

# Sensitive Data Handling in the Social Sciences

CESSDA & UK Data Service Use Case

Hervé L'Hours [herve@essex.ac.uk](mailto:herve@essex.ac.uk)

## CESSDA UKDS - About

- CESSDA- Consortium of European Social Science Data Archives
  - 22 member countries and one observer
  - Explore use case specification across Service Providers
- UK Data Service
  - UK Service Provider to CESSDA
  - partnership between the Universities of Essex, Manchester, Southampton, UCL, Edinburgh and Jisc since 2012
- Lead partner: UK Data Archive
  - Originally Established 1967 at the University of Essex

## Context

- Data about humans in the Social Sciences
- Secure Remote Access
- Access via Safe Room
- Hybrid file-driven and RDF Linked Data Ecosystem
- ISO27000 Information Security
- UK Digital Economy Act Certification
- EOSC PID Policy & evolving expectations of PID kernel metadata

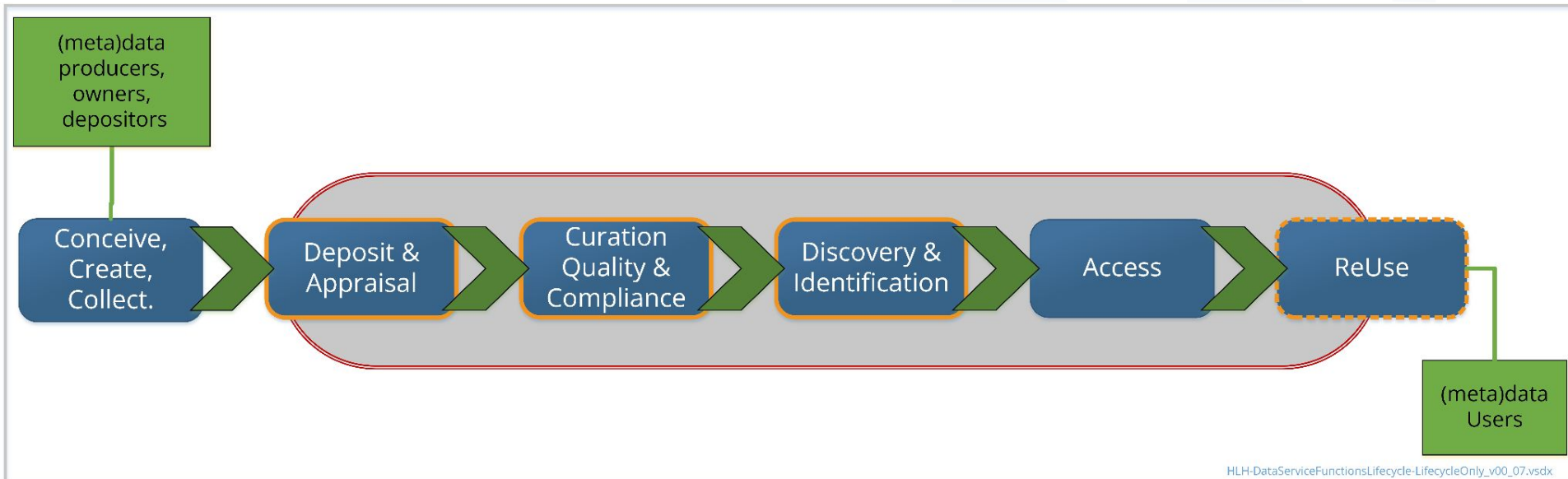
## Scope & approach

Digital Object Models (data & metadata), kernel metadata, rights management

Sensitive data and non-sensitive data classification and division

- At deposit
- During initial curation and, later active preservation (if sensitivity concerns change)
- Handling synthetic data, subsetting, outputs for checking etc

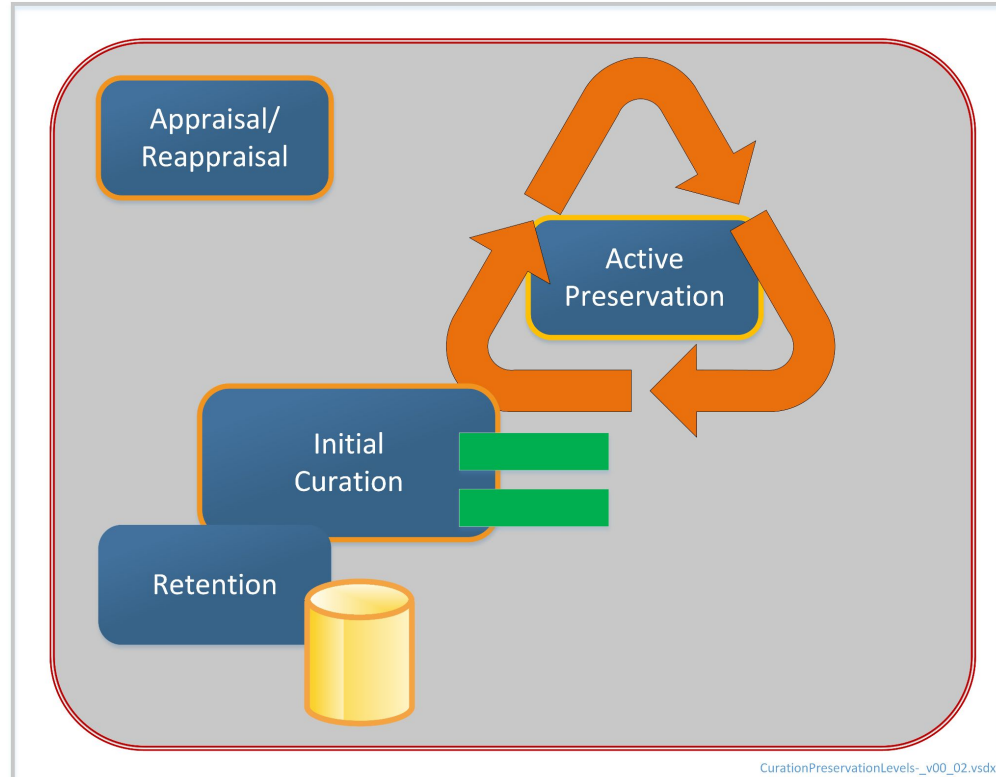
# Lifecycle Perspective



## Trustworthy & Transparent Research Data Management

What is appraised & reappraised to be:

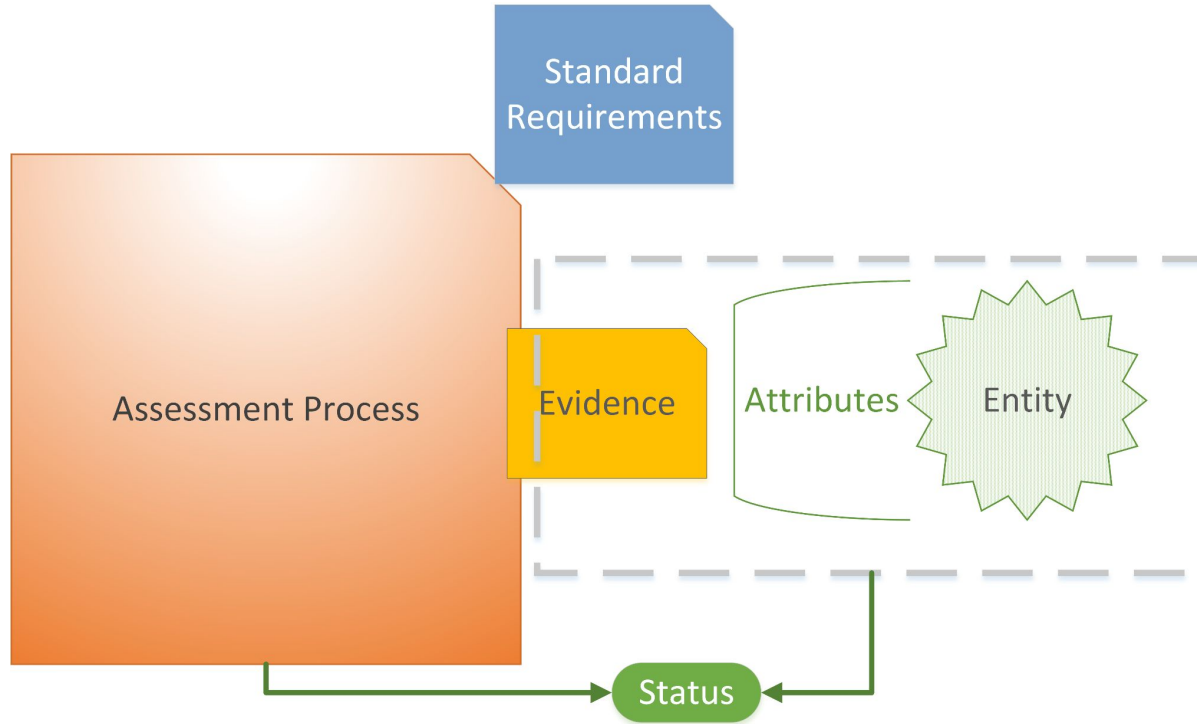
- retained
- curated
- preserved



## Scope & approach: Considerations

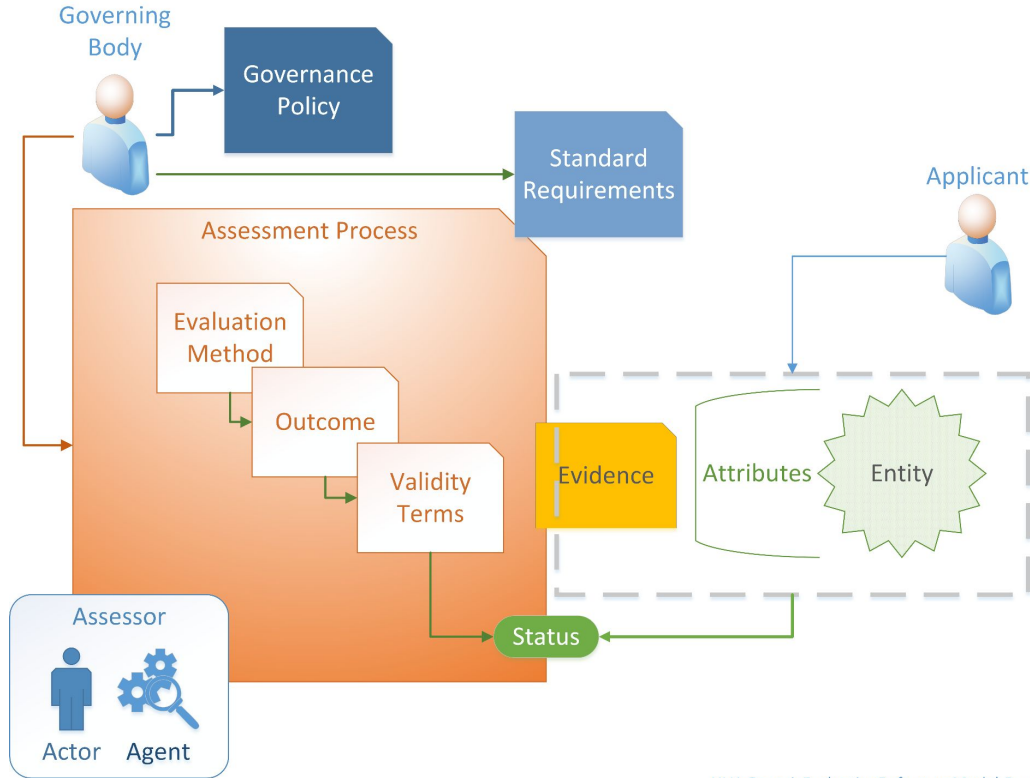
- **Classifications of Sensitivity**
  - Personal data, commercially, environmentally, culturally sensitive
  - Data vs. metadata
- **Assertions of Sensitivity**
  - Data owner-prescribed rights (self-declaration of sensitivity)
  - Intrinsic properties of the data ( empirical run-time analysis)

# Sensitivity Evaluation

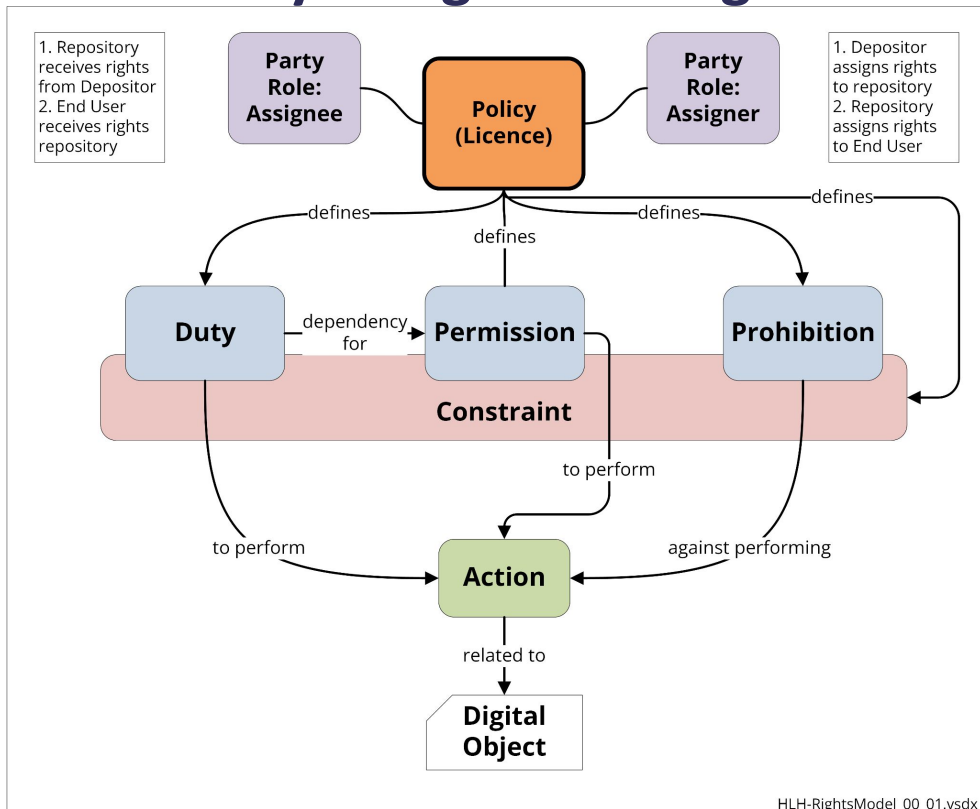




# Sensitivity Evaluation



# Sensitivity & Rights Management



## Scope & approach: Considerations

- **PIDS Resolving & Routing to...**
  - Multiple serialisations e.g. XML, JSON etc. or RDF (RDF/XML, JSON-LD, Turtle etc.)
  - vs html landing pages & catalogue records
  - cf: machine actionability
- **Granularity & Version Management**
  - Sub-“object”: variables (statistics), questions (surveys) etc
- **PID Syntax and Semantics**
  - RDA PID Kernel recommendation, & FAIRCORE4EOSC

## Five Safes Framework

- Safe Data
- Safe Projects
- Safe People
- Safe Settings
- Safe Outputs



## Expected outcome & added-value

- Practical guidance on optimal PID usage (access & management) for sensitive data
- Better aligned and documented practices
- Consistent handling of sensitive data within and between secure environments
- More efficient use of PIDs for sensitive data to benefit research & deliver societal and economic impact