New data curation impacting the renewal of Basecol

Y.A. Ba, M.L. Dubernet, C.M. Zwölf, N. Moreau

ASOV Meeting 15-17 Mars 2017

Plan

- Introduction

- Reasons of redesign

- Procedure

  - Quality Check of data

  - History integration & improvement of the database

  - Mechanisms used to introduce new data

- Conclusion

Introduction

- Basecol is a database devoted to collisional data of atomic and molecular physics

- It is labeled AA-ANO5

- It is supervised by an international working group of molecular physicists and astrophysicists

- It is a web service into VAMDC (Virtual Atomic and Molecular Data Center)
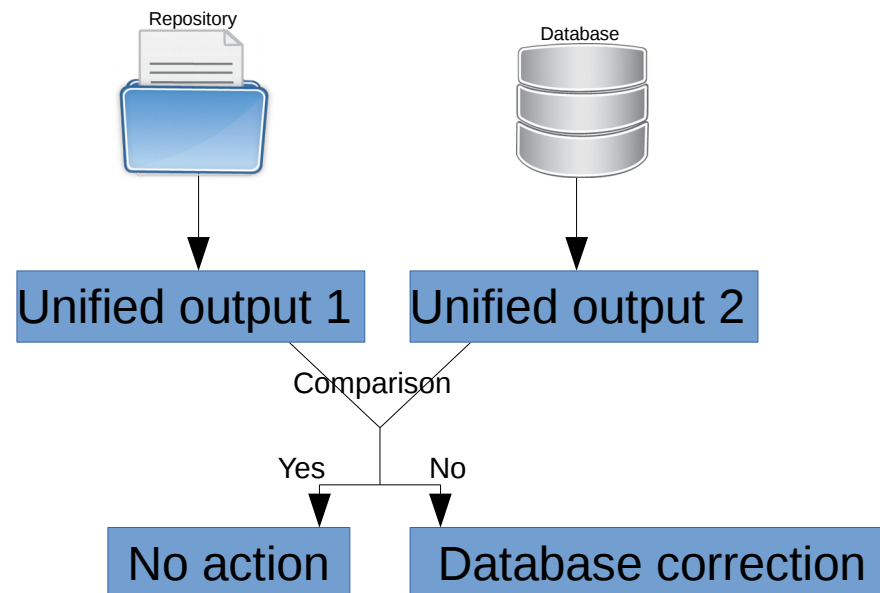
Reasons of redesign

- Basecol exists since 2004

- Its main infrastructure has become obsolete

- System maintenance is difficult

Hence our initiative to build a new infrastructure

Procedure : quality check of data

- Compile the repository → replication and preservation of data

- Comparison with the database

- Correct the database → ensure the data quality

Repository

Database

Unified output 1   Unified output 2

Comparison

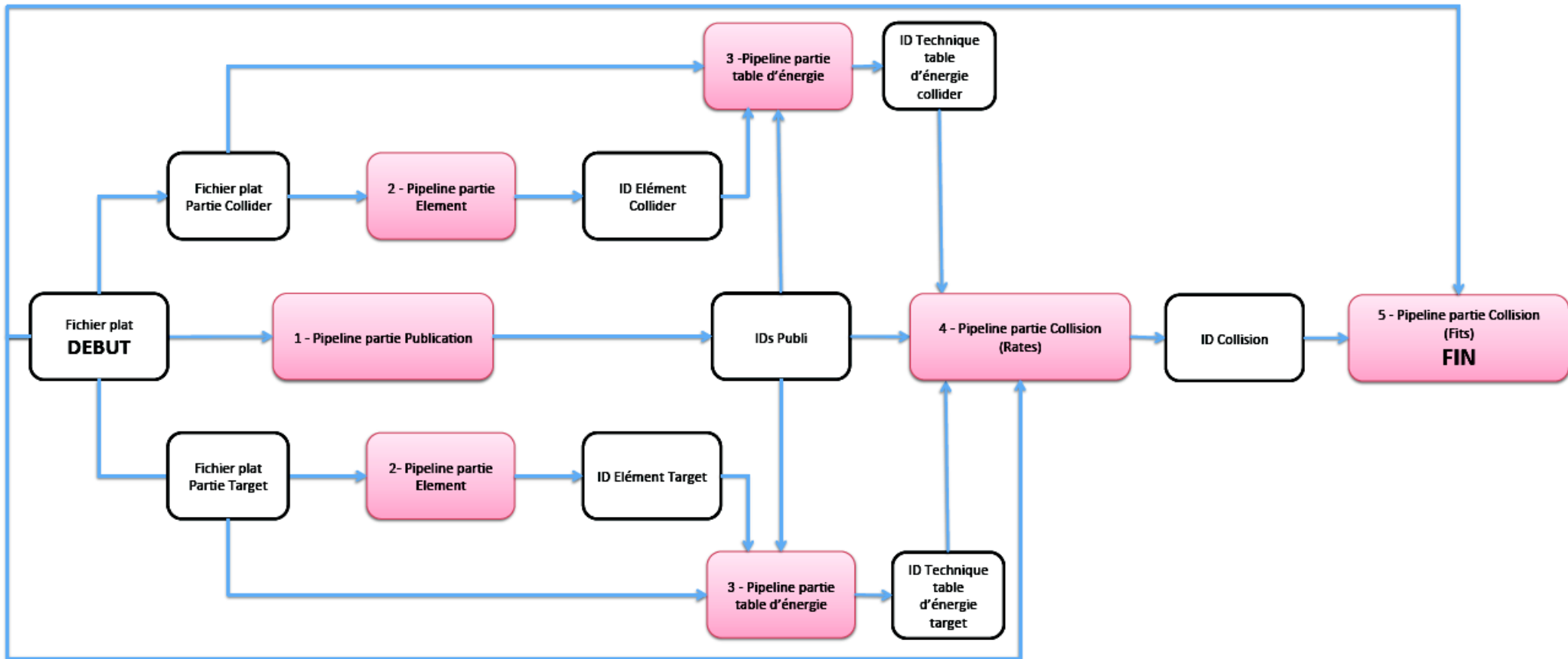Yes   No

No action   Database correction

Procedure : history integration & improvement of database

- Inclusion of the energy origin

- Inclusion of timestamp → data versioning

- Keep old versions → ensure the data recovery

- Documentation of the changes → traceability

Procedure : Mechanisms used to introduce new data

- Interactive software to create and modify data into Basecol database :

  - Implemented in Java 1.7, available on git with documentation

  - Input → formated file for each type of data (species, publications, energy tables, transitions and fits files)

  - For each type, we have a process to handle the required modification or datatype creation

  - output → modified database

Procedure : Mechanisms used to introduce new data

Procedure : Mechanisms used to introduce new data

- Specificity of the software :
  - Check that files are well formated
  - Check if some data match with the description of metadata
  - Possibility to cancel the execution without any changes

Conclusion

- We started this work since september 2015

- Our efforts regroup all services linked to Basecol (web services into VAMDC, website...)

- Today with the new Basecol we guarantee :

    – Quality of data

    – Good versioning

    – Traceability

    – Easy maintenance