF</b>	2010 : projet <b>CORA</b> naissance <b>HLG-BAS</b>	2012 : projet <b>CORE</b>



## L'utilité des normes en statistiques (2)

- › Après les données, la demande se porte aussi sur les métadonnées : besoin de comparabilité, de qualité des données
- › Ce n'est pas un hasard que les normes statistiques ont vu le jour dans les data archive ou dans le monde des organismes internationaux. Il y a bien des usages répétés et communs à plusieurs organismes
- › Les organismes internationaux ou les data archive reçoivent et échangent beaucoup de données de multiples organismes. Ils ont besoin d'un format standard et même plus, d'une organisation standard de leur documentation ou de leur manière de structurer l'information (SDMX ou DDI)



## L'utilité des normes en statistiques (3)

- › ● DDI (Data Documentation Initiative)
- › ● SDMX (cf. exposé de M. Gauduffe)
- › ● Dublin core (norme d'archivage)
- › ● Neuchâtel (terminologie et modèle conceptuel de métadonnées pour les nomenclatures)
- GSBPM (Generic Statistical Business Process Model ou modélisation de l'activité métier de l'ensemble du processus statistique)



# GSBPM (Generic Statistical Business Process Model)

L'opération statistique typique qui suit l'ensemble de ces étapes, c'est l'enquête statistique.

**Les premières étapes jusqu'à « collecter » concernent les données individuelles. Les étapes 5, 6 et 7 sont plutôt sur les données agrégées.**

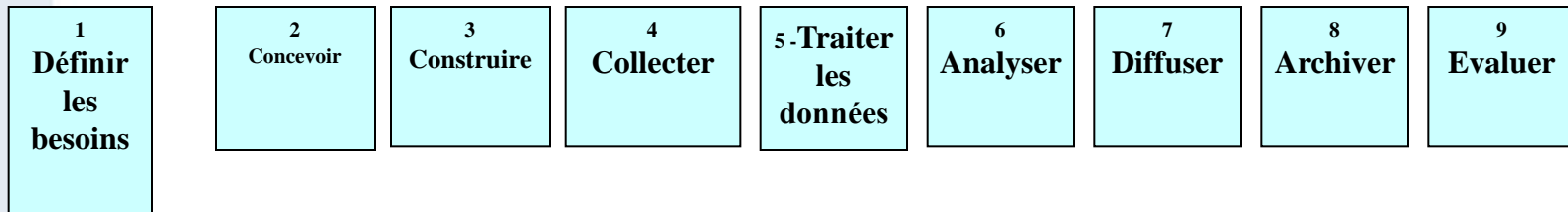
SDMX tout à fait adéquat pour les données agrégées (5, 6 et 7) mais n'est pas adapté aux données individuelles.

DDI est adapté aux données individuelles (1, 2, 3 et 4). En particulier, pour élaborer un questionnaire

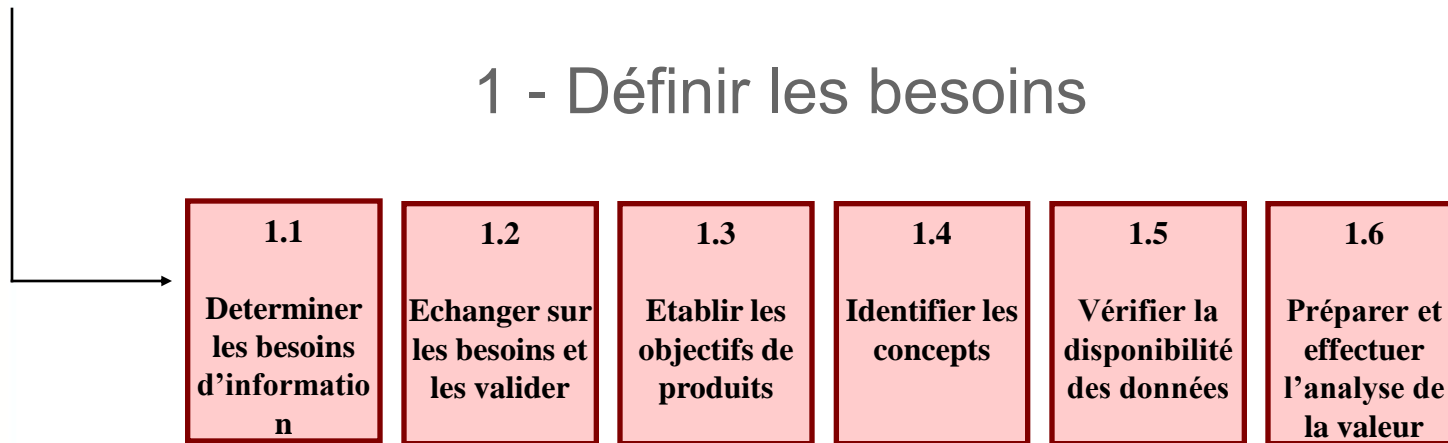
<b>1 - Définir les besoins</b>
<b>2 - Concevoir</b>
<b>3 - Construire</b>
<b>4 - Collecter</b>
<b>5 - Traiter</b>
<b>6 - Analyser</b>
<b>7 - Diffuser</b>
<b>8 - Archiver</b>
<b>9 - Évaluer</b>



# PROCESSUS ET SOUS PROCESSUS

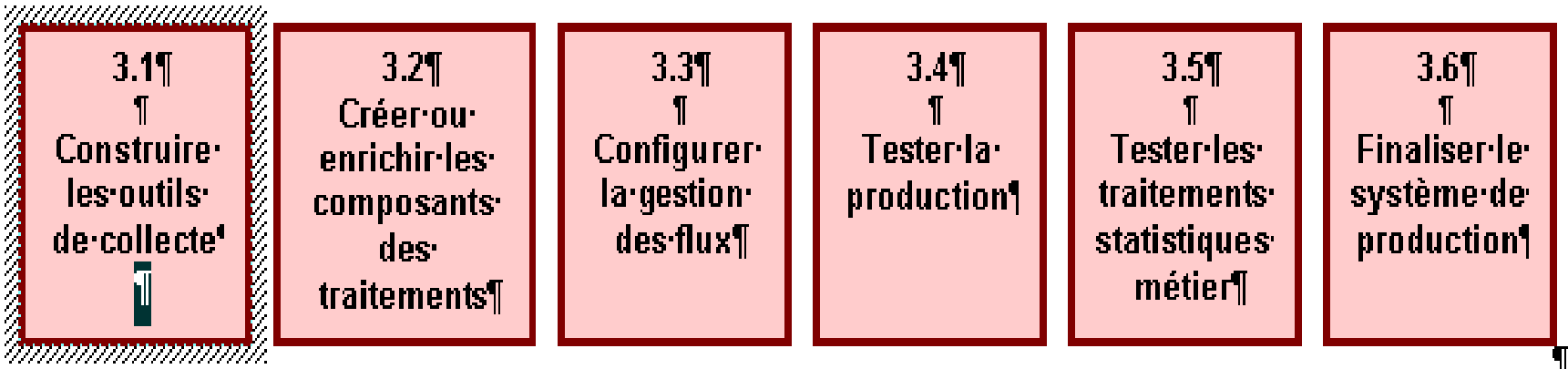
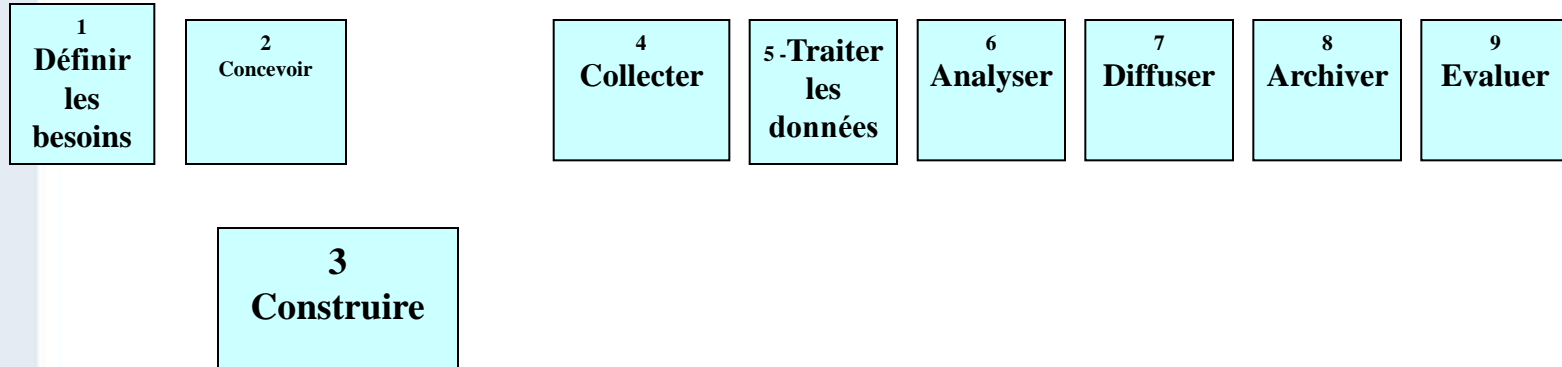


## 1 - Définir les besoins



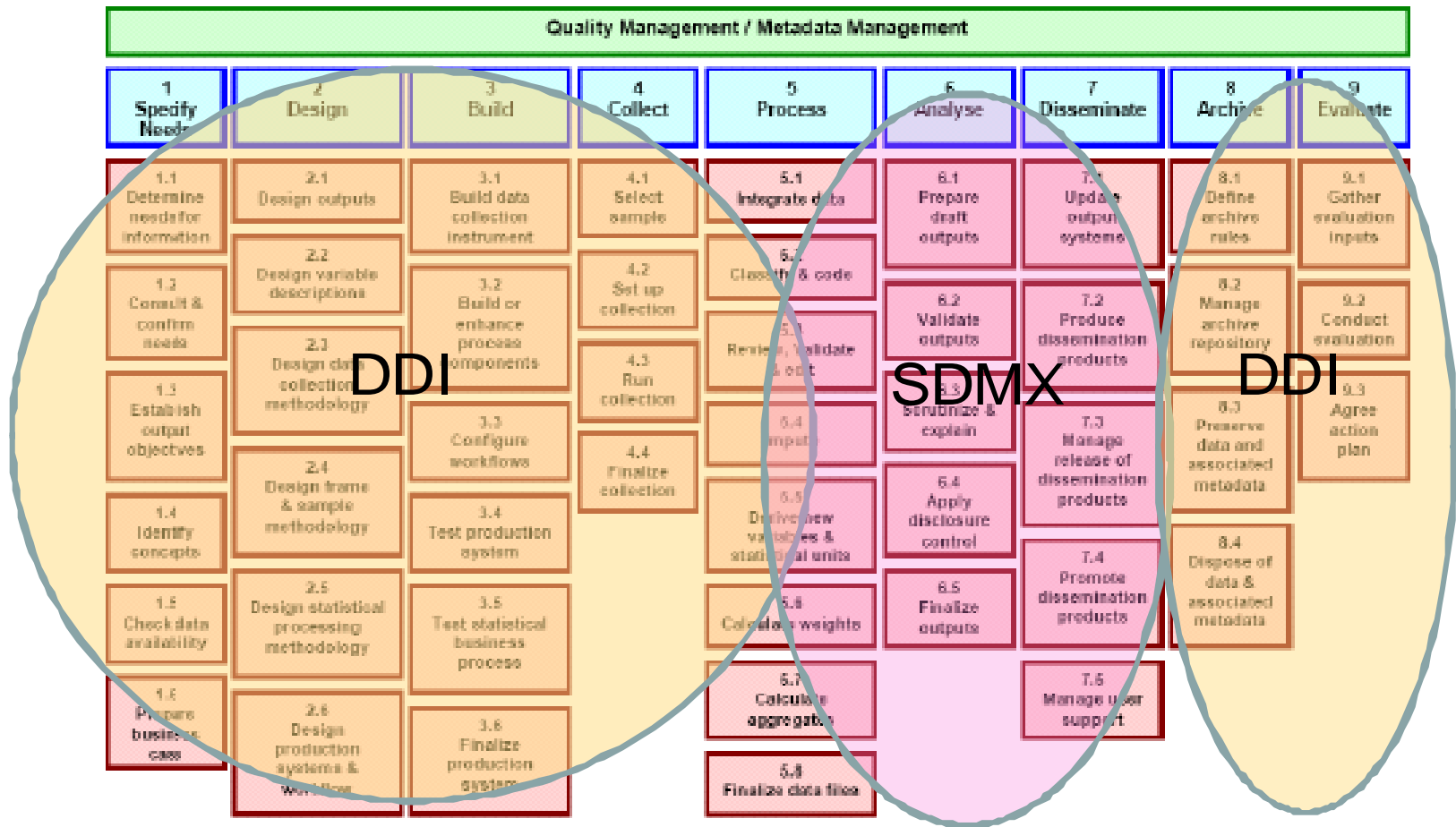


# PROCESSUS ET SOUS PROCESSUS





# GSBPM, SDMX et DDI





# Les objectifs de RMÉS

- 1 - S'orienter vers un référentiel de métadonnées statistiques** : c'est l'information dans RMÉS qui fera référence (comme aujourd'hui la rubrique définitions du site insee.fr)
- 2 - Favoriser la mutualisation et la normalisation en gestion** : éviter de mal réinventer, faciliter la transmission de savoir faire, le portage entre différentes applications, etc.
- 3 - Disposer d'un suivi des opérations statistiques**



# Les premières réalisations prévues dans le référentiel de métadonnées statistiques à l'Insee

- › La gestion des concepts avec leurs définitions en anglais et en français
- › La gestion des nomenclatures : établissement d'une nouvelle nomenclature, table de passage entre ancienne et nouvelle nomenclature
- › Les questionnaires d'enquêtes, avec les questions, les filtres, les variables attachées aux questions, les modalités des variables



# Les nomenclatures et RMÉS

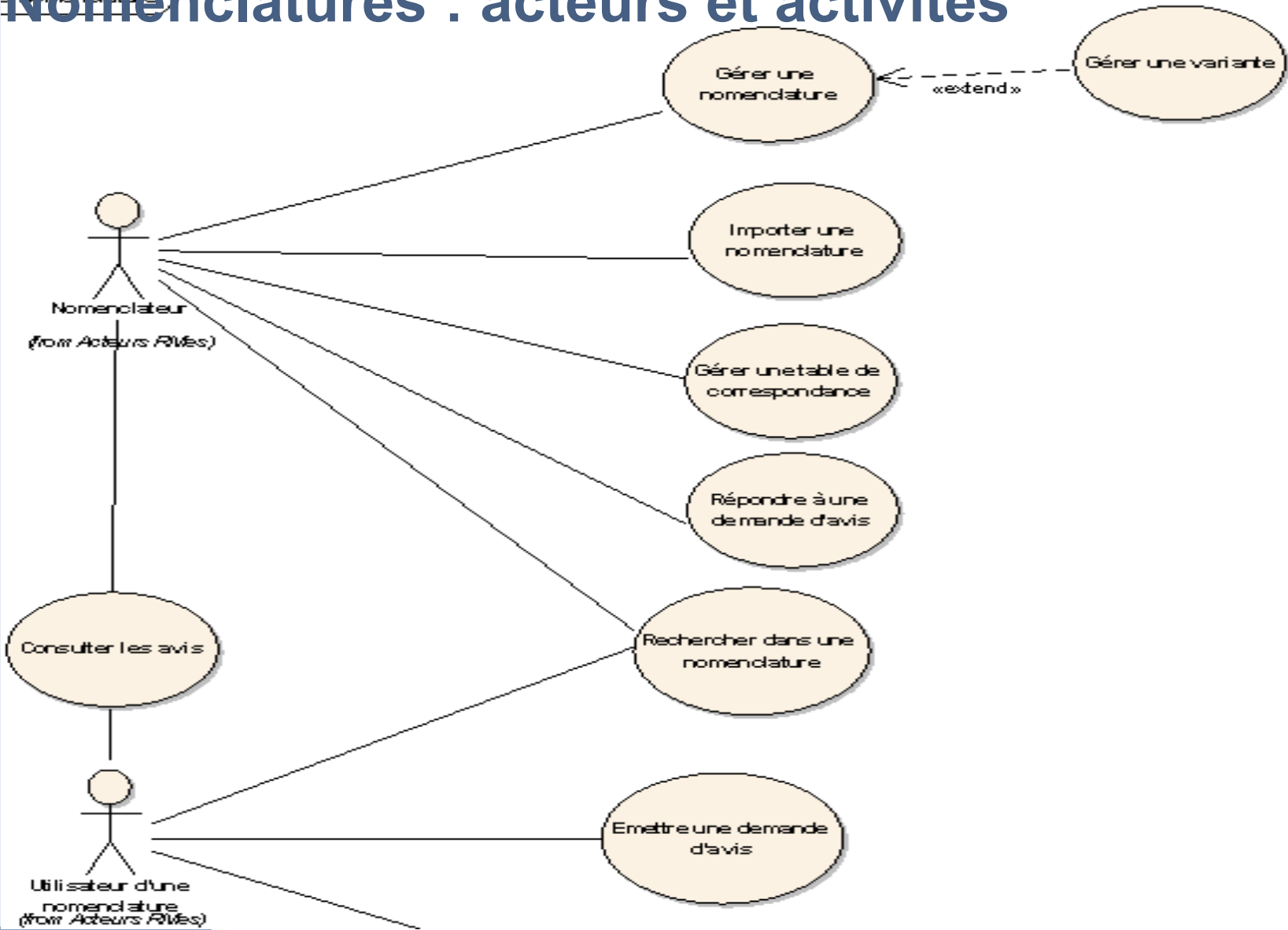
› Deux activités métier ont été identifiées :

⇒ 1. Gérer les nomenclatures officielles, les tables de passage et les variantes

⇒ 2. Mettre à disposition les nomenclatures et les tables de passage



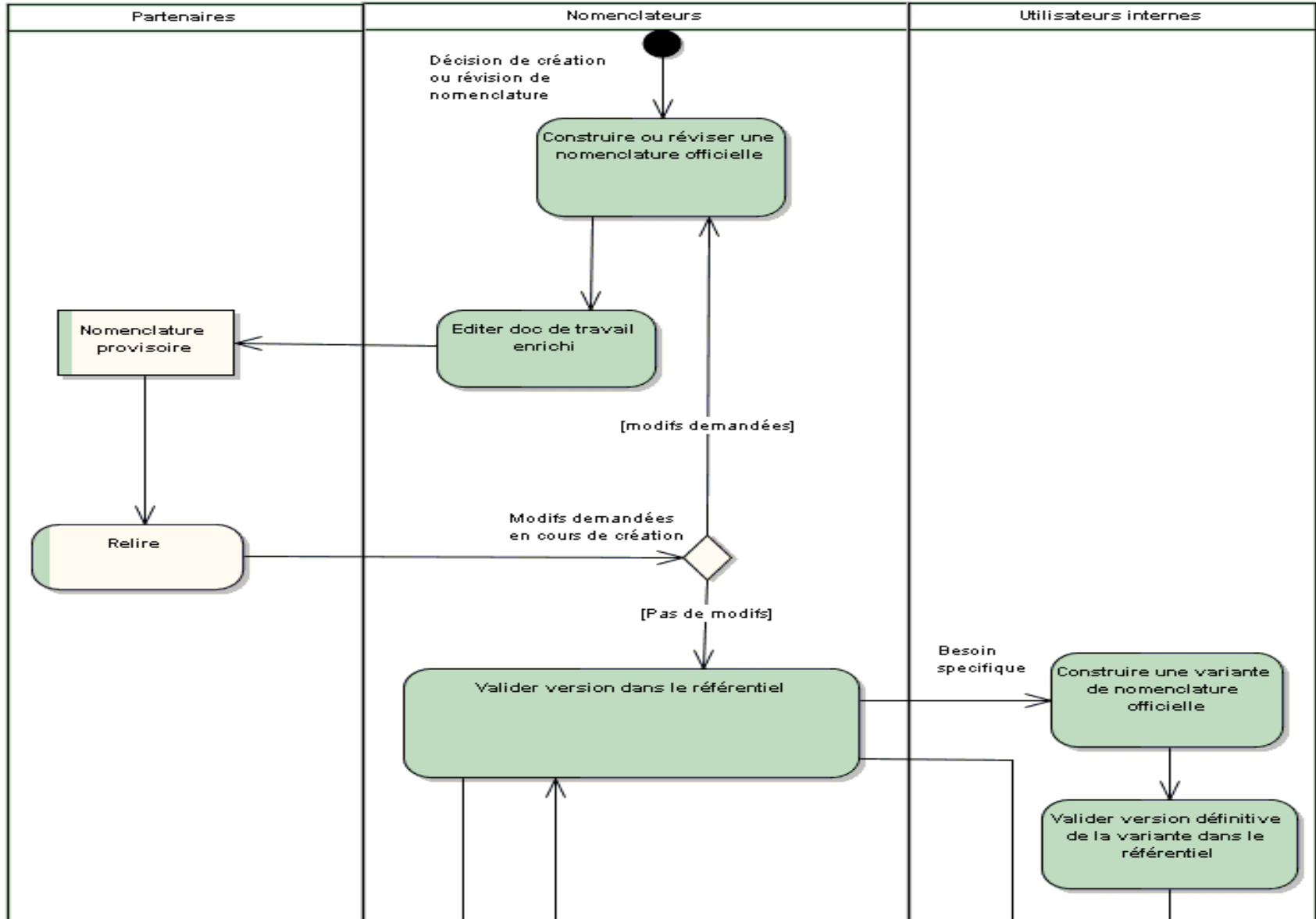
# Nomenclatures : acteurs et activités



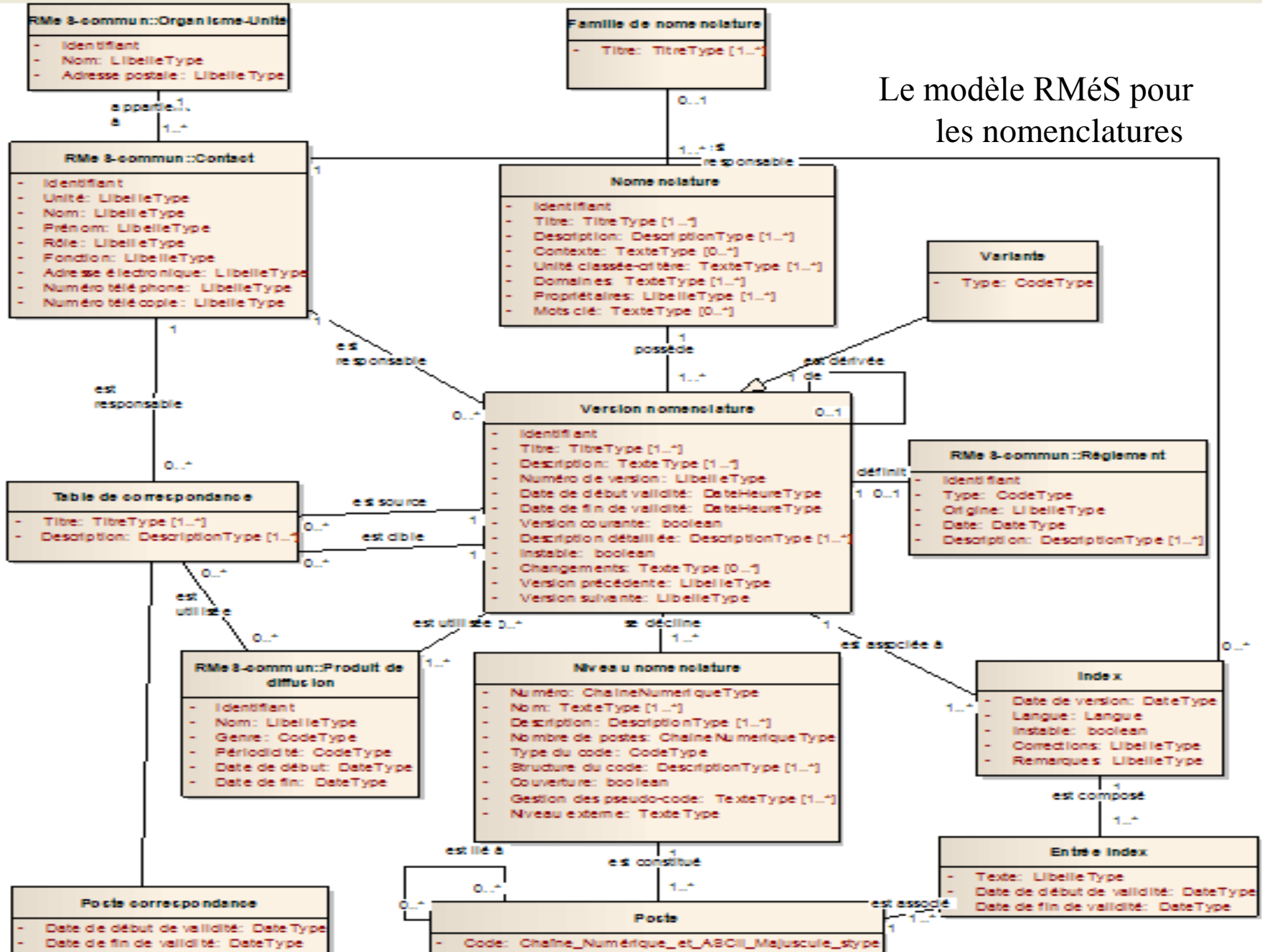


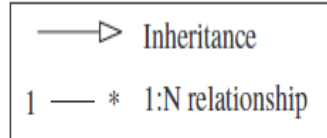
# Construire une nomenclature officielle

act Gérer les nomenclatures officielles, les tables de passage et les variantes

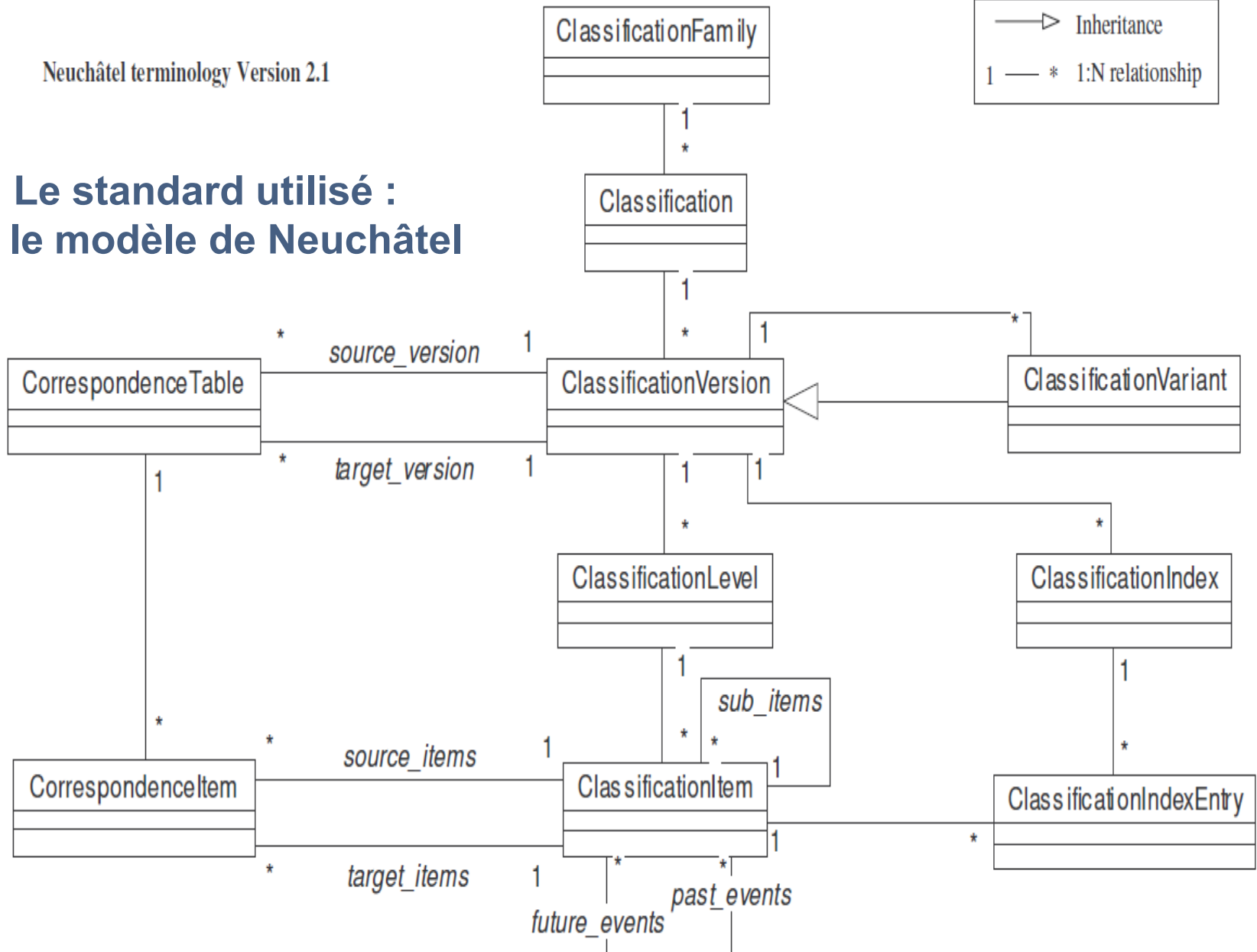


# Le modèle RMÉS pour les nomenclatures





## Le standard utilisé : le modèle de Neuchâtel





## Neuchâtel = modèle standard de nomenclatures

- › Modèle mis au point par les pays du Nord
- › Modèle adopté par la plupart des Instituts statistiques nationaux en Europe.
- › Modèle commun et définitions communes
- › ⇒ **Partage d'outils ou d'applications développés selon ce modèle**



## Les enquêtes : pourquoi d'abord les questionnaires ?

- › Plus de 100 enquêtes ou exploitations de fichiers administratifs
- › Questionnaires, questions, variables : un passage obligé
- › Usage répété pour une même opération statistique : enquête emploi trimestrielle
- › Usage commun à beaucoup de statisticiens à l'Insee et dans les Services Statistiques Ministériels
- › ⇒ terrain fertile pour la réutilisation de questions, de variables pour la même opération statistique d'une année sur l'autre mais aussi entre opérations statistiques
- › ⇒ terrain fertile pour l'utilisation d'un standard : DDI



## Intérêt de DDI comme standard : pas seulement un format d'échange mais un MODÈLE

- › L'information est structurée. Par exemple, le questionnaire est lié aux
- › questions :
- › blocs de questions



## Intérêt d'un standard comme DDI : pas seulement un format d'échange mais un MODÈLE

- › **Le cas de l'ESA** (enquête structurelle annuelle) auprès des entreprises
- › 128 questionnaires différents (selon le secteur, la taille, ...) utilisant 100 blocs de questions dont de nombreux communs aux 128 questionnaires
- › Chaque année, au lieu de retravailler dans word les 128 questionnaires, ce qui est très coûteux...
- › Modifications uniquement dans les blocs concernés et les 128 questionnaires sont générés automatiquement

*Merci de votre attention*

Marc Cohen-Solal  
Division Coordination de la  
méta-information  
4 Avril 2012

