

# Data Conservancy Packaging Tools

10<sup>th</sup> International Digital Curation Conference  
Tim DiLauro & Jonathan Petters



**Data**Conservancy

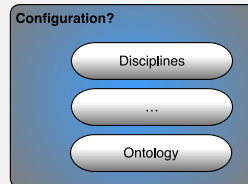
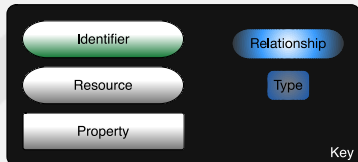
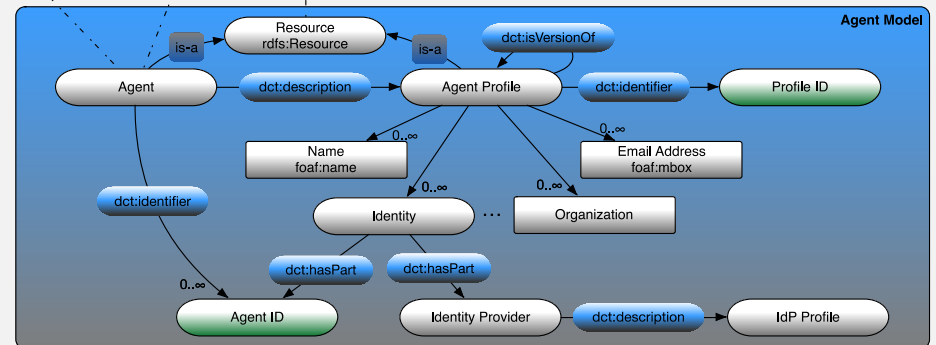
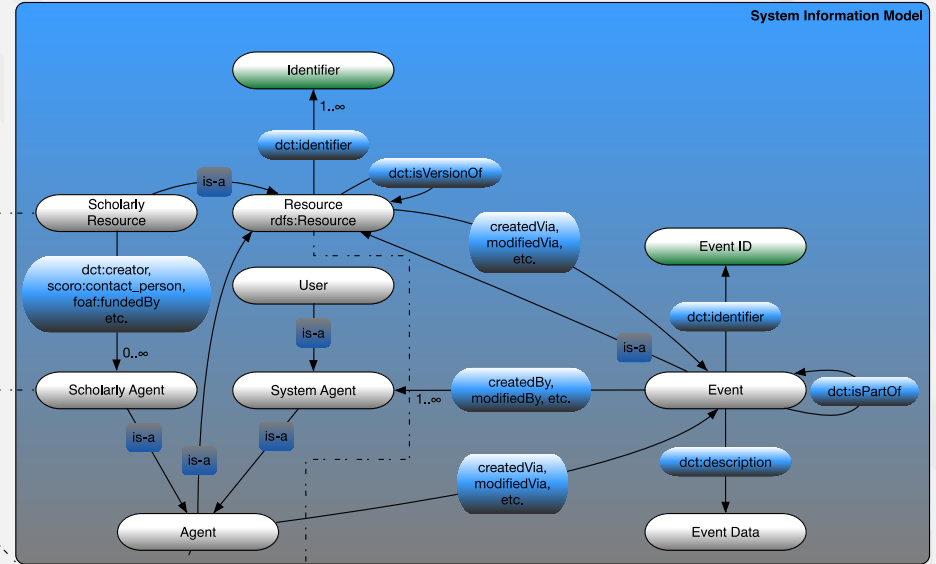
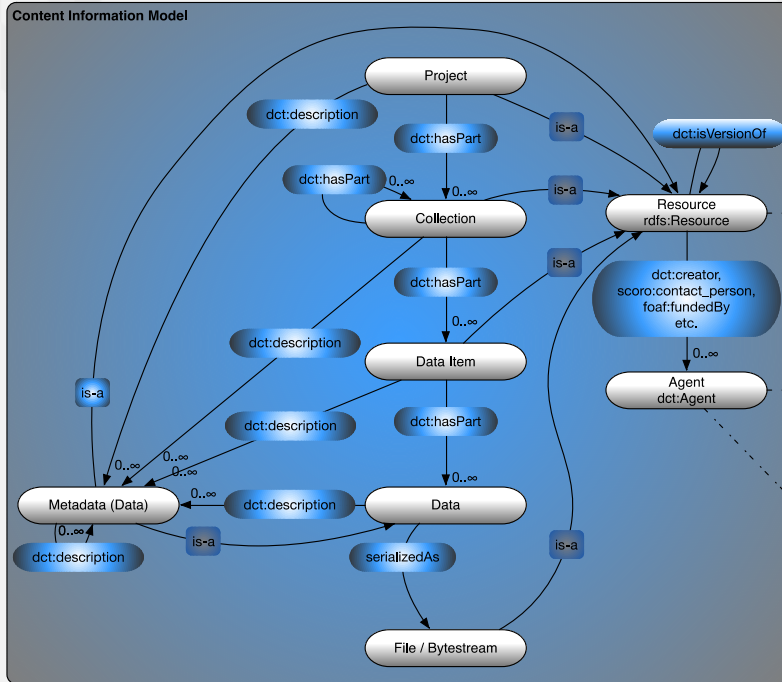


# Goals & Needs

- Capture relationships for packaged entities
  - Make transfer easier
  - Ensure successful transfer to archive
  - Support content reorg while ensuring integrity of original content
  - Support more than one package model and serialization
- 
- Content business model
    - - internal package description model
    - - BagIt+ORE serialization

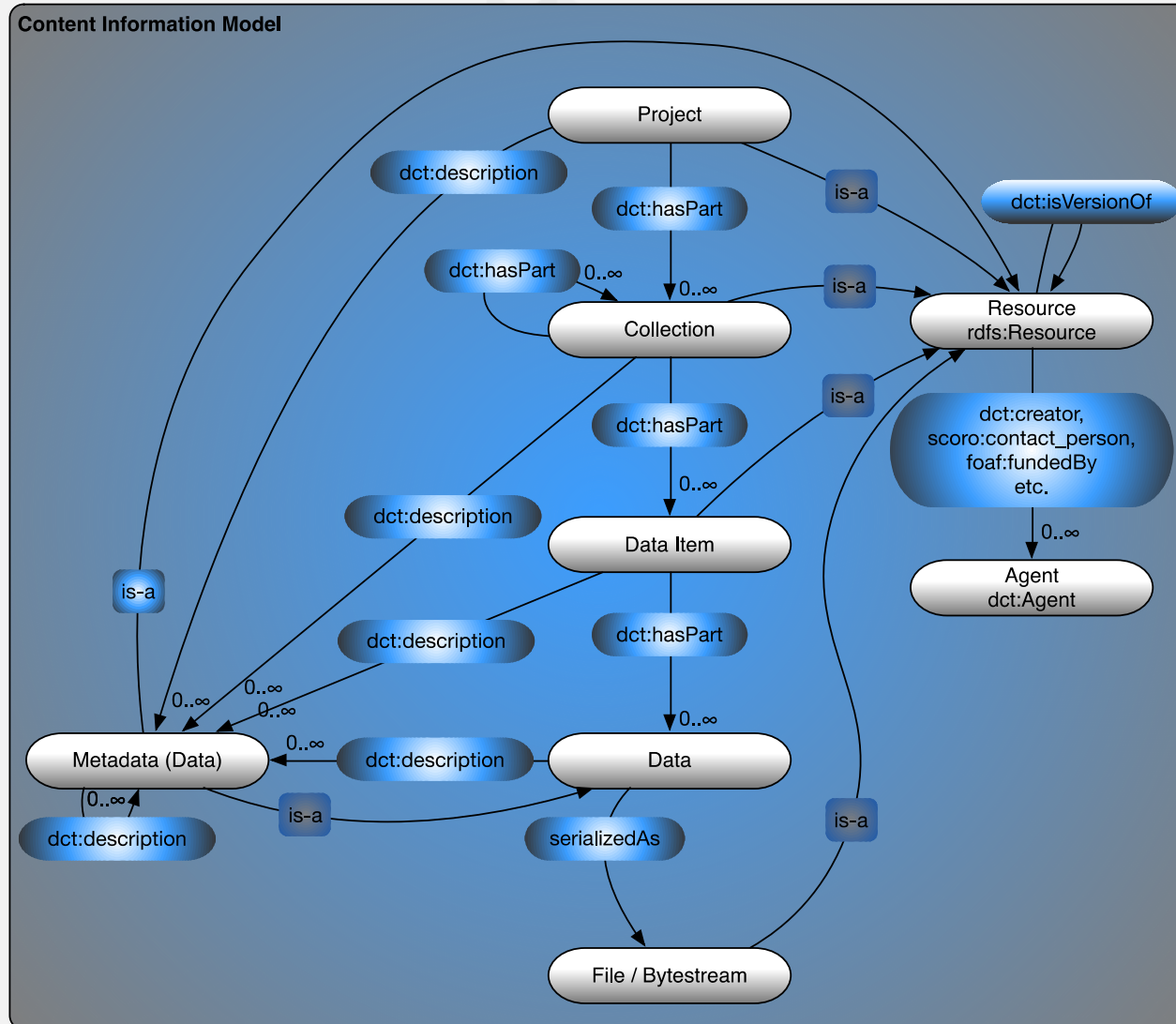


# Rough Data Model





# Content Information Model

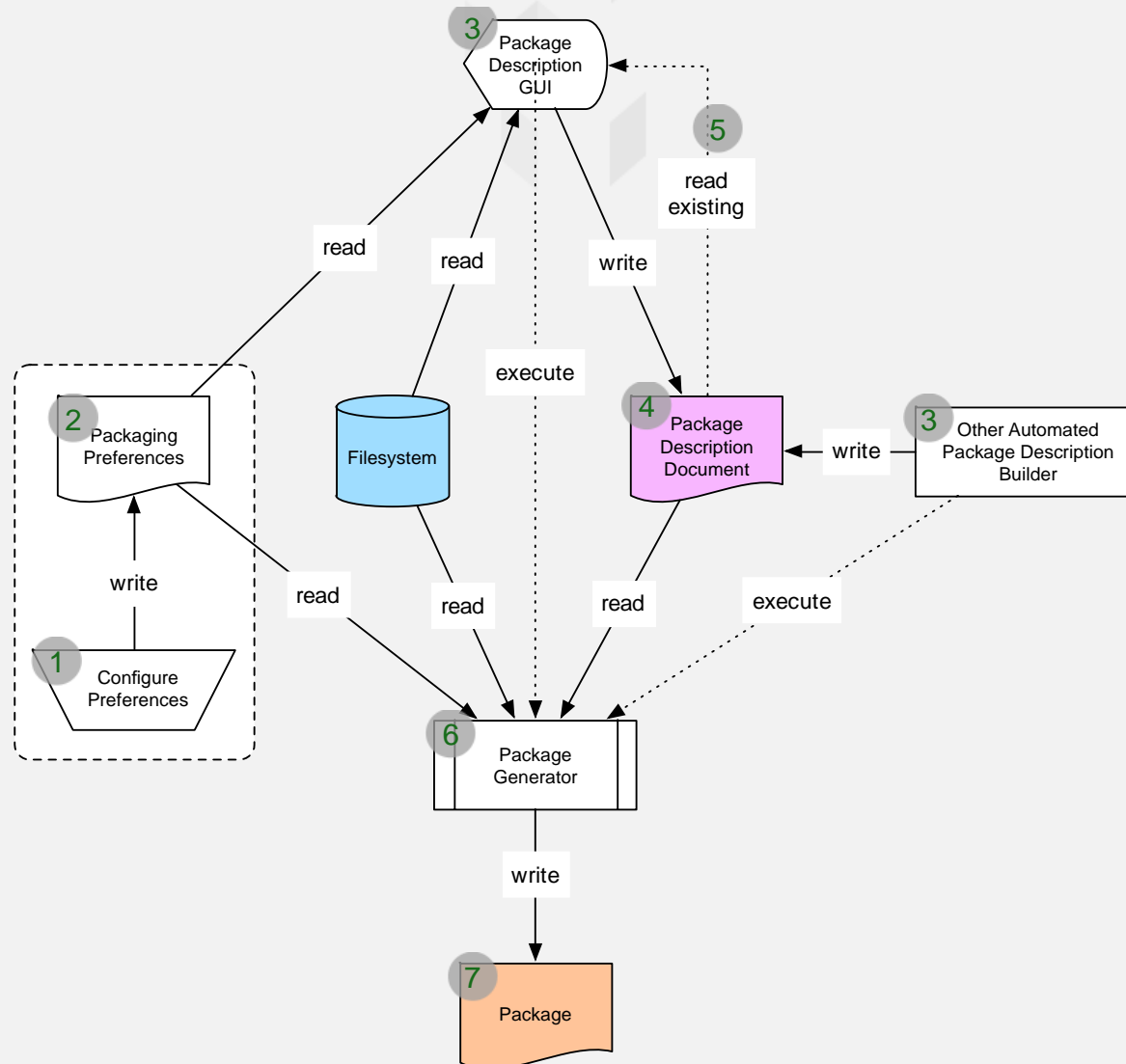






# Capabilities

- Ability to serialize package to a single file, avoiding some OS-specific issues such as hidden files and misinterpreted characters during research data transfer.
- Captures checksum(s) for each packaged bytestream, allowing integrity verification.
- Captures object graph, allowing for expression of relationships among packaged research data objects and external resources.





# Demo



# Challenges

- Building on one platform and enhancing on another
  - Depends on parallel file system structure
- Use of packaging tools from external drive across multiple platforms
  - Currently requires multiple installations on the drive
- Too easy for file name collision between package description document and serialized package
  - Package description doesn't include file extension





# Future Work

- Package editing w/ original content verification (in the queue)
- Default file extension for package description document (done, but not yet released)
- Ontology-driven configuration (under way)
- Configuration from a web service endpoint
- Improve platform-independence and interoperability
- Additional package models and serializations (e.g., LDP, Research Object Bundle)
- Additional package description models/serializations (e.g., LDP)



JOHNS HOPKINS  
UNIVERSITY

THE SHERIDAN  
LIBRARIES



- <http://dataconservancy.org>
- <http://dmp.data.jhu.edu>