

Achieving Interoperability and Reusability of Metadata: Two Organisations and the Same Reference

- ❑ *National Institute of Statistics and Economic Studies (INSEE)*
- ❑ *Secure Data Access Center (CASD)*



Halima Bakia, Ifaliana Rakotoarisoa and Thomas Dubois



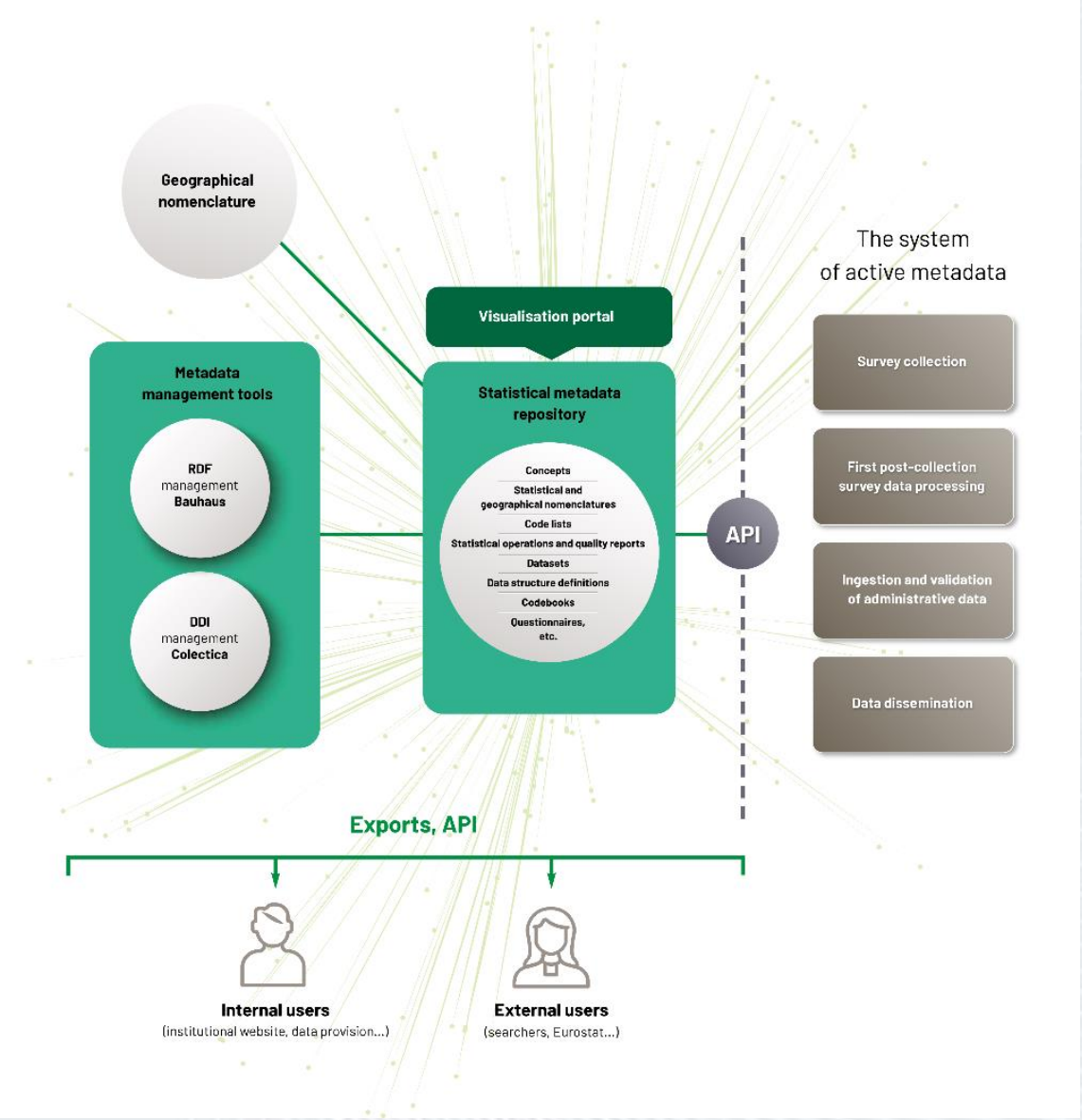
Outline

- Presentation of the two organizations
 - INSEE and the RMÉS Metadata Repository
 - CASD French secure data center
- Current situation
 - Previous Works Between INSEE and CASD
 - Current Process for Disseminating Data and Metadata
- Experimentation: an Opportunity Offered by a Metadata Standard (DDI)
 - Experimented Workflow and Beyond
 - Expected Gains
- Conclusion & Next steps



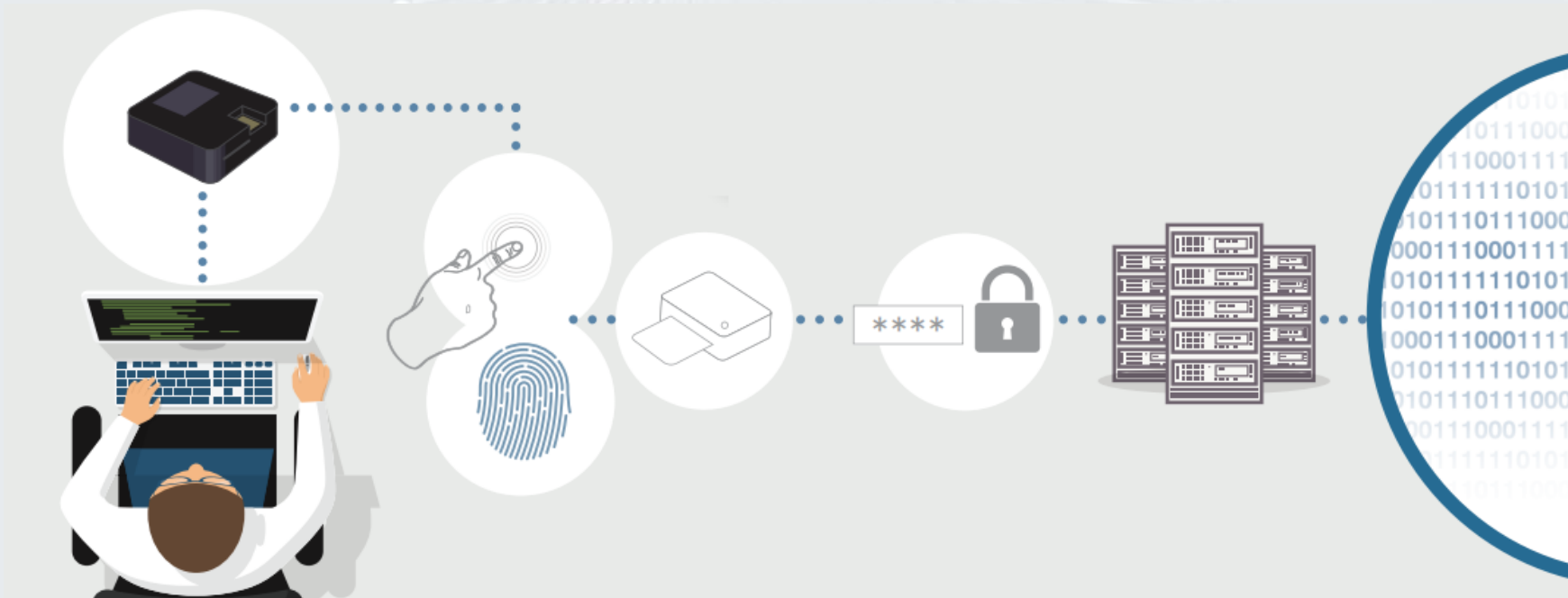
Two Organizations:
a Data Producer (INSEE)
and a Data Disseminator (CASD)

The RMÉS Metadata Repository



The French Secure Data Access Center (CASD)

- Organize and set up services of secure access to confidential data for users pursuing non-profitable research, study, evaluation or innovation purposes
- Secure terminal called an SD- Box
- Individual access smartcard and biometric authentication



Key Figures

513

data sources

made available in a secure way (see the complete list of available sources)

1503

managed and hosted projects

997

institutions

5257

secure data users since launch

+400

referenced publications and communications

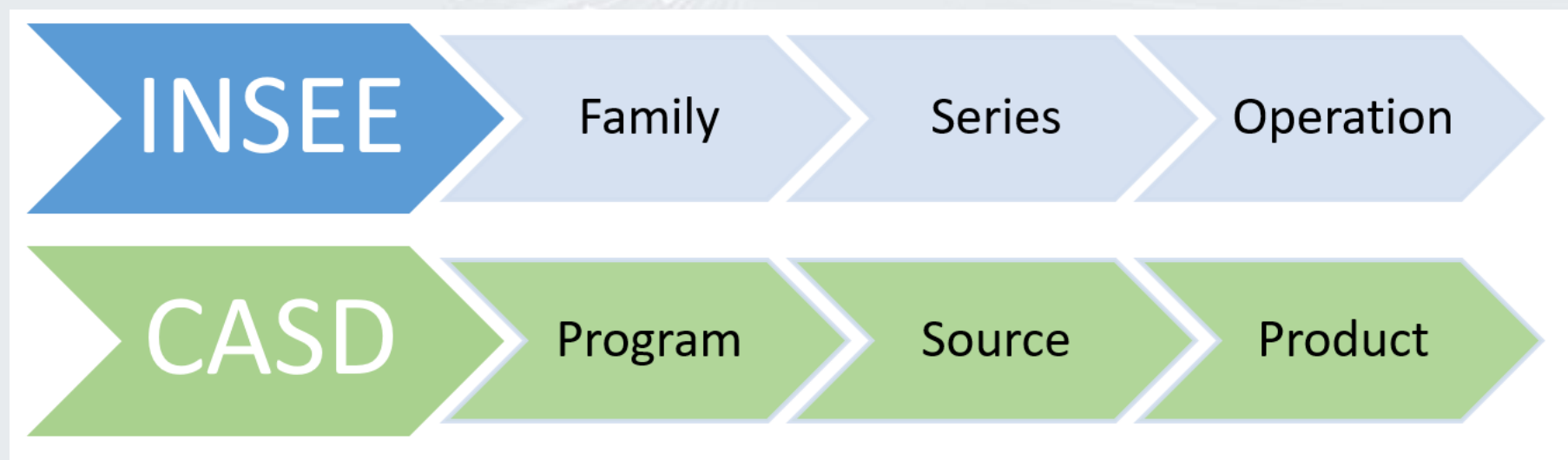
(articles, book chapters, books, theses, reports, conference papers, etc.)

➤ Documentation

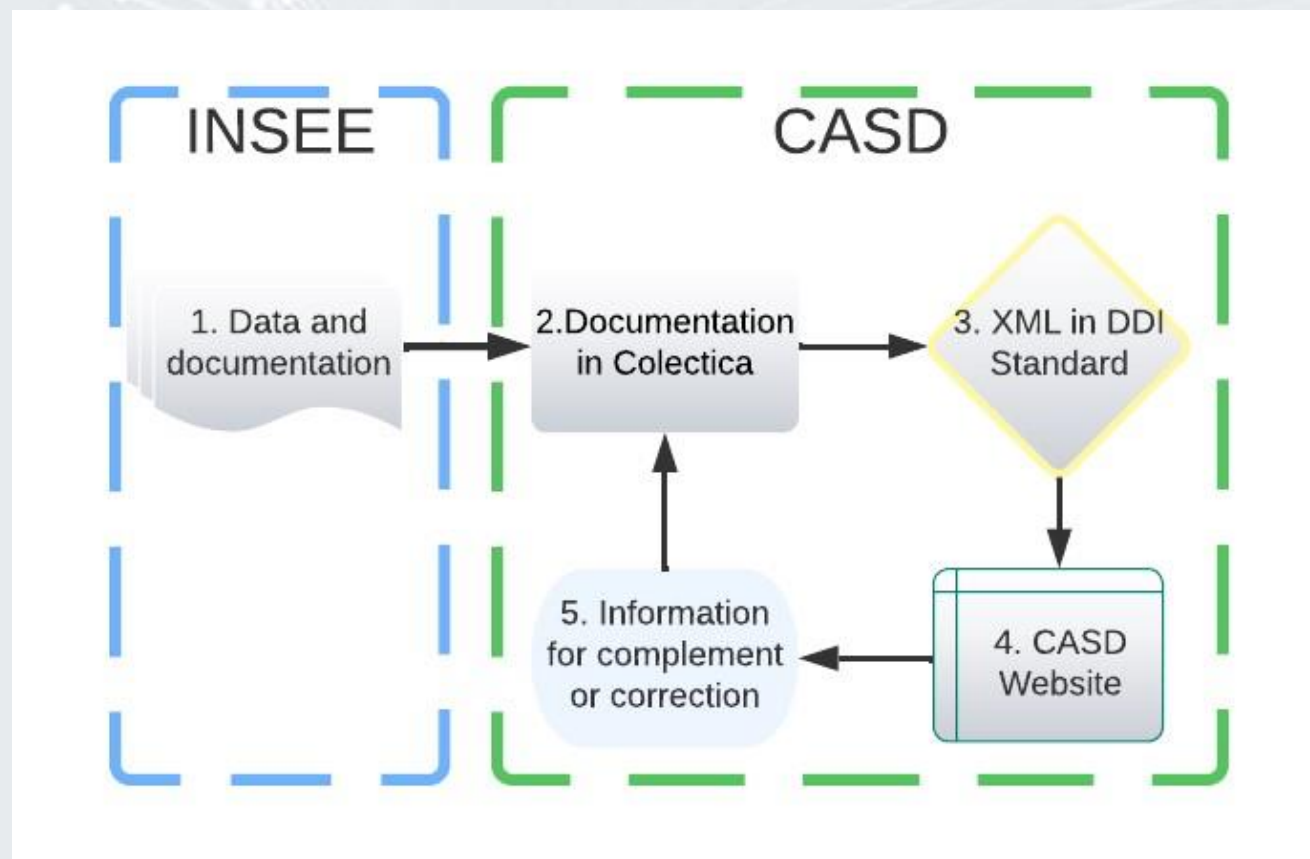
- DDI (Data Documentation Initiative) standard
- Colectica Software
- Access to data sources <https://www.casd.eu/en/data-used-at-casd/>

Previous Works Between INSEE and CASD

- Mapping Between INSEE and CASD Model



Current Process for Disseminating Data and Metadata

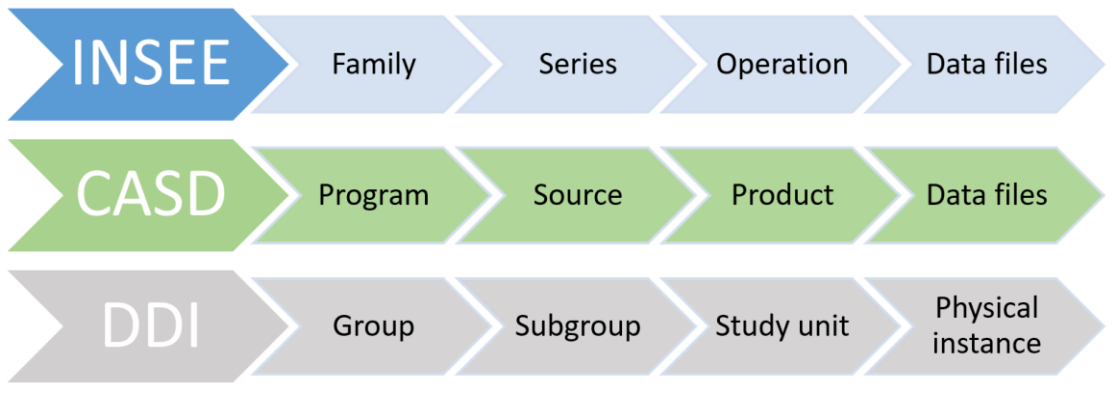
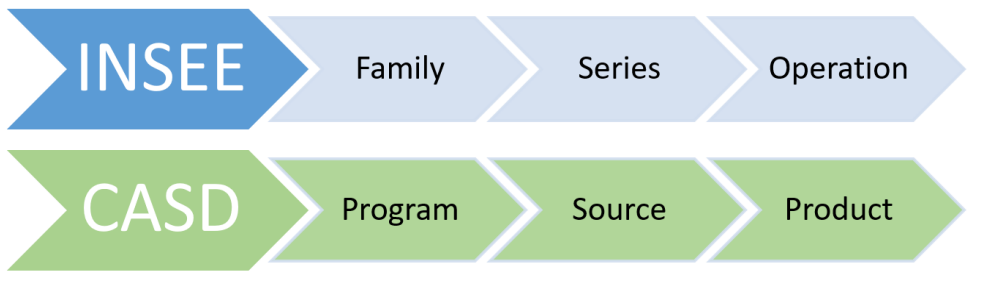




Experimentation

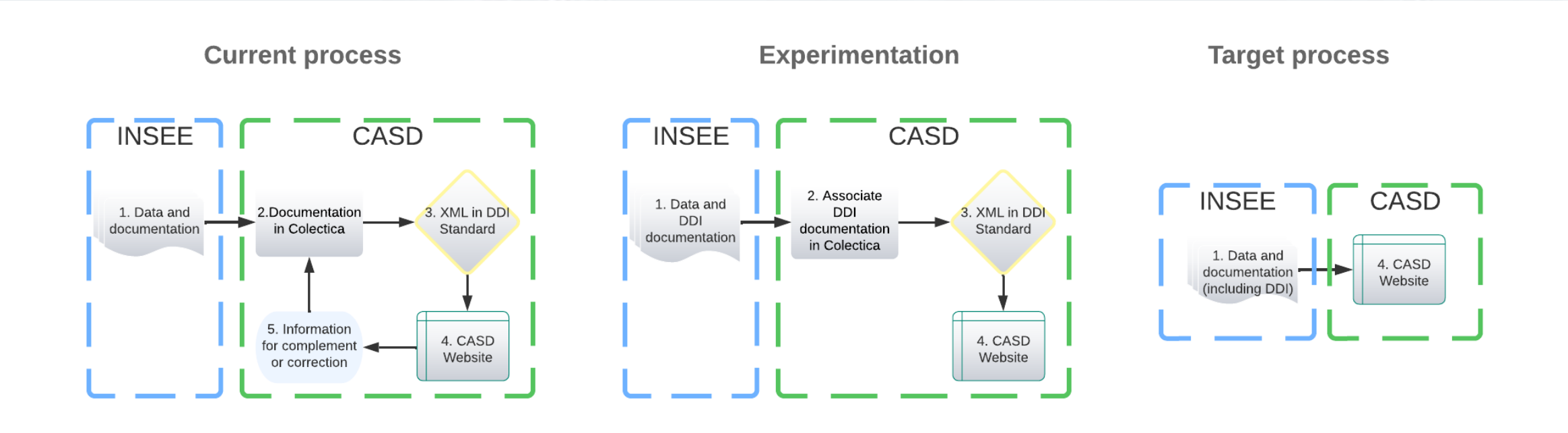
Mapping Between INSEE and CASD Model

- Matching the documentation's architecture of INSEE and CASD with DDI standard enables the exchange of XML files
- Experimentation focused on metadata at the lowest level: Physical instance



Process of Disseminating Data and Metadata

- CASD website requires documentation at study unit level: manipulation via the Colectica software
- Experimentation on “All employees databases - job position data 2019”



Expected Gains

- on workload
- on timeliness
- on reliability
- on relevance
- on accessibility and clarity of the information
- on coherence



Conclusion

Conclusion

- Main purpose:
 - Provide users with an overview of the confidential data available in the secure area
 - Description of the variables and modalities
- The benefits of sharing the same standard:
 - Facilitates reusability
 - Speeds up documentation display
 - Reduction of the reinterpretation of the meaning of variables
- Next steps
 - Consolidate this process and encourage its generalization within the other producers from the French Official Statistical System
 - At this point in time none of the other producers providing the data through CASD use the DDI standard

Thank you for your attention

CASD website: <https://www.casd.eu>

CASD

