EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Optimization of Accessibility and Quality of Metadata for Researchers: an Example of Collaboration Between INSEE and CASD

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

Halima Bakia, Ifaliana Rakotoarisoa and Thomas Dubois

Creative Commons Attribution 4.0 International (CC BY 4.0)

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL

Outline

- Presentation of the two organizations
  - INSEE and the RMéS Metadata Repository
  - CASD French secure data hub
  - Previous Works Between INSEE and CASD
- Experimentation
  - Mapping Between INSEE and CASD Model with DDI
  - Current Process for Disseminating Data and Metadata
  - Experimented Workflow and Beyond
  - Expected Gains
- Conclusion & Next step

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

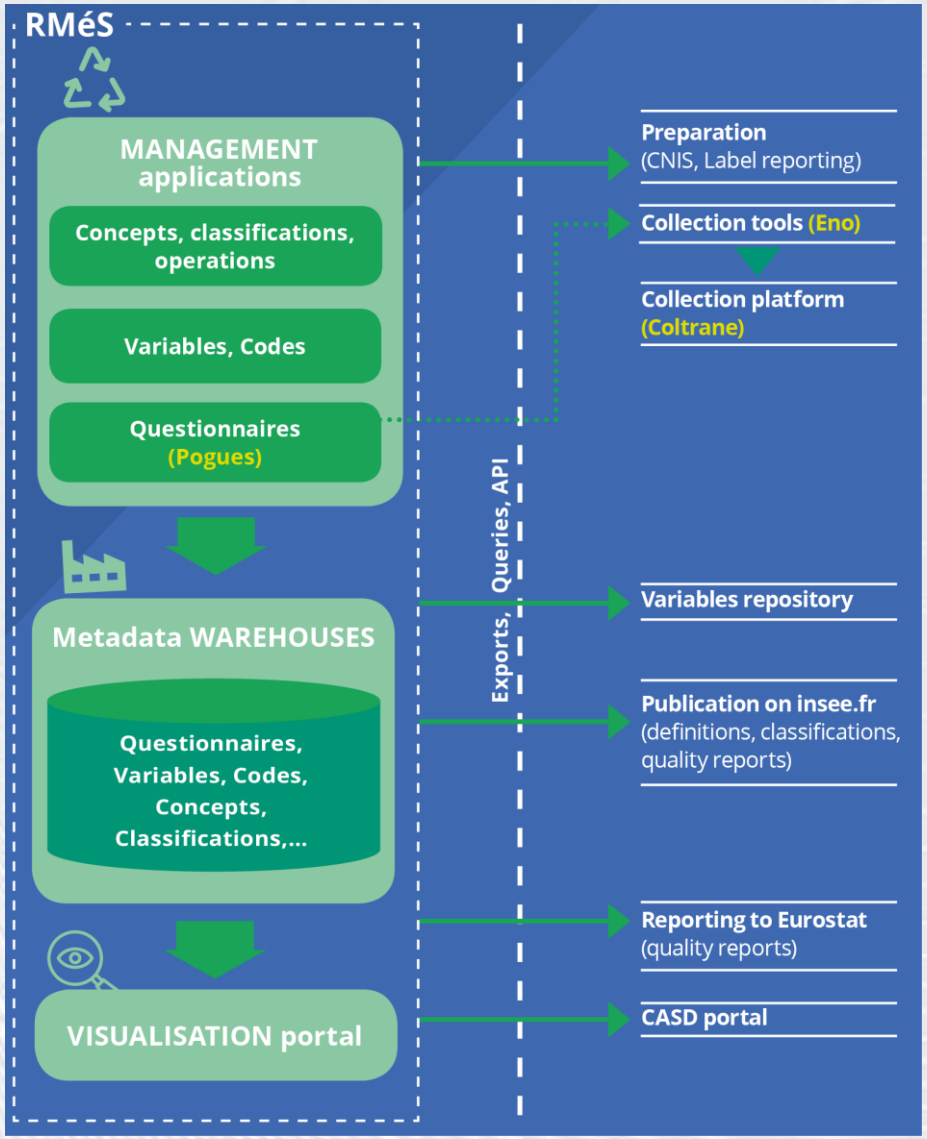
The conference is partly financed by the European Union

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

CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES

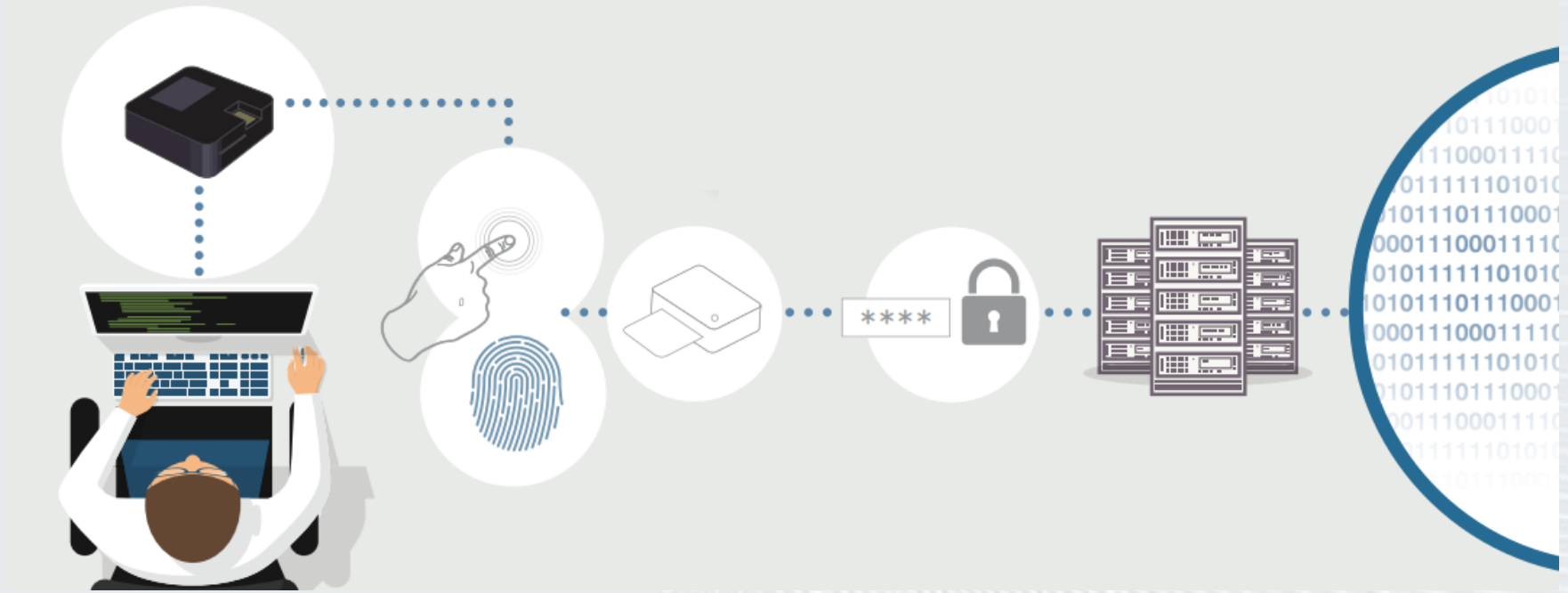
The RMéS Metadata Repository

Statistical metadata are essential to understanding the statistical output that they describe. By standardising metadata, they can be accessed and reused easily. The producer of a single statistical operation need not look any further. But for an organisation such as INSEE, which handles a large number of statistical operations, centralisation is necessary. It ensures uniformity in the representation of cross-functional metadata and the consistency of information provided, using identification rules.

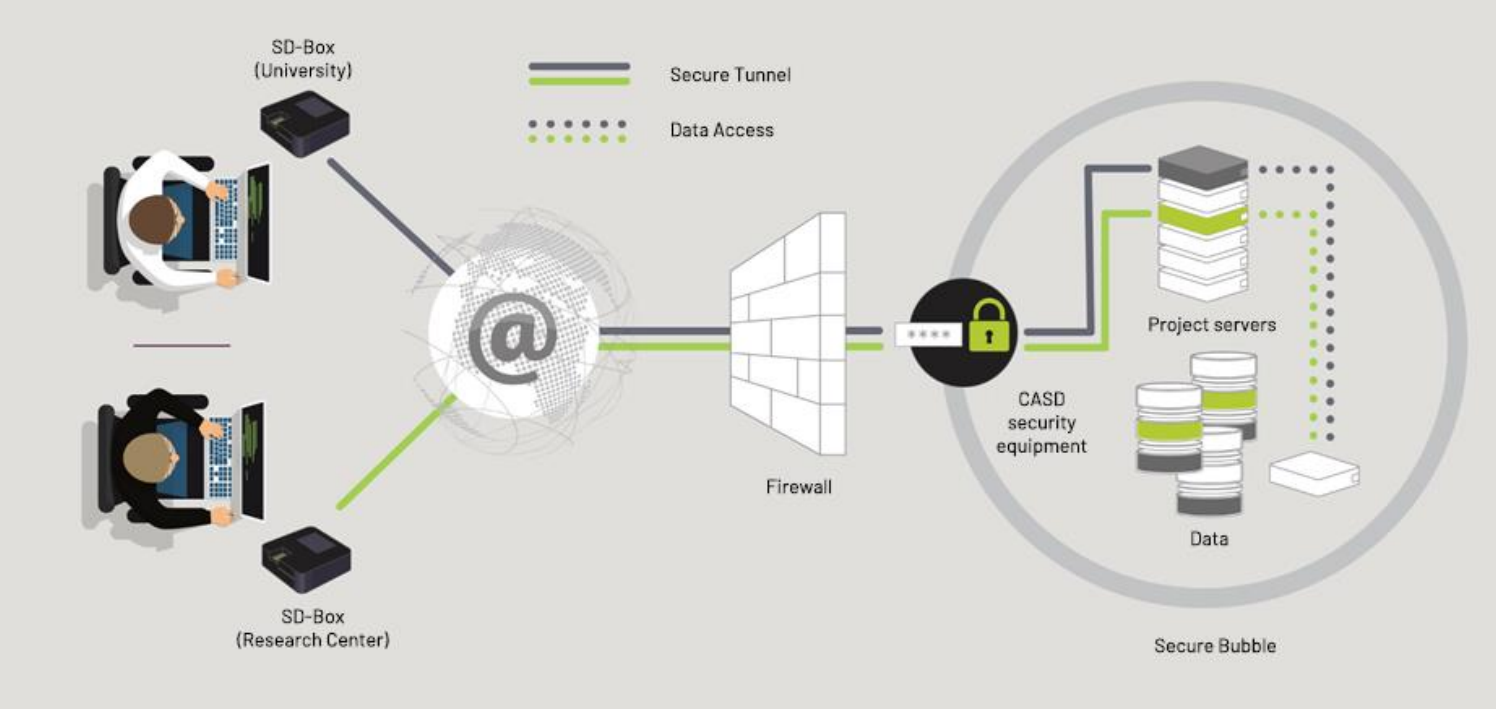


The French Secure Data Access Center (CASD)

- 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



- The SDBOX is linked by a secure tunnel to the user's project work environment
- Each secure bubble is isolated from all the other bubbles
- No file can be retrieved directly by the user or inserted in the work environment (no copy/paste, no printing, no USB key, no internet connection...)



➢ Key Figures

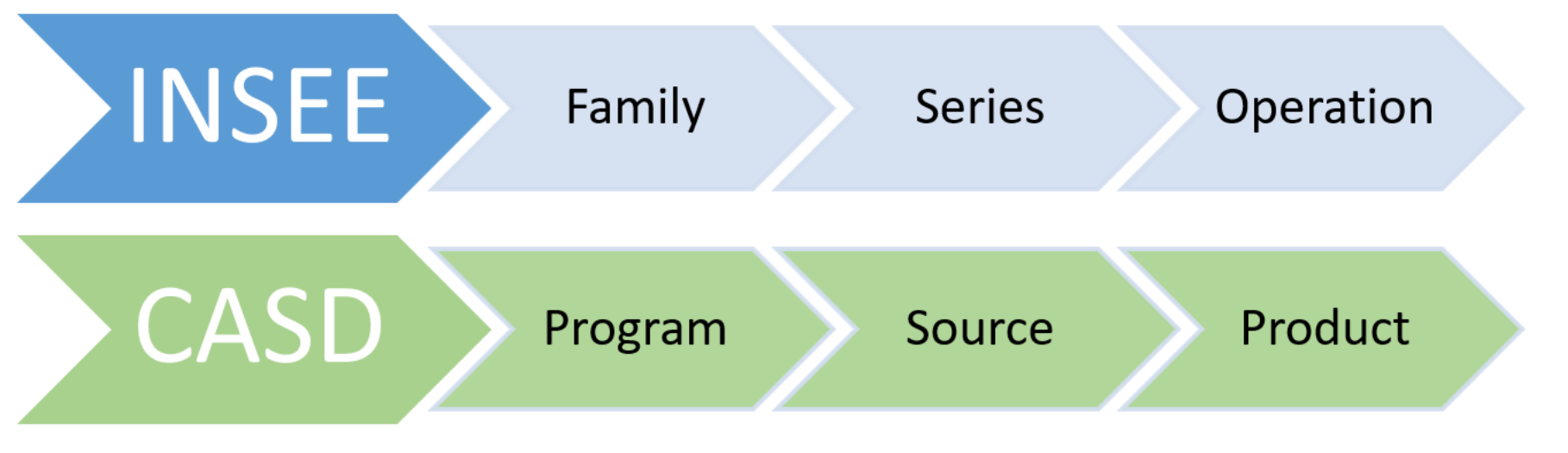
518	1542	1010	5460	+400
sources de données mises à disposition de façon sécurisée (voir la liste complète des sources disponibles)	**projets gérés et hébergés**	**institutions utilisatrices**	**utilisateurs de données sécurisées** depuis son lancement	**publications et communications référencées** (articles, chapitres d'ouvrages, ouvrages, thèses, rapports, conference papers, etc.)

➢Documentation

- Data Documentation Initiative standard (DDI Lifecycle)
- 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

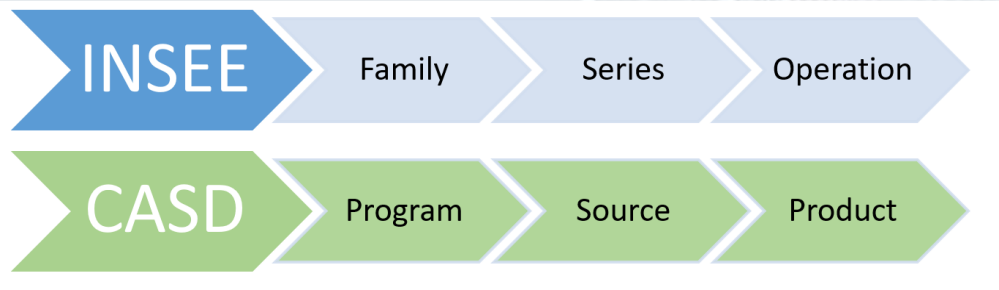


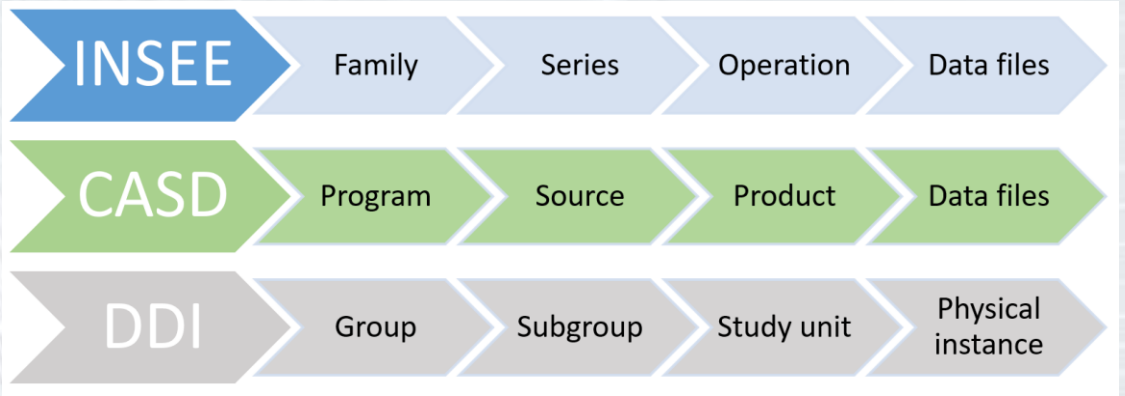
Experimentation

CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES

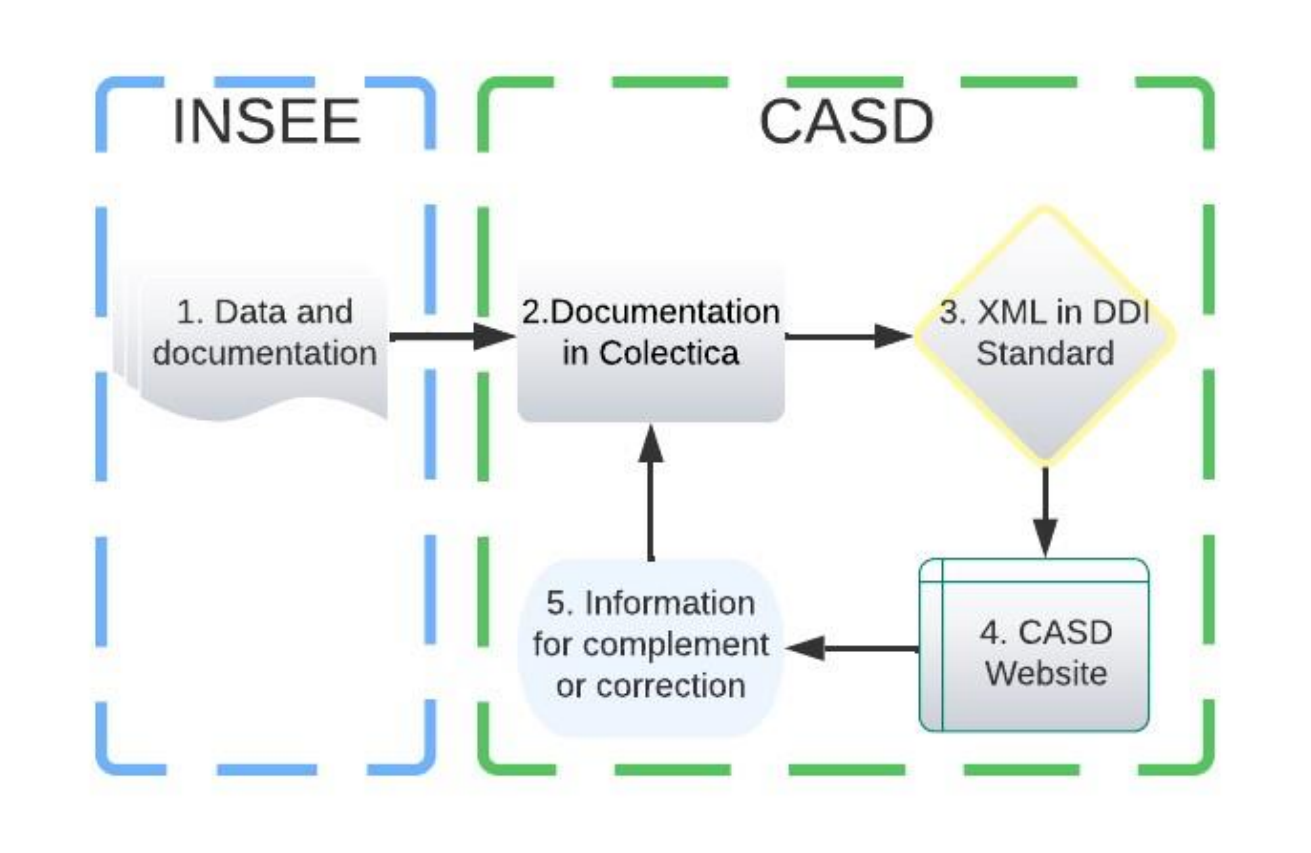
Mapping Between INSEE and CASD Model with DDI

- 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



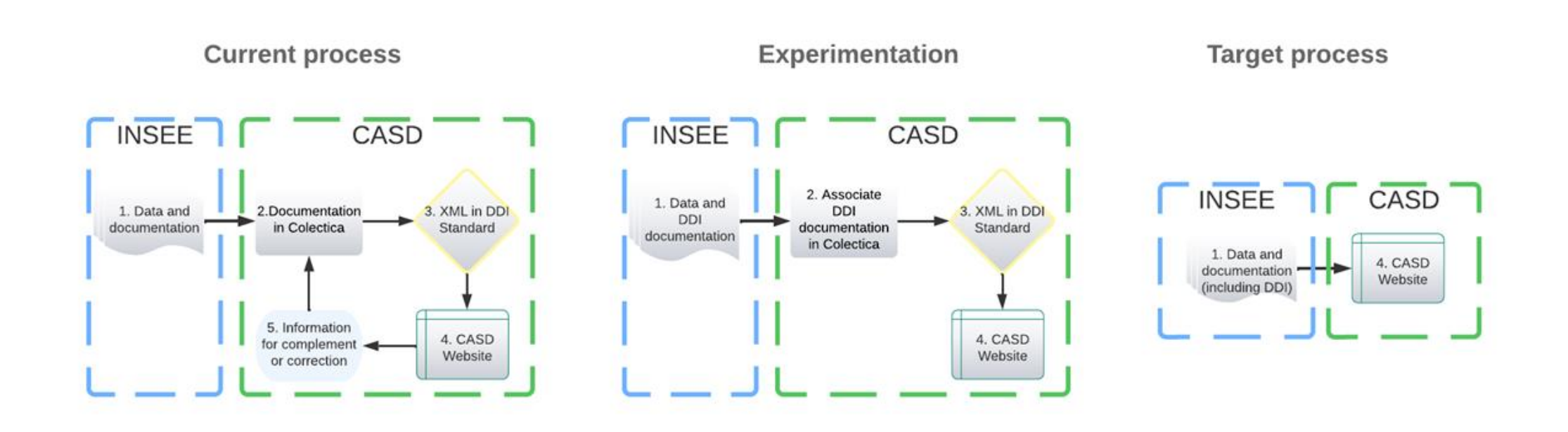


Current Process for Disseminating Data and Metadata



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

What's next?

- Ongoing work
  - INSEE has retrieved 163 metadata files in XML format
  - Re-use to back-document their data files in the RMéS repository
- 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

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL

Thank you for your attention

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

CASD

CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union