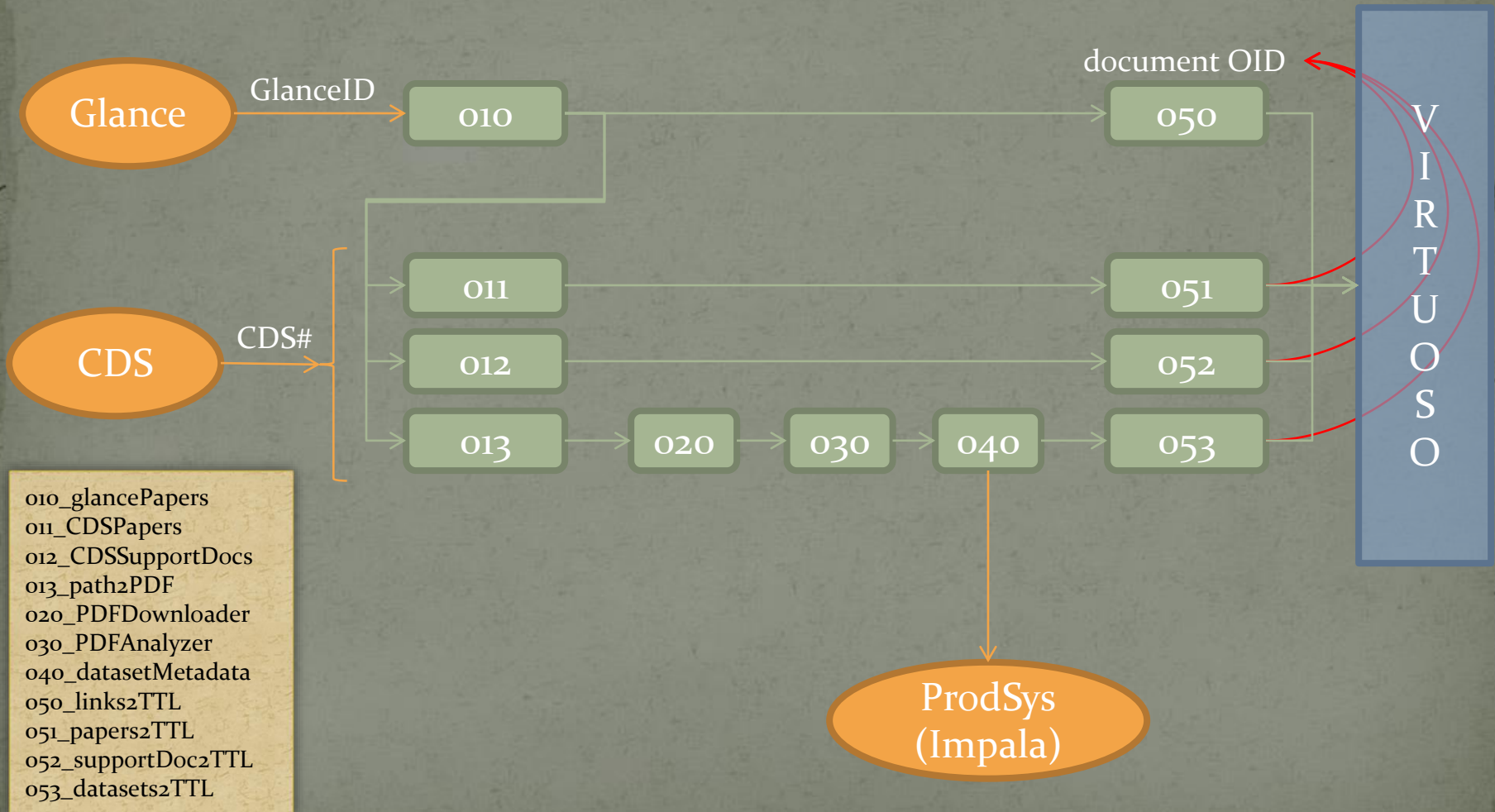


# DKB Dataflow modification

---

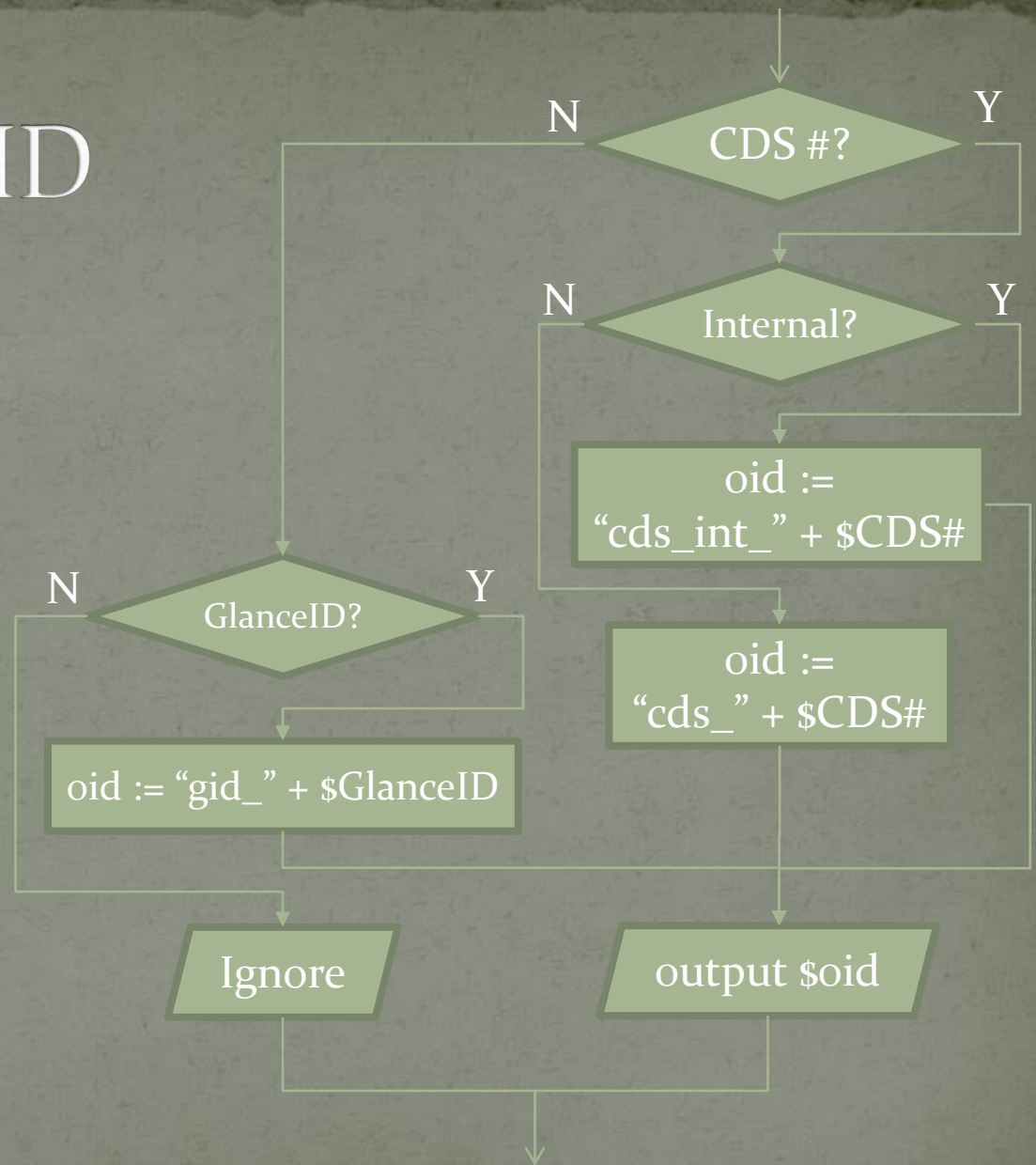
DocumentContent introduction

# How it works now

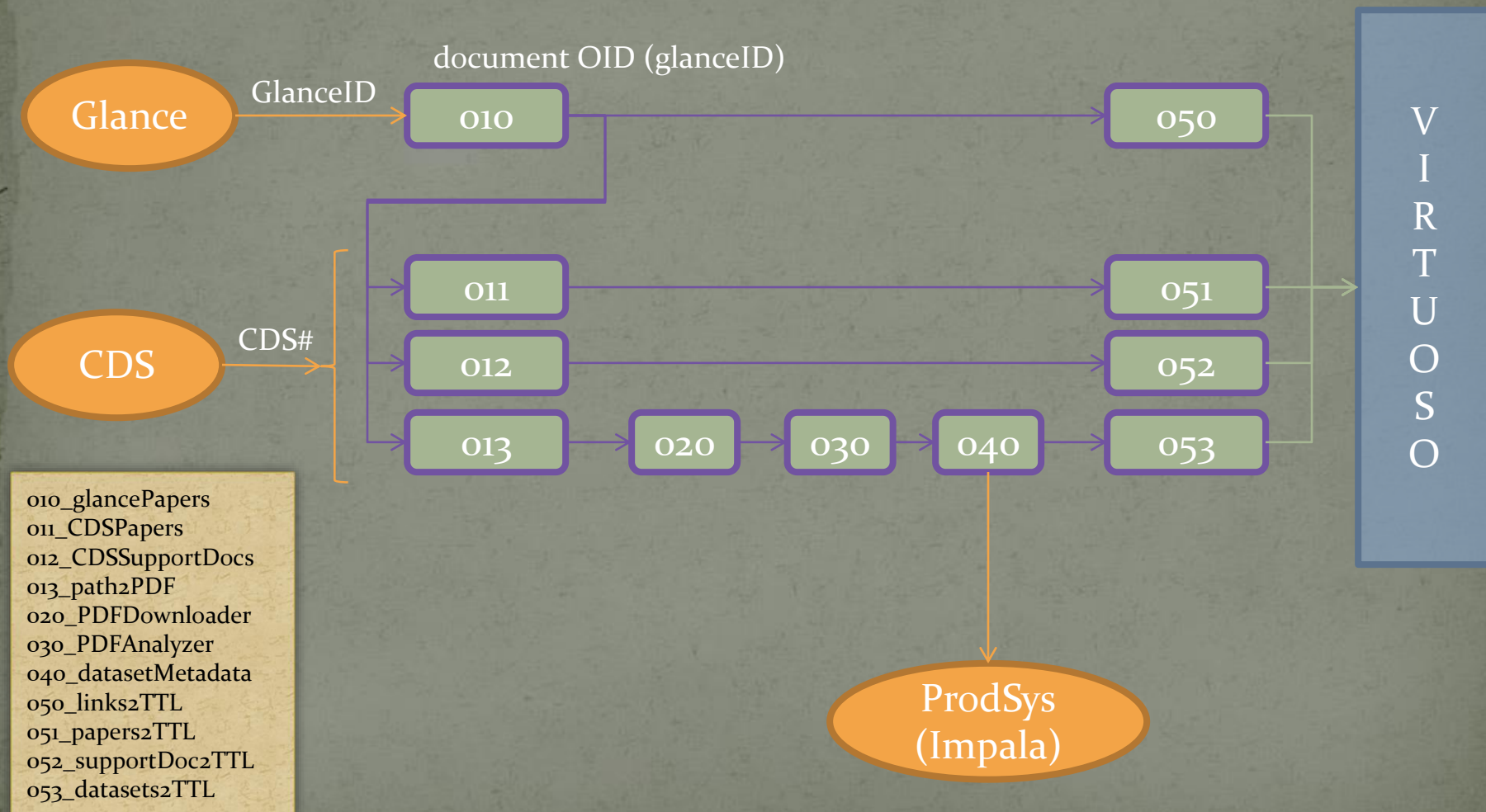


# Document OID

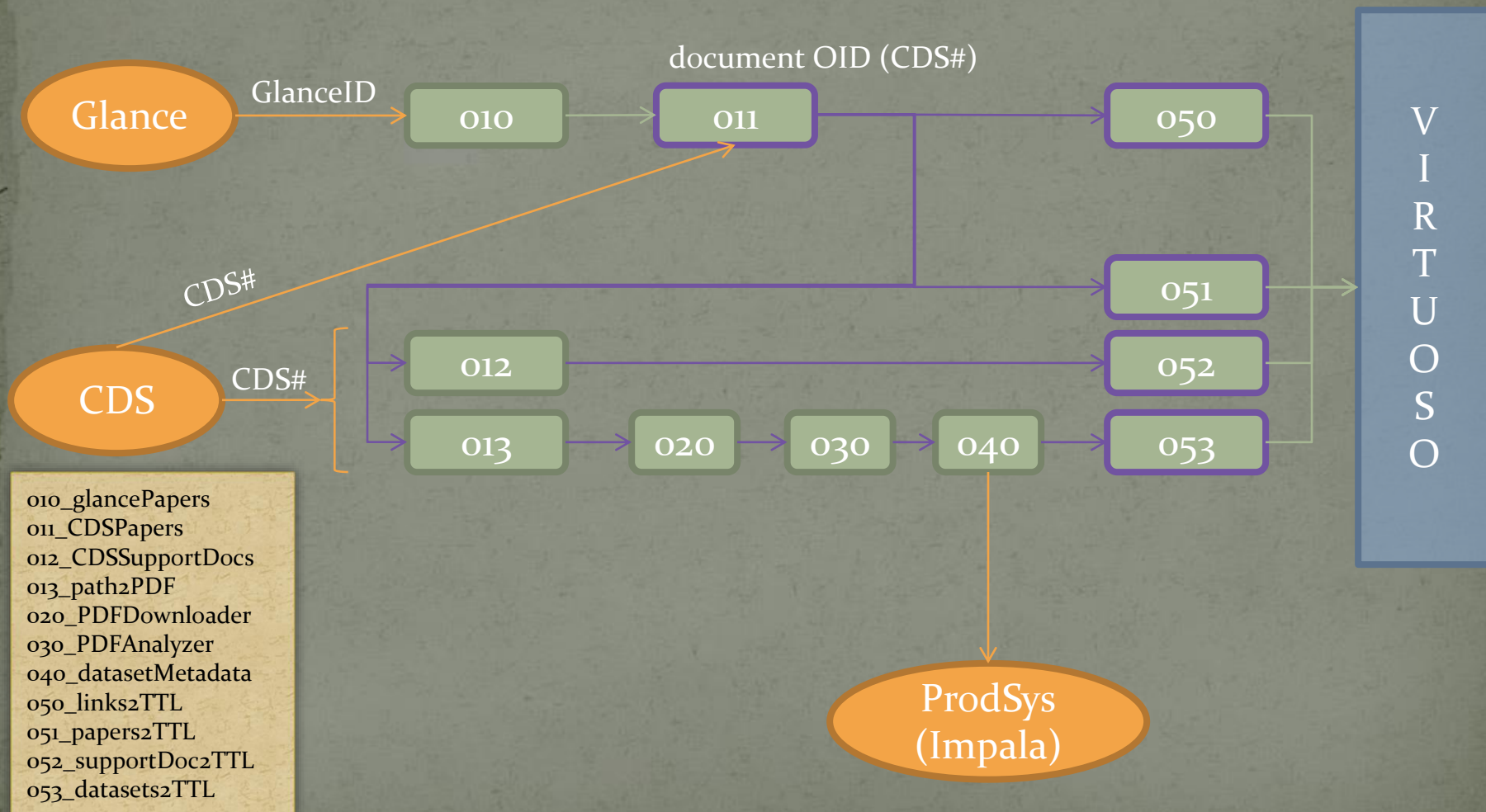
- Now: UUID
- V1: Glance ID
- V2: CDS #
- V3: priority choice



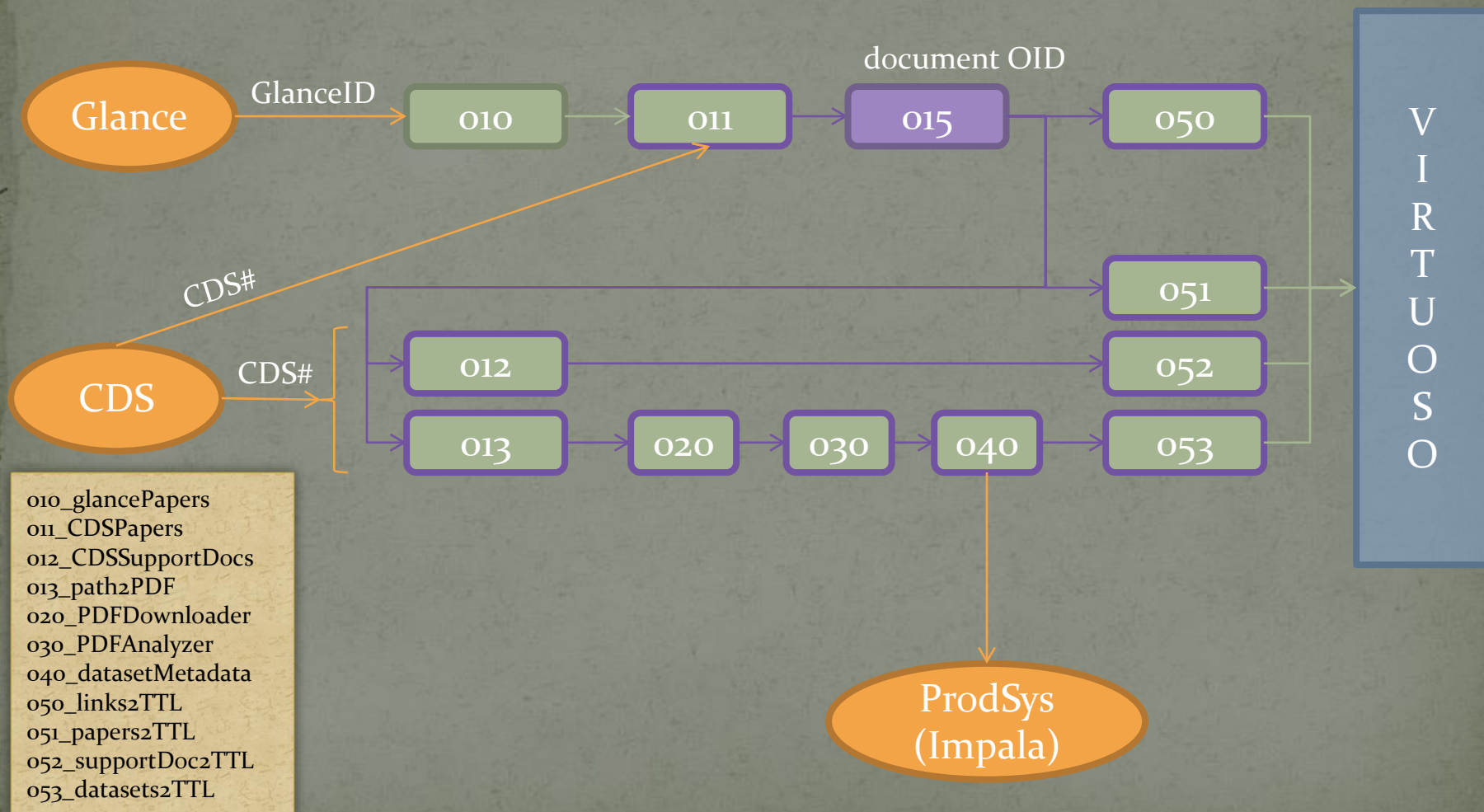
# Modification 1: DocumentOID (v.1)



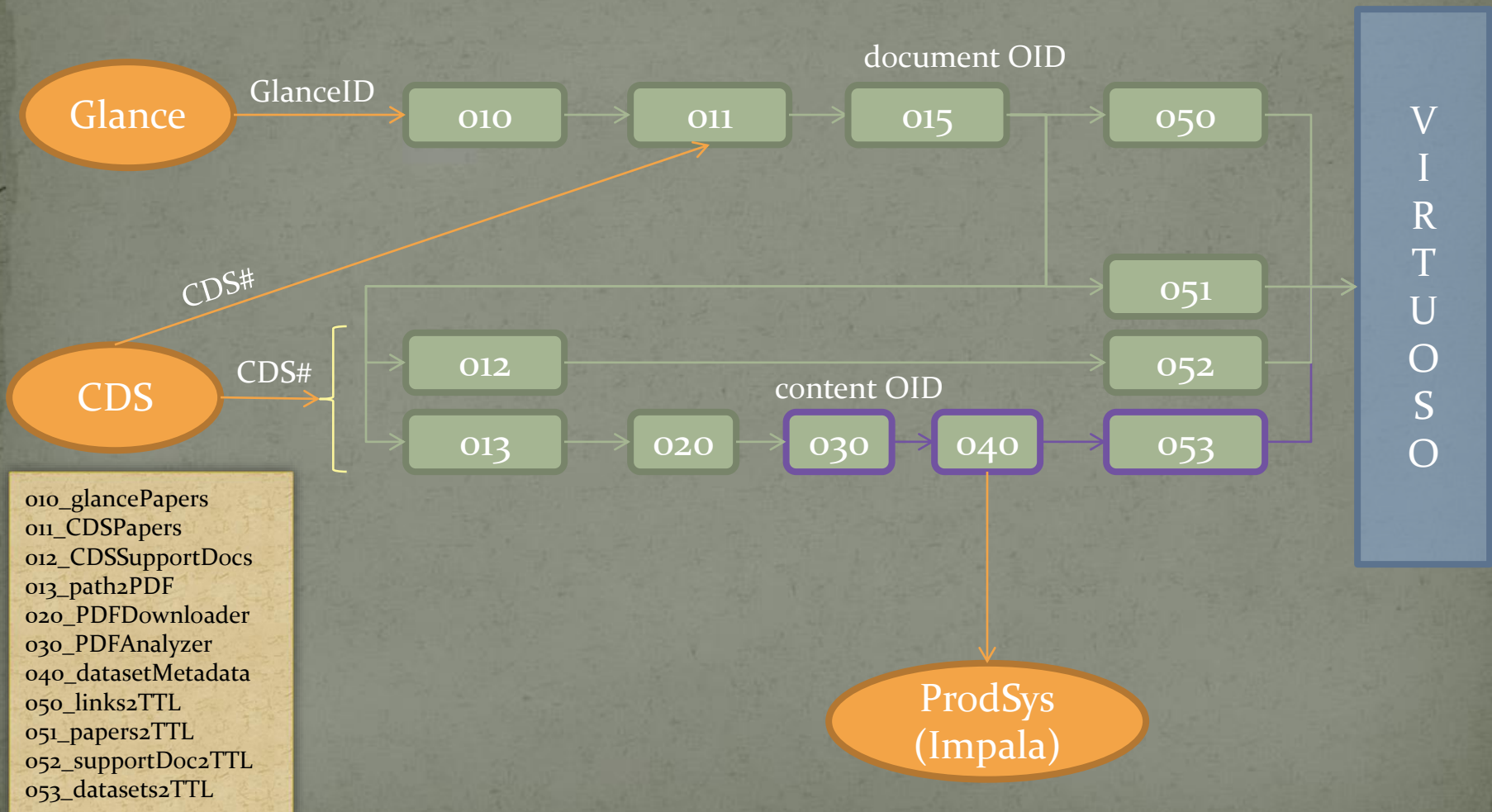
# Modification 1: DocumentOID (v.2)



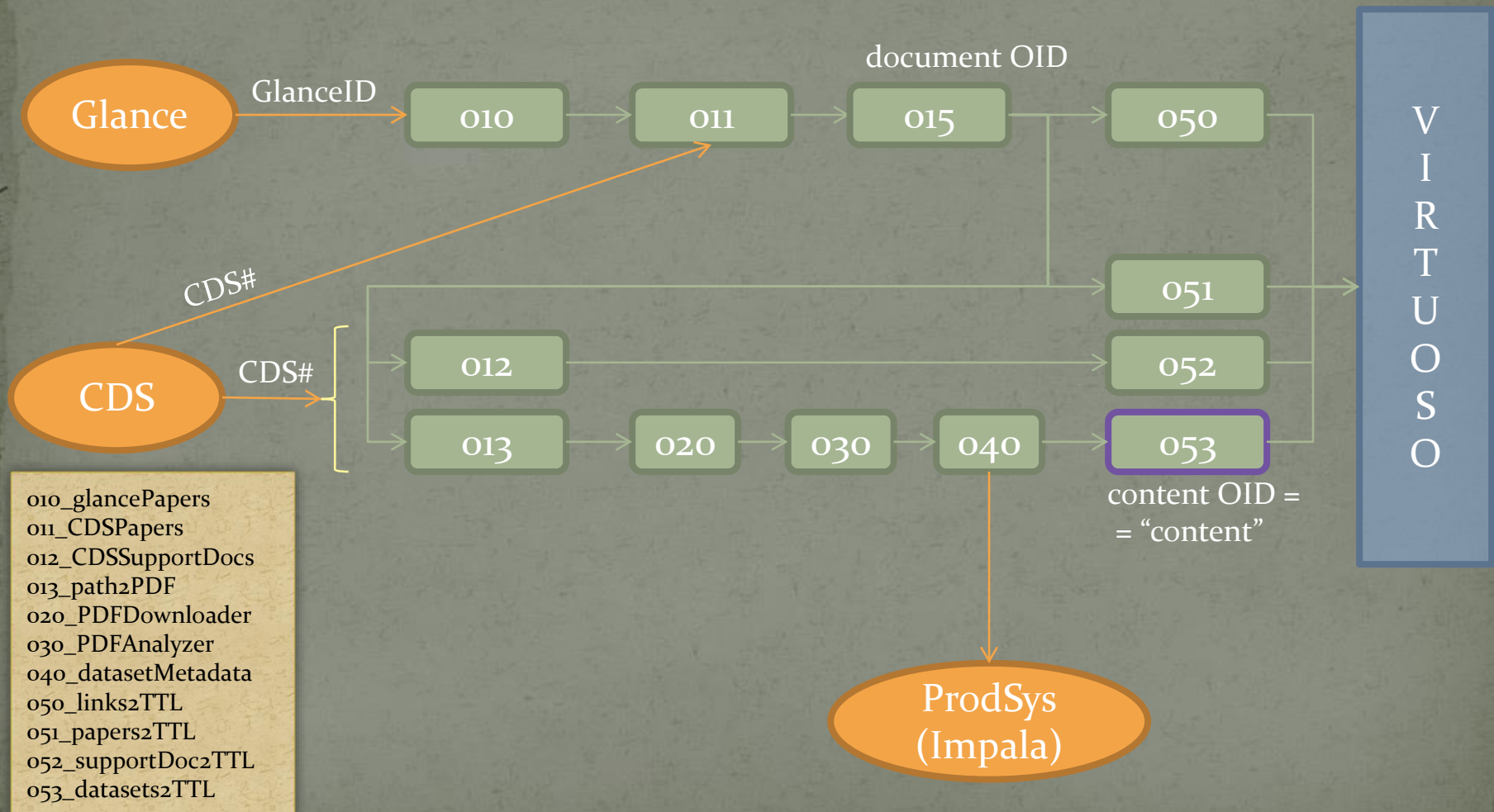
# Modification 1: DocumentOID (v.3)



# Modification 2: DocumentContent



# Modification 2: DocumentContent (lazy mode)





# DocumentContent: discussion

- Lazy mode:
  - OID = “content”
  - content = all metadata found
- Normal mode:
  - OID to be generated at document analysis stages
  - contents (v.1):
    - datasets (real) – list of datasets with real data
    - datasets (mc) – list of datasets with MC generation data
    - table X – information extracted from table X
    - text – other information (extracted from the plain text)
  - contents (v.2):
    - logical pieces of the paper (chapters, sections, tables, ...)

*The End*