

Distr.
GÉNÉRALE

CES/AC.71/2003/16 (Summary)
4 décembre 2002

FRANÇAIS
Original: ANGLAIS

**COMMISSION DE STATISTIQUE et
COMMISSION ÉCONOMIQUE POUR
L'EUROPE (ONU)**

COMMISSION EUROPÉENNE

**CONFÉRENCE DES STATISTICIENS
EUROPÉENS**

**OFFICE STATISTIQUE DES
COMMUNAUTÉS EUROPÉENNES
(EUROSTAT)**

**ORGANISATION DE COOPÉRATION ET
DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE)
DIRECTION DES STATISTIQUES**

**Réunion mixte CEE/Eurostat/OCDE sur la gestion des systèmes d'information statistique
(Genève, 17-19 février 2003)**

Point II: Impact des mesures et normes techniques sur la qualité des données

**MODÈLE ORIENTÉ OBJET DU CYCLE DE VIE DES DONNÉES
POUR LES SYSTÈMES D'INFORMATION STATISTIQUE**

Document d'appui

Présenté par Eurostat¹

Résumé

1. Les systèmes d'information statistique ont pour fonction de traiter les ensembles de données recueillis auprès des fournisseurs de données afin de produire des statistiques. Ces ensembles de données sont traités sous la forme de flux de données, dont la continuité est régie par les processus automatisés d'exécution des tâches. À chaque étape, les ensembles de données se trouvent dans un état particulier et peuvent être agrégés en ensembles de données plus grands ou plus abstraits. Le document présente un métamodèle orienté objet, qui décrit plusieurs aspects du cycle de vie des ensembles de données dans le contexte des systèmes d'information statistique. La notation UML est utilisée pour la présentation des modèles.

¹ Établi par Georges Pongas (georges.pongas@cec.eu.int) et François Vernadat (françois.vernadat@cec.eu.int).

2. Tout comme n'importe quel institut de statistique, Eurostat collecte auprès de nombreuses sources des données brutes (microdonnées ou données agrégées), qui sont ensuite validées et traitées en vue d'obtenir des informations statistiques à valeur ajoutée à l'usage des décideurs, des responsables politiques, des journalistes, des chercheurs ou des particuliers. Les données sont reçues et traitées sous forme d'ensembles de données, qui passent par plusieurs étapes et états et qui sont filtrées et agrégées avant de pouvoir être publiées en tant qu'information nouvelle utile. C'est ce que l'on appelle le «cycle de vie des données (CVD)» à Eurostat.

3. Le document fait état d'un métamodèle orienté objet reflétant divers aspects du cycle de vie des ensembles de données traités par les systèmes de production de statistiques. Les modèles sont conçus en se fondant sur l'expérience d'Eurostat mais ils ne sont en aucun cas propres à cet organisme de l'Union européenne et ils peuvent servir de métamodèles ou de modèles de référence à d'autres instituts de statistique. Les modèles sont présentés selon la notation UML.

4. Le document expose certains éléments de modélisation en vue de la définition d'«objets activité» (business objects) statistiques dans le domaine de la statistique officielle. Il s'agit d'un cadre conceptuel établi à partir de l'analyse de 12 grands systèmes de production de statistiques d'Eurostat. Ce modèle peut être utilisé pour affiner davantage la définition des «objets activité» statistiques et comme guide pour les phases de l'analyse des besoins dans le contexte du développement de systèmes statistiques.

5. Dans le cas particulier d'Eurostat, il importe d'utiliser un modèle uniforme qui précise l'état des ensembles de données, commandé par un modèle de métadonnées commun tel que défini dans le document, pour permettre aux systèmes extérieurs d'obtenir des informations sur le stade où se situe le cycle de vie des données. Étant donné que peu de travaux pratiques ont été réalisés à ce sujet jusqu'ici, il est encore possible de proposer une architecture logicielle sur laquelle s'appuyait ce genre de modèle de métadonnées commun pour le cycle de vie aux fins des systèmes de production de statistiques d'Eurostat. En tout état de cause, le système assurant le suivi du cycle de vie des données doit être alimenté par les systèmes de production. Les informations sur le cycle de vie communiquées par les systèmes de production doivent être indépendantes de leurs formats de données internes.

6. Les travaux futurs d'Eurostat consisteront à mettre au point un gestionnaire opérationnel du CVD basé sur le métamodèle exposé dans le document en vue d'enregistrer et de contrôler l'évolution des flux de données au sein d'Eurostat.
