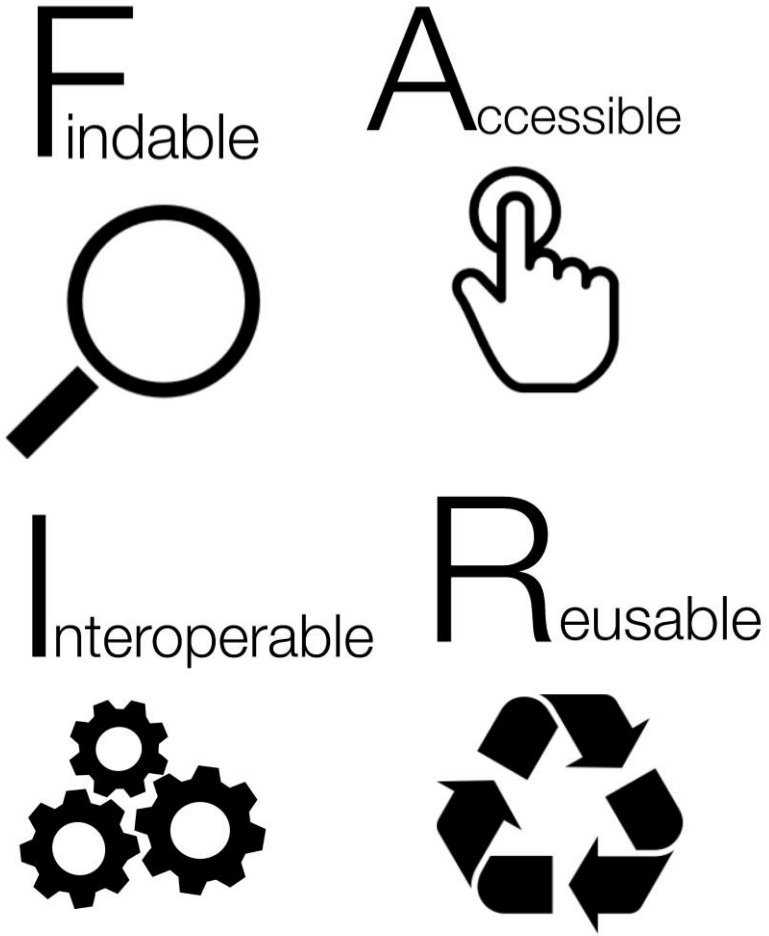
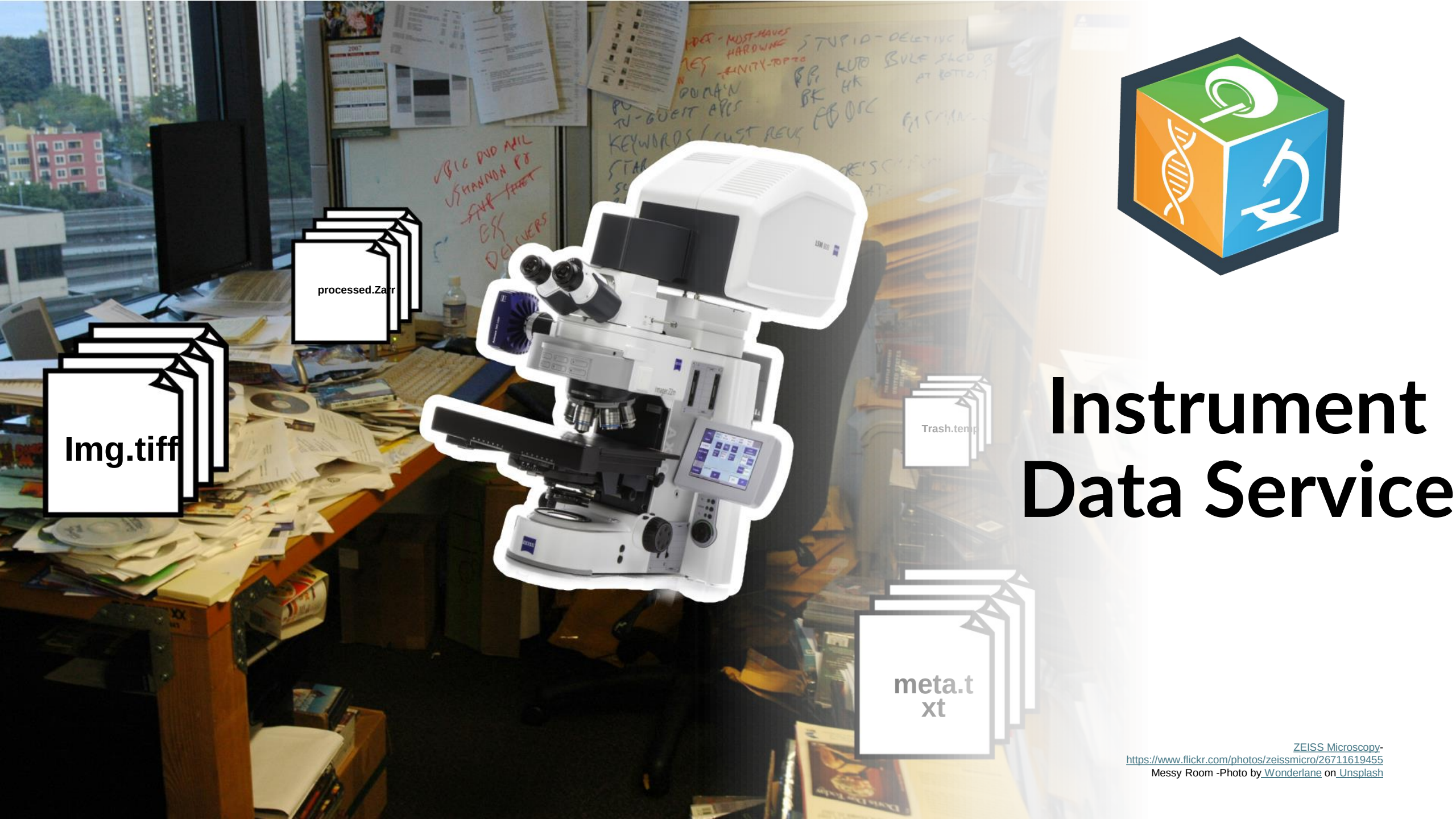
# HERO-Crate

Heuristically Encrypted Research Object Crates

# Secure F.A.I.R Metadata

- Everyone wants their Data to be F.A.I.R
  - But this requires strategies and infrastructure (Metadata!)
- Sensitive metadata should be:
  - As transparent as possible
  - As protected as required






# Instrument Data Service

ZEISS Microscopy-https://www.flickr.com/photos/zeissmicro/26711619455
Messy Room -Photo by Wonderlane on Unsplash

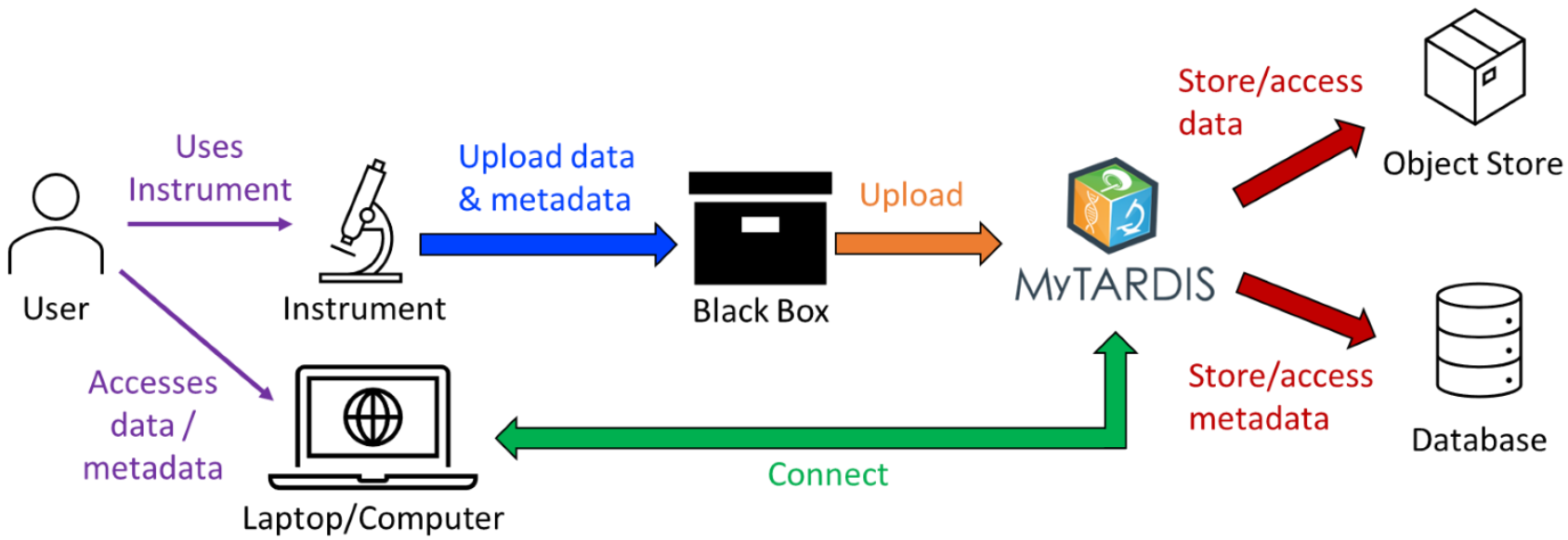
# Instrument Data Service

ZEISS Microscopy-
https://www.flickr.com/photos/zeissmicro/26711619455
Paper Archive Photo by Nana Smirnova on Unsplash

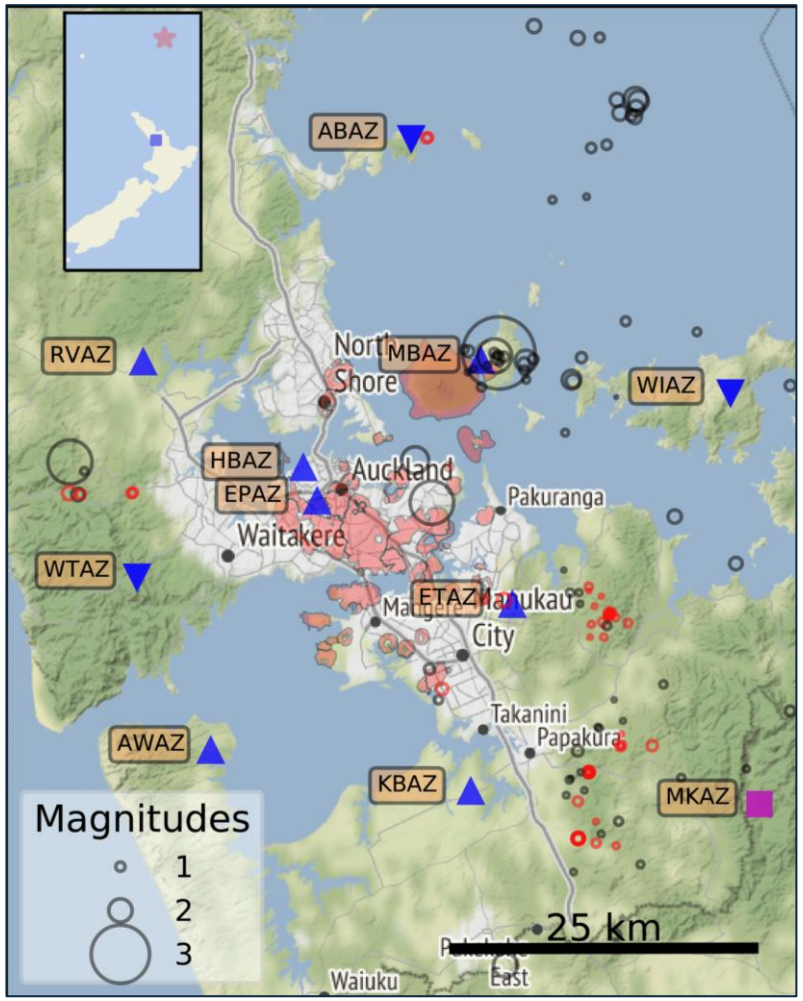
# Ingestion into the IDS

Metadata Collected:

- National Health Index Number
- Date of Birth
- ICD-11 Diagnostic Codes



# Disaster Recovery



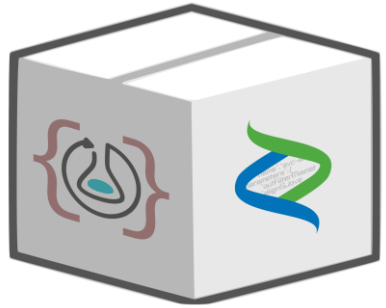


van Wijk, K., Chamberlain, C. J., Lecocq, T., & Van Noten, K. (2021). Seismic monitoring of the Auckland Volcanic Field during New Zealand's COVID-19 lockdown. *Solid Earth*, *12*(2), 363–373. https://doi.org/10.5194/se-12-363-2021

# Need a metadata backup and ingestion solution that is both F.A.I.R and Secure

# F.A.I.R Metadata Packaging

RO-Crate



- Metadata packaged with data
- It's Linked JSON

WorkflowHub

RoHub

# It's Linked JSON

## Data Entities


```
{
"@id": "Vault/pancreatoblastoma/bam/",
   "@type": "Dataset",
   "projectOwner": [
       {
           "@id": "https://orcid.org/0000-0001-7760-1240"
       }
   ]
},
{
   "@id": "Vault/pancreatoblastoma/raw/rna/1806KHP-
0132/A0006L_1.fastq.gz",
   "@type": "File"
},
```


## Context Entities


```
{
   "@id": "https://orcid.org/0000-0001-7760-1240",
   "@type": "Person",
   "email": "JCarberry@psychoceramics.brown.com",
   "identifier": "jcar034",
   "name": "Josiah Carberry",
   "organization": [
       {
           "@id": "#UOA_FMHS"
       }
   ]
},
{
   "@id": "#UOA_FMHS",
   "@type": "Organization",
   "name": "University Of Auckland Faculty of Medical and
Health Science"
}
```

# GnuPG and OpenPGP

- Established standard
- Portable and Lightweight
- Wrapper libraries in Python and JS (nice for RO-Crate)

# HERO-Crate

Heuristically Encrypted Research Object Crates

A.K.A OpenPGP-Crate

(https://uoa-eresearch.github.io/GPG-ro-crate-profile/)

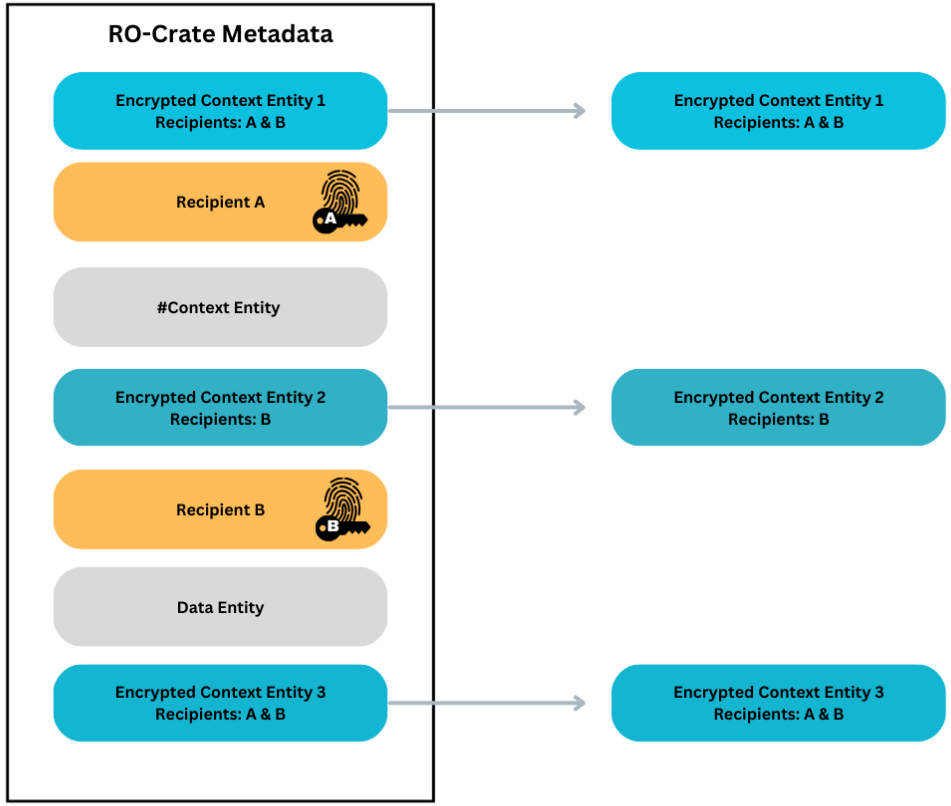
Love, James (2024). OpenPGP-Crate Whitepaper. The University of Auckland. Standard. https://doi.org/10.17608/k6.auckland.27288039.v1

# Encrypting Metadata

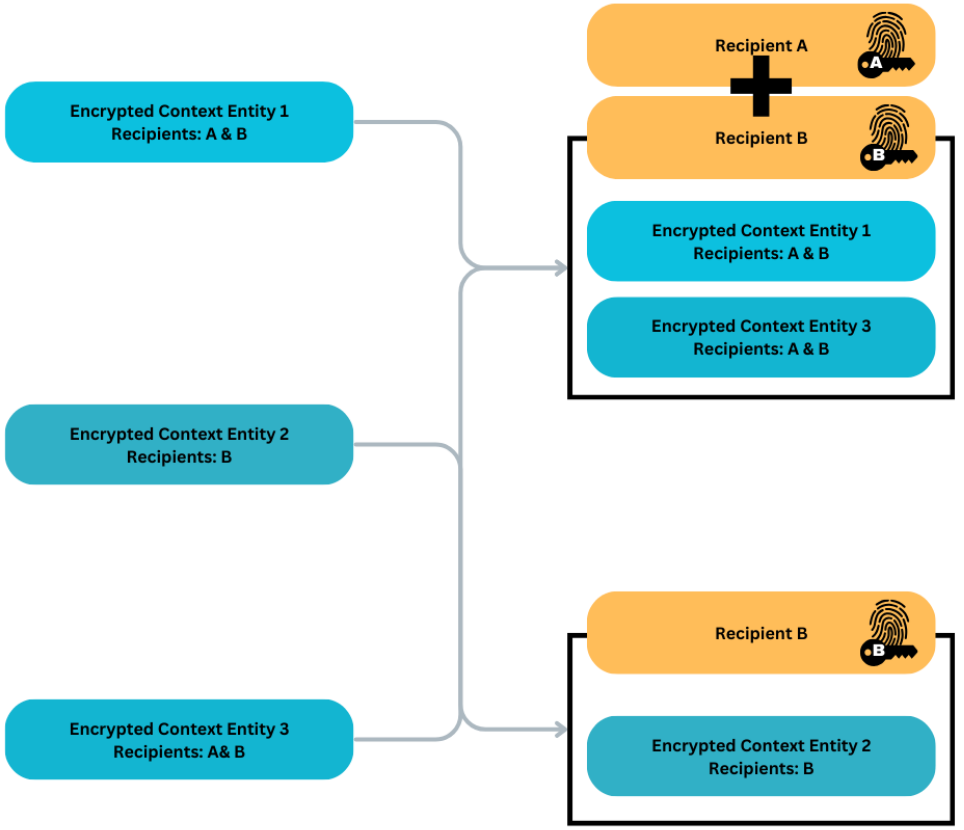
- **In-Memory**
  - Any Context Entity can be an *Encrypted Context Entity*
  - Before writing to disk GPG encrypts all Encrypted Context Entities as PGP messages
- **On-Disk**
  - PGP Messages are stored in *Encrypted Graph Messages*
  - When reading the RO-Crate *Encrypted Graph Messages* are decrypted (if possible) into *Encrypted Context Entities*

# Metadata Encryption

## 1. Extract All Encrypted Context Entities

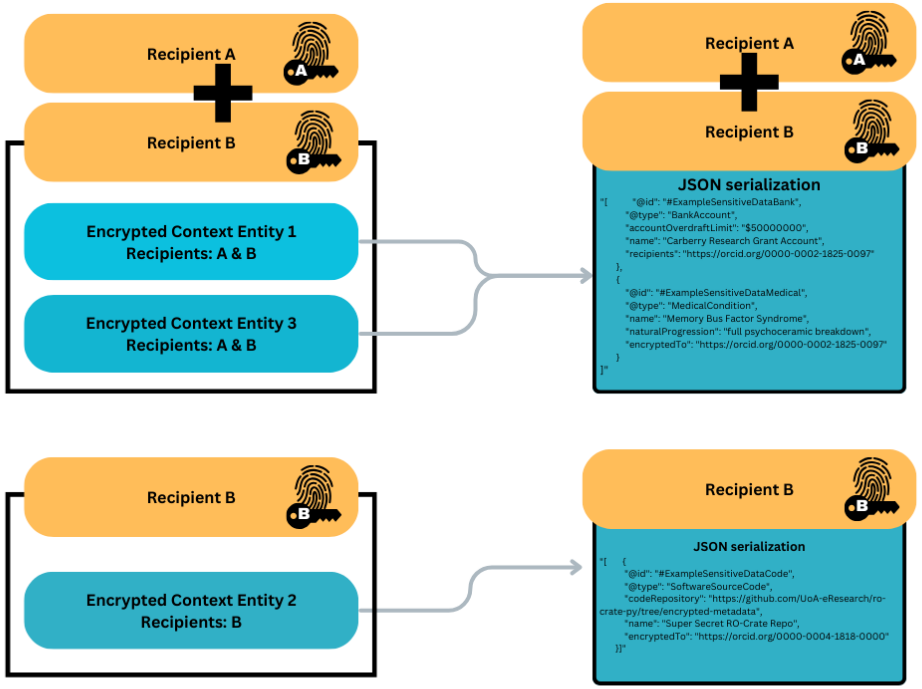



## 2. Aggregate Encrypted Context Entities by Shared Recipients

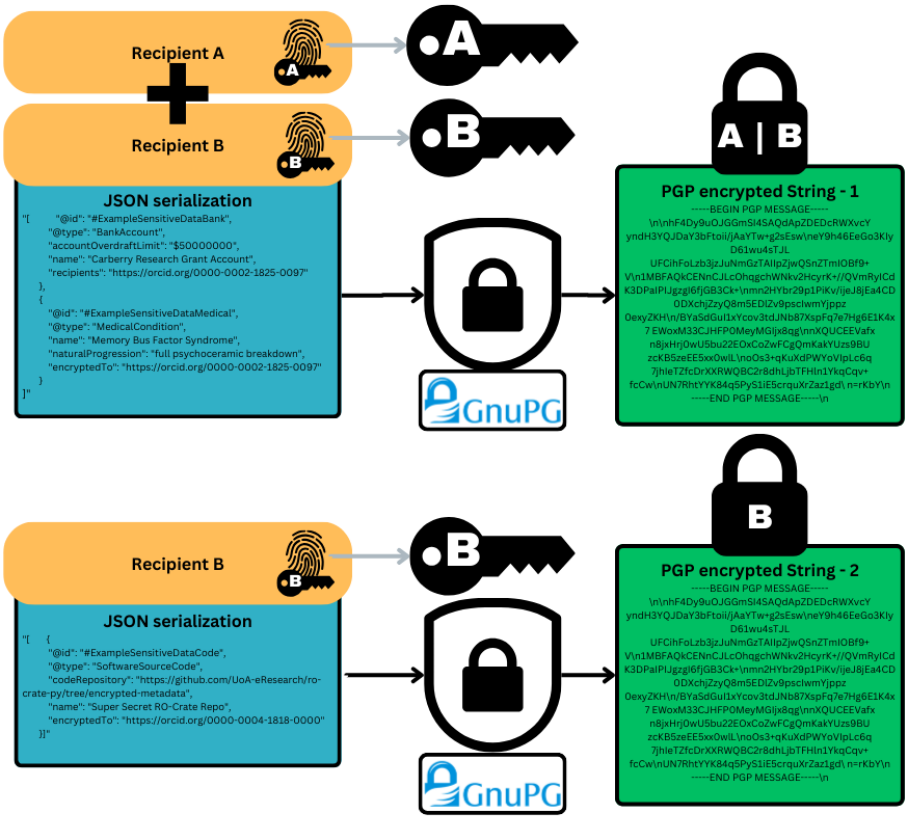


# Metadata Encryption

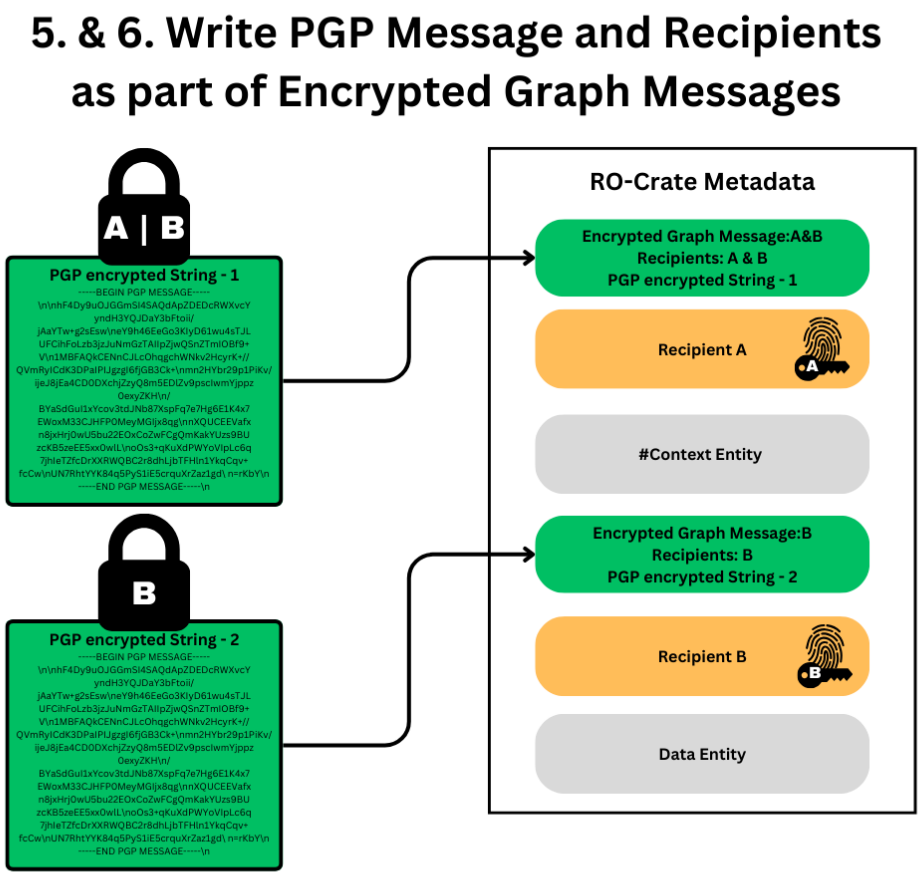
## 3. Serialize Entities To JSON LD




## 4. Encrypt JSON using Recipient's Public keys



# Metadata Encryption



# Encrypted Graph Messages

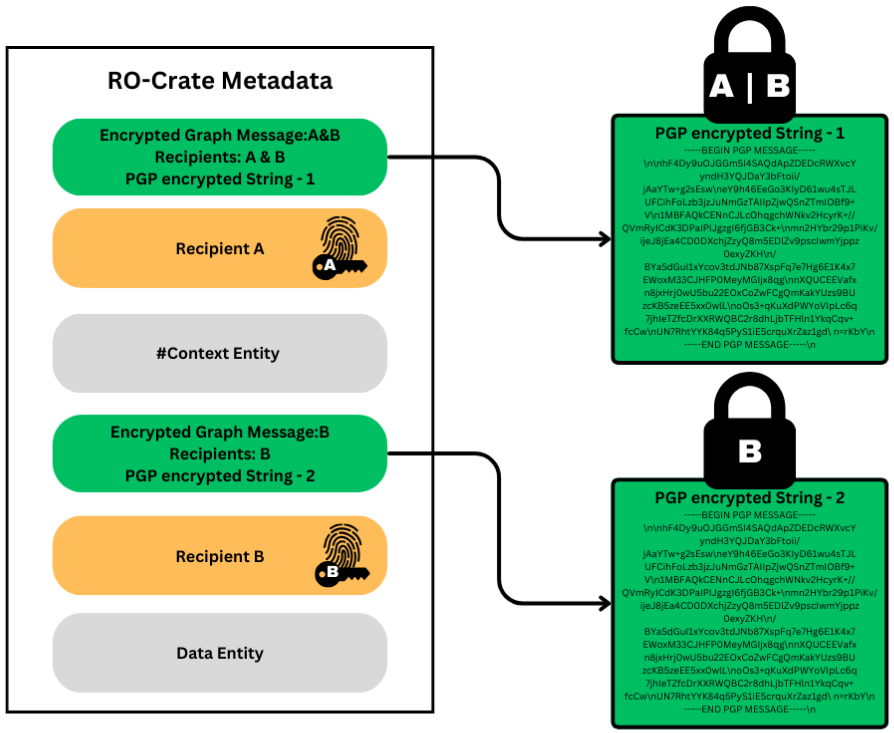

```
{
   "@id":
"#Encrypted_MessageA86F04EAD1342A90F538ED7F0221D767C9AEE494
",
   "@type": [
       "SendAction",
       "EncryptedGraphMessage"
   ],
   "actionStatus": "PotentialActionStatus",
   "deliveryMethod": "https://doi.org/10.17487/RFC4880",
   "encryptedGraph":,
   "encryptedTo": [
       {
           "@id": "https://orcid.org/0000-0001-7760-1240"
       }
   ]
},
```



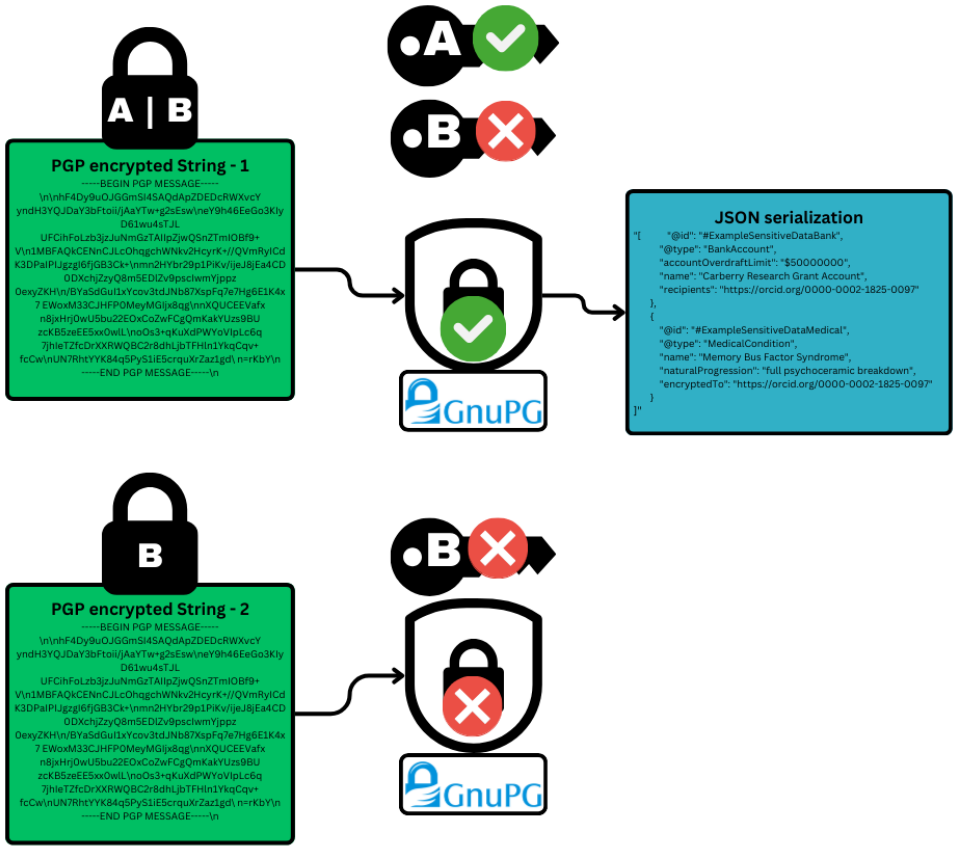
```
-----BEGIN PGP MESSAGE-----
hF4DVhe2+C+HB+0SAQdAzFc3uYBXXvfuoOWThBCyy+Hlae6JFhil9yH3Kx
DIjzowMHq3CsQhwqr9qDrQASCVDu1pICHczxF9ySj2gIlkXGZ701QRLpiHq
GuA4ygFKFUs1MCIAQkCEB/dXpQeMH/v2Qet1e+gibfpilFVQKYfRYIvZeT4
TZ7KuwHHn2ylI2pHsQe/RUshmneAyzyE70sYvXxJuXZXJfaS8/dAw1spVvt
qCjG7Fm3jIXITmFW8XLnAG+YrAJHTWDCPmkeMYDeLhybEG5FKgmk0ZsHjbf
SRMhfaKa2wptUIKsCnyNAHh0ln98bMt3R14REeFuJIvzss61/WXWFf8jfyz
GzDsapebr/bfGv/YhI5u2uu0Q+MfjKLpeyRPmCz1IWaCxN2JvCHee9oYUJJ
PJ7ql1Jg7cwVzjLDdm+CwkvMccrLcf9b88WLeTq0gTfi/KJIDsayHEPYPJZ
KrpiEtbHZeOq2lShHPpX/qqUQB0rptlZ5brq588q7rBl+DpTPkRmutXemBq
7OLbBw33A+3jJMbEQRVFOzEqXRoyOtJgNj3EPZ6w===vc2Q
-----END PGP MESSAGE-----
```

# Metadata Decryption

## 1. & 2. Extract PGP strings from Encrypted Graph Messages in the RO-Crate

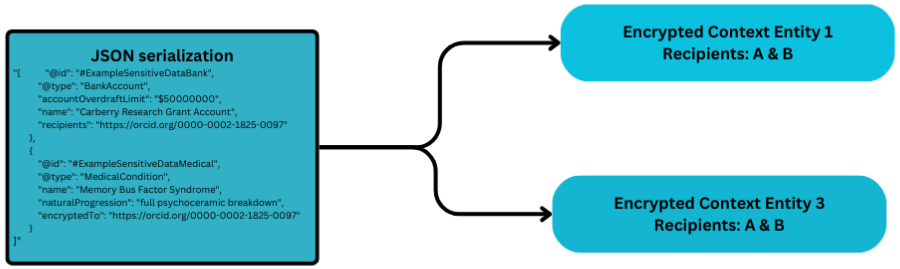



## 3. Attempt To Decrypt the PGP Strings Using Available Private Keys

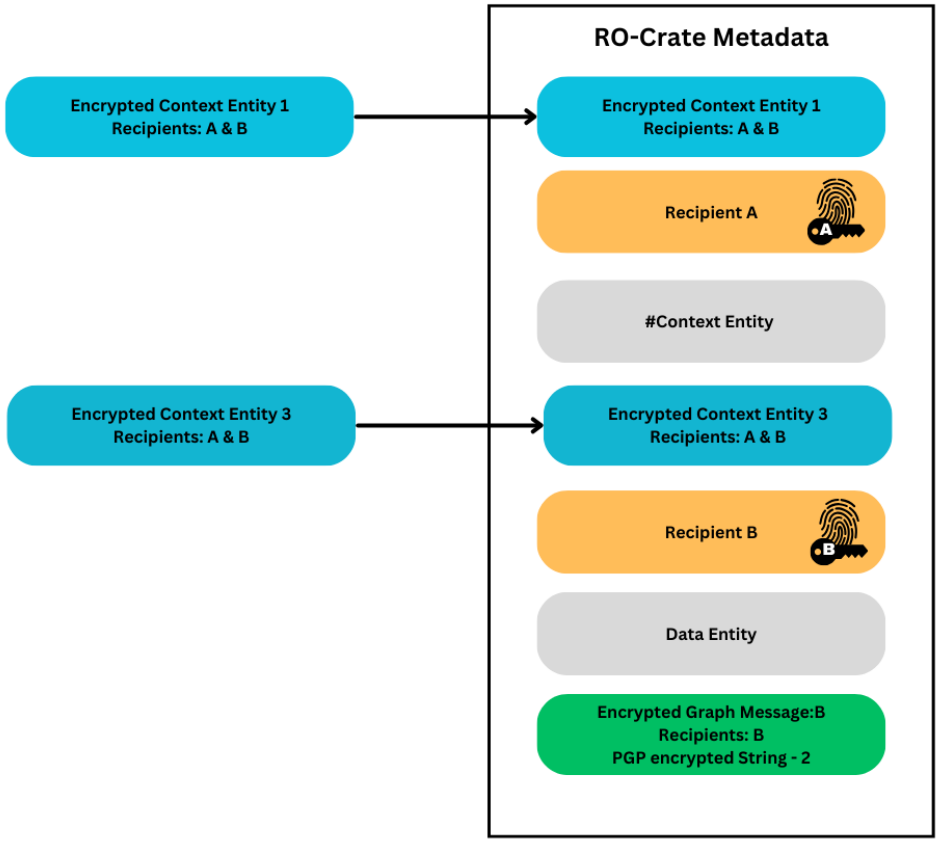


# Metadata Decryption

## 4. Reconstruct Encrypted Context Entities From Raw JSON

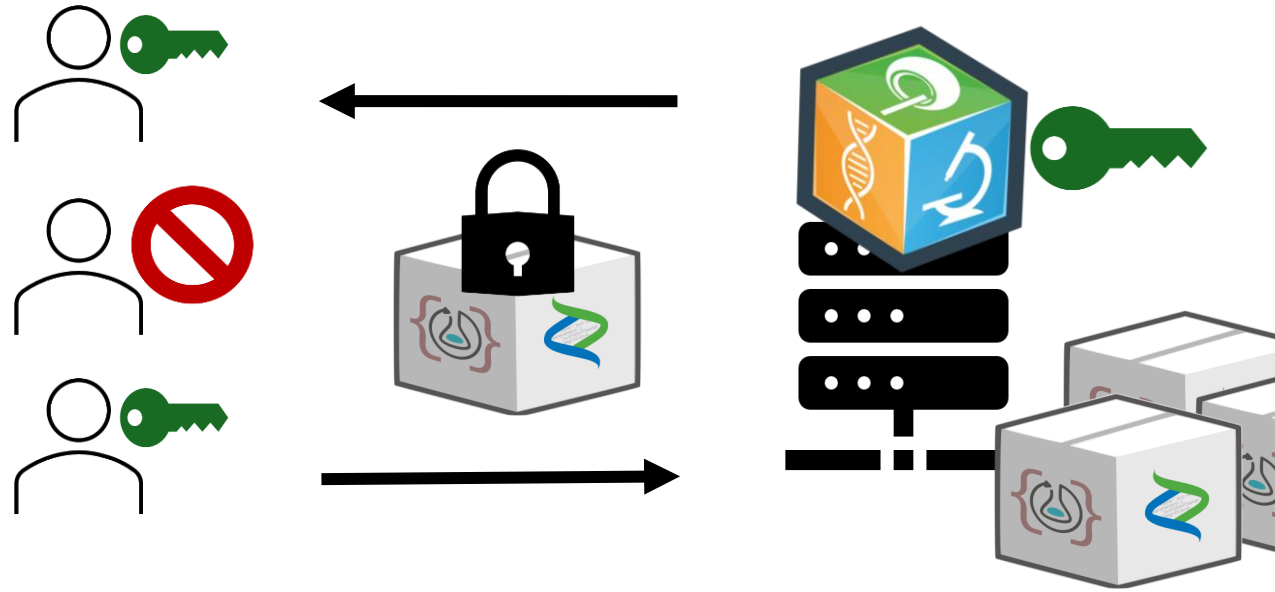



## 5. Re-insert Encrypted Context Entities into RO-Crate



# IDS Metadata backup

- RO-Crate backup of all metadata for each dataset
- Metadata alongside data
- Encryption based on ACLs and User information
- MyTardis internal keys



# Other Applications

- Non-instrument Data Archives
- Workflow Protection & Provenance
- Secure Research Environment
- Technical means of enforcing 5-Safes rules

# Thank you