DOI and User Upload

Alexander Kemp - Software Engineer - DSEG

DOI Work to Date

- Rough proof of concept demo 'ed in Feb (thanks to a broken collarbone!)

- Working with Diamond to refine requirements through [Use Cases](#)

- Speaking to ISIS and ESRF to understand how they do it

- Broken down into two main areas of functionality

  - Automatic minting based on the investigation end date

  - User-defined minting of DOIs

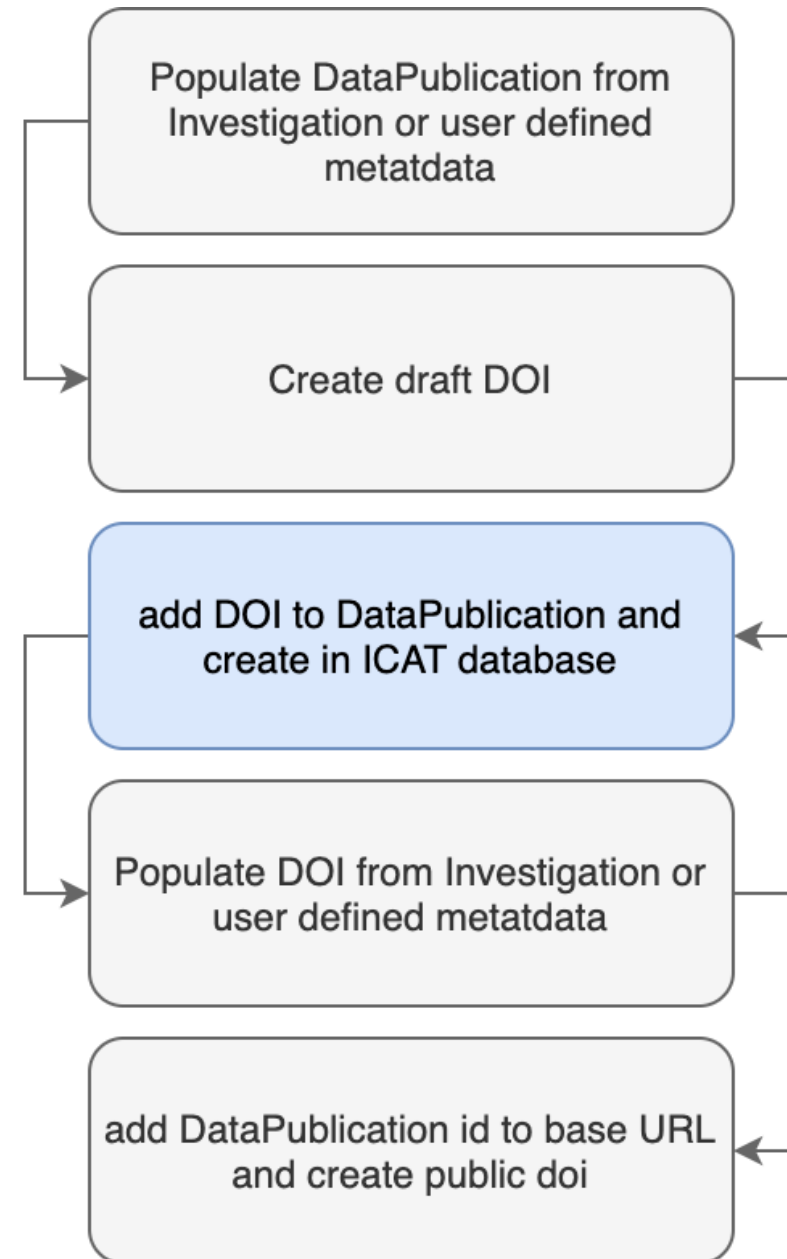- Plan on releasing a final POC end of June and Starting Upload Dev
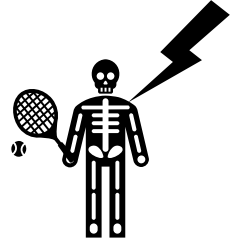
How?

- Two main cases:
  - That a script will need to automatically create Data Publications based on information provided in an Investigation.
  - That user (PI) will want to create a Data Publication based on a set of Visits / Datasets / Datafiles with information or metadata that they provide.

- Fast API
- Microservice
- Containerised
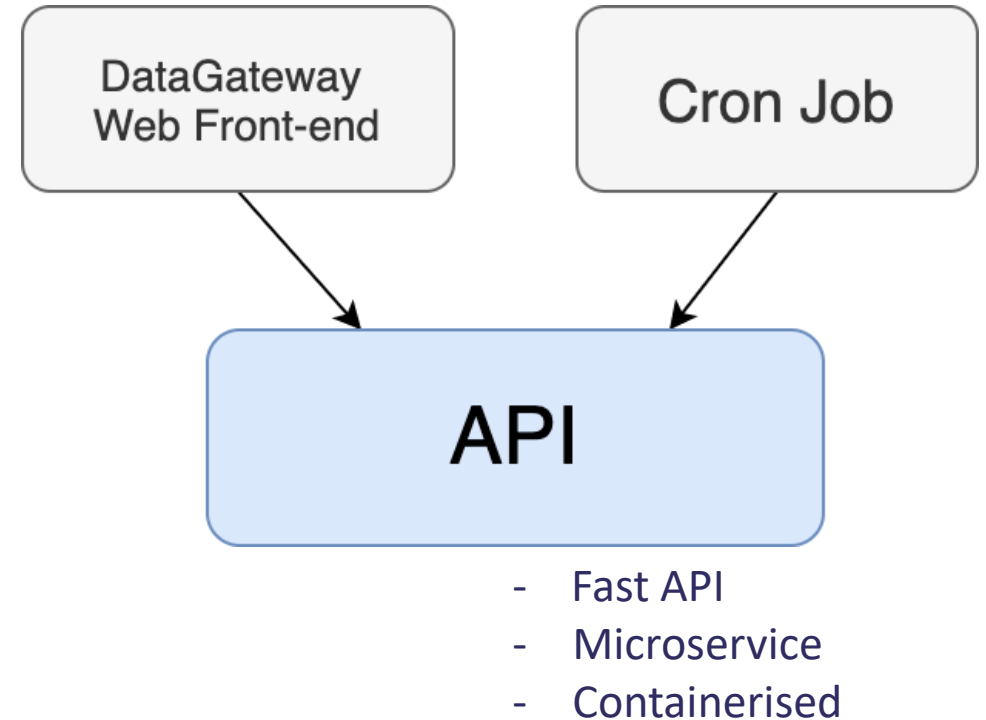
What?

- Creating a DataPublication touches:
  - Dataset
  - DataCollectionDataset
  - DataPublicationUser
  - Investigation
  - Affiliation
  - Datafile
  - Facility
  - User
  - DataCollectionDatafile:
  - InvestigationUser
  - DataCollection
  - DataPublicationType
  - DataPublicationDate
  - DataCollectionInvestigation

- DP needs a DOI, but a DOI needs a DP for the landing page URL

Populate DataPublication from Investigation or user defined metatdata

Create draft DOI

add DOI to DataPublication and create in ICAT database

Populate DOI from Investigation or user defined metatdata

add DataPublication id to base URL and create public doi

Difference in Facility Approach

- Diamond: Difference in Reqs

  - Publish at a later date

  - Create from Investigations / Datasets / Datafiles

  - We do more of the ETL on our side

- ISIS:

  - Only automatic minting of whole investigations, based on when a data file is entered into ICAT.

  - No User-defined DOIs, but they will hopefully be moving over to this new DOI tool.

Difference in Facility Approach..cont

- ESRF:

    - They only produce DOIs for datasets

    - An email is also sent to the participants of the investigation.

- Deltas documented [here](#) if anyone wants more details

Event Flow for Automatic Process
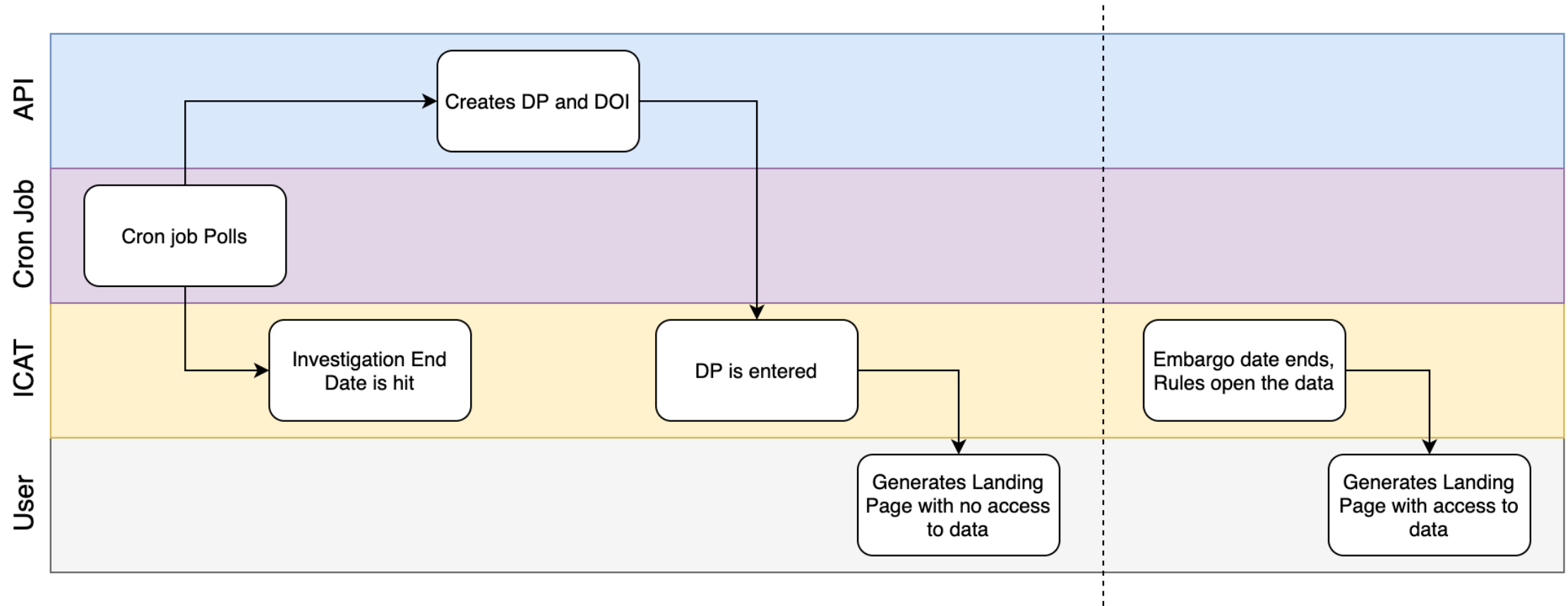
UI Flow for Manual Process

- In [Figma](#)

Changes to Schema?

- Need to maintain landing pages for removed DataPublications.

    - So need to track deleted DPs that are created erroneously

    - Can't remove DOIs so they'll still need to point to a landing page where the user will be informed that the Publication has been removed.

- Need a flag to denote proprietary data

- Not implemented yet

    - Will add issues in github

Upload work to Date

- Provide the functionality to allow Diamond user to supplement their data online

- Using this as part of my CS MSc

- Currently looking at underlying protocols and frameworks

- Met with Zenodo SRE, who pointed us towards www.tus.io

- Current Challenges include:

    - Choosing right protocol

    - Imposing limits

    - Making sure it's not abused

Upload work to Date…cont

- DOI: Users shouldn't be able to supplement their data once a DP has been created.
  - Given an investigation with a DOI and DP, a user shouldn't be able to add to a Publication
  - Solution probably to create a new DP for all uploaded data and relate it to other DP/DOIs