



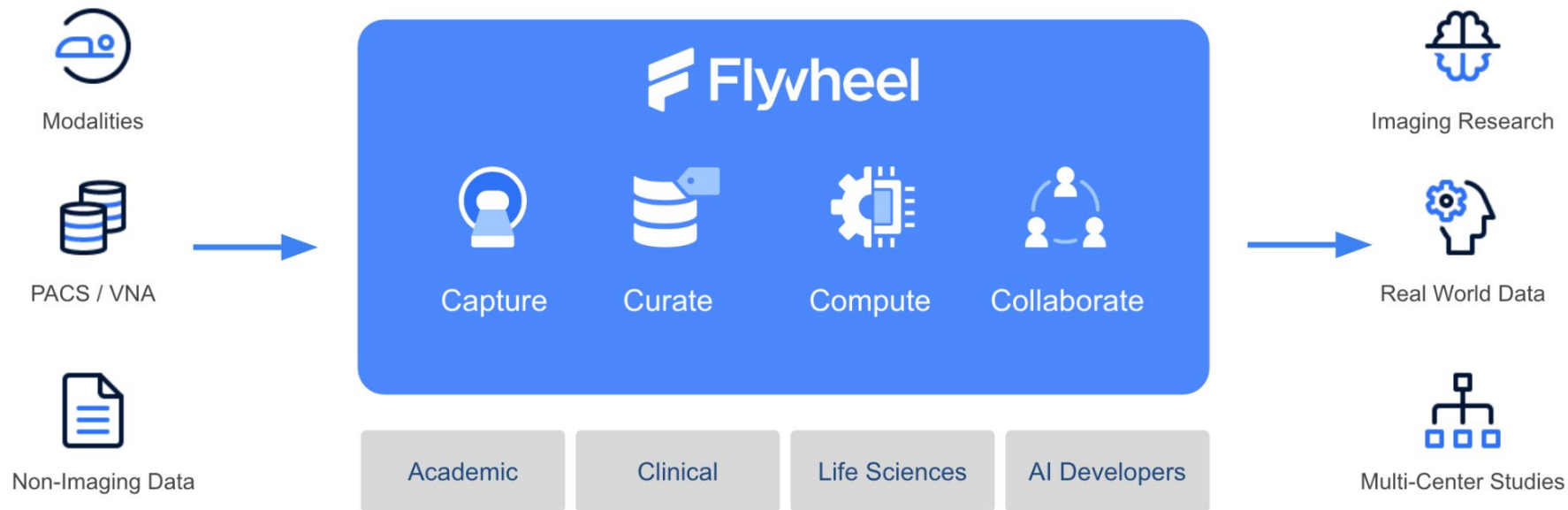
The Biomedical Research Data Platform
Powering Healthcare Innovation

The Flywheel Platform for Intelligent Image Anonymization

MIDI Workshop
May 22, 2023

Daniel Marcus
Chief Scientific Officer, Flywheel
Professor of Radiology, Washington University

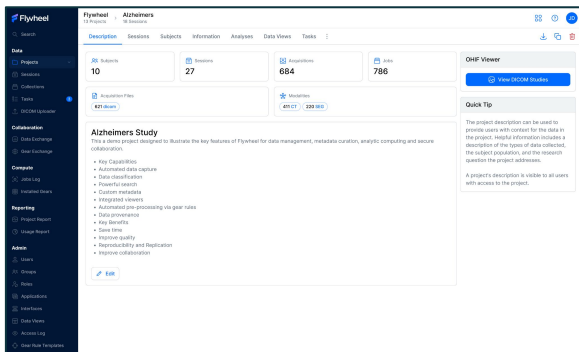
Flywheel - Medical imaging data management & collaboration



The Flywheel product suite

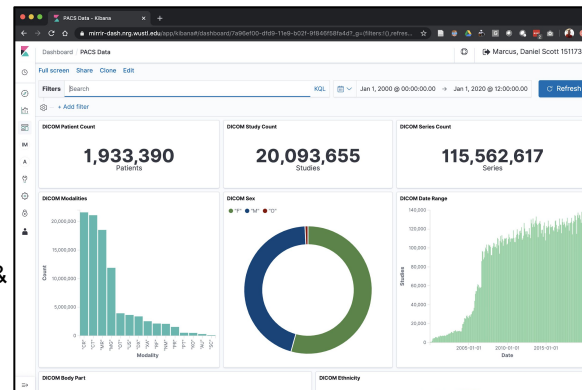
Productivity Flywheel Enterprise

The flagship platform for data management, image analysis, and scalable computation.



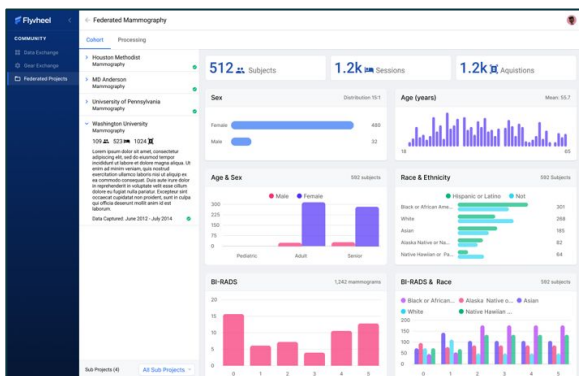
Capacity Flywheel Discovery

The clinical platform for cohort management, imaging feature extraction, end-to-end translational research, and clinical business intelligence.



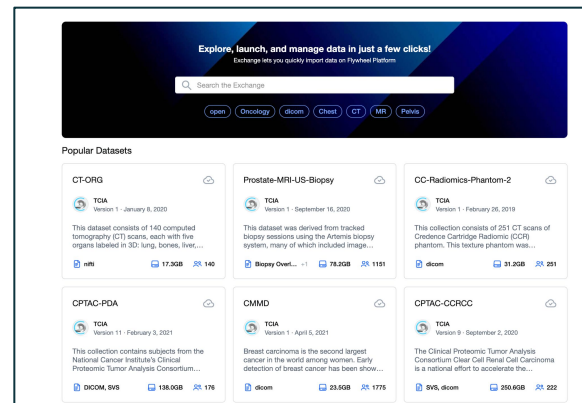
Availability Flywheel Exchange

A marketplace and collaboration platform for sharing data and algorithms and to enable federated learning and analysis.

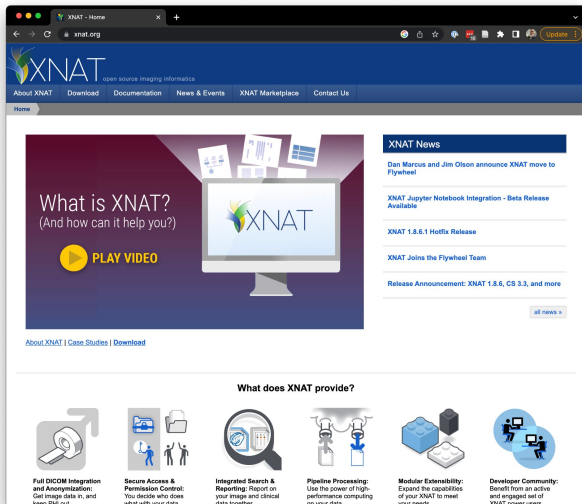


Ubiquity Flywheel Cloud

Multi-tenant Flywheel services as a self-service entry point to the Flywheel ecosystem.



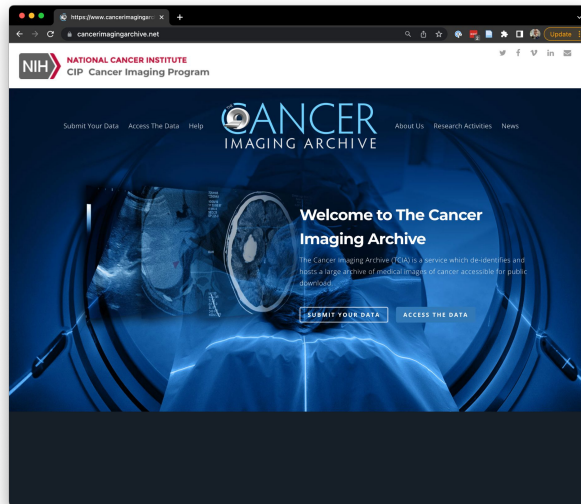
Flywheel open science initiatives



The screenshot shows the XNAT website homepage. At the top, there's a navigation bar with links for 'About XNAT', 'Download', 'Documentation', 'News & Events', 'XNAT Marketplace', and 'Contact Us'. The main content area features a large banner with the text 'What is XNAT? (And how can it help you?)' and a 'PLAY VIDEO' button. Below the banner, there's a section titled 'What does XNAT provide?' with six icons representing different features: Full DICOM Integration and Annotations, Secure Access & Permission Control, Integrated Search & Reporting, Pipeline Processing, Modular Extensibility, and Developer Community. A 'XNAT News' sidebar on the right lists recent announcements, including the move to Flywheel and the release of XNAT 1.8.6.1.

XNAT

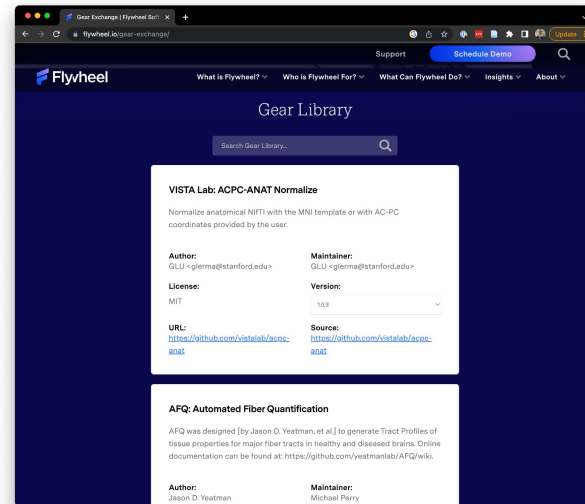
Open source XNAT provides a powerful informatics tools accessible to all researchers, expanding Flywheel's impact and powering our federated data network..



The screenshot shows the Cancer Imaging Archive (CIP) website. The header includes the NIH logo and the text 'NATIONAL CANCER INSTITUTE CIP Cancer Imaging Program'. The main content area features a large banner with the text 'Welcome to The Cancer Imaging Archive' and a subtitle 'The Cancer Imaging Archive (CIP) is a service which de-identifies and deposits a large archive of medical images of cancer accessible for public download.' Below the banner, there are two buttons: 'SUBMIT YOUR DATA' and 'ACCESS THE DATA'.

Open Data

Open data provides a shared communal resource for research and method development. Flywheel Discovery and Enterprise support creation and distribution of open data sets. Open data are shared through Flywheel Exchange.



The screenshot shows the Flywheel Gear Library page. The header includes the Flywheel logo and navigation links for 'What is Flywheel?', 'Who is Flywheel For?', 'What Can Flywheel Do?', 'Insights', and 'About'. The main content area features a search bar and a list of gears. The first gear is 'VISTA Lab: ACPC-ANAT Normalize', which is used to normalize anatomical NIFTI with the MNI template. The second gear is 'AFQ: Automated Fiber Quantification', which is used to generate Tract Profiles of tissue properties for major fiber tracts in healthy and diseased brains. Both gears list the author and maintainer.

Open Compute

Open gears enable tools to be shared, powering common computing methods across data sets. Open gears can be deployed on Flywheel products to power open science.

Flywheel's existing De-ID capabilities

The screenshot shows a web browser displaying the 'De-Identification (de-id) overview' article on the Flywheel website. The page has a dark blue header with the Flywheel logo and navigation links for 'Contact Support' and 'Sign in'. A search bar is located below the logo. A left sidebar contains a 'WHAT'S NEW' section with links to 'Welcome to Flywheel', 'Announcements', and 'Release Notes', and a 'USER GUIDE' section with links to 'Signing in', 'Adding data', 'Creating and managing projects', 'Form Builder', 'Viewing images', 'Gears', 'Gear Rules', 'Task Management', 'Create a config file for the fw ingest command', 'Search', 'Downloading data', 'BIDS resources', and 'Command-Line Interface (CLI)'. The main content area features the article title 'De-Identification (de-id) overview', an 'Updated · 3 months ago' timestamp, and social media sharing icons. The article text explains that de-identification is the process of removing or transforming personally identifiable information (PII) and provides an overview of the workflow in Flywheel. It details how de-identification works, including the use of de-id profiles and the configuration of de-identification in Flywheel. It also lists two main de-identification workflows: 'De-identification at the edge' and 'De-identification gear', each with a brief description of their use cases.

Flywheel Enterprise De-Identification Overview

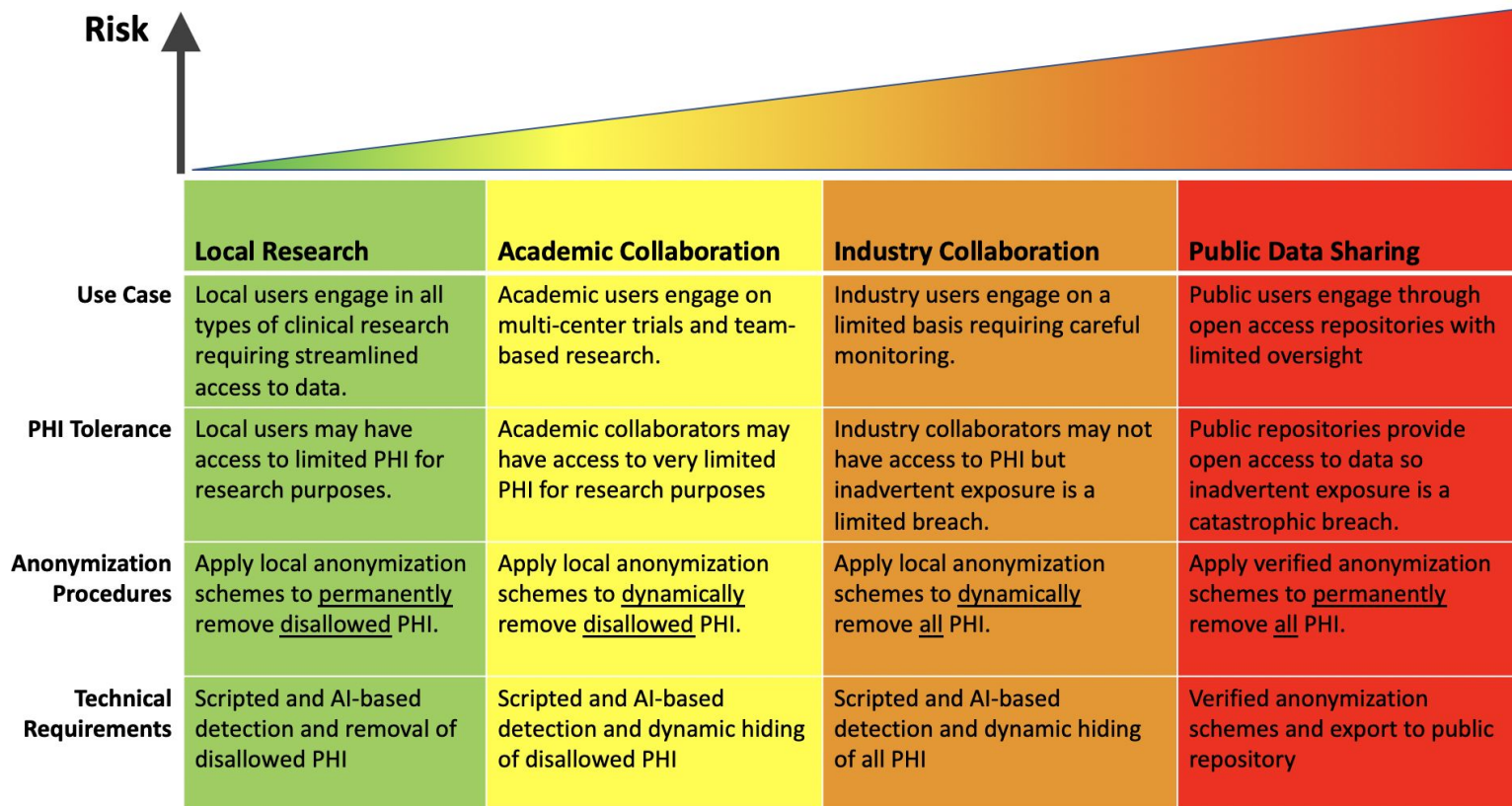
The screenshot shows a web browser displaying the 'XNAT Anonymization 101' course page on the XNAT Academy website. The page has a dark blue header with the XNAT logo and navigation links for 'Explore courses', 'Register', and 'Sign in'. The main content area features the course title 'XNAT Anonymization 101', a description stating 'This course will teach you the importance of, and techniques for, image data anonymization in XNAT', and a yellow 'Enroll in XNAT Anonymization 101' button. Below the main content, there is a section titled 'About This Course' with a table listing course details: Course Number (XANON-101), Self-Paced, and Estimated Effort (10:00/week). The 'About This Course' section also includes a welcome message and a description of the course content, which includes lectures, quizzes, and practical exercises. A 'Course Goals' section is also visible at the bottom of the page.

XNAT De-Identification Overview

Some background

- NCI SBIR Phase II: “Intelligent Image Anonymization with XNAT”
 - Awarded to Radiologics, now operating under the Flywheel Banner
 - Broader than XNAT
 - It’s not like we weren’t already intelligent... but we have tended to brute force our way to de-ID.
- The problem
 - Text-based information in file metadata
 - Text-based information in image pixels
 - Facial features inherent in neuroimaging
 - Other features in the images
- What’s changed
 - Larger data sets
 - More available data
 - NIH data sharing requirements
 - More creative computing approaches (e.g. federated learning)

Flywheel's approach: intelligent risk-based anonymization



Flywheel's approach: intelligent risk-based anonymization

The screenshot displays the XNAT web interface for configuring site-wide anonymization. The browser address bar shows the URL: `harvard-pilot.radiologics.com/app/template/Page.vm?view=admin#tab=session-upload-import-anonymization`. On the left, a sidebar contains navigation options under "Advanced XNAT Settings" (Pipeline Settings, DICOM SCP Receivers, File System, Async Operations) and "Other" (Manage Investigators, Miscellaneous). The main content area is titled "Anonymization Script (Site Wide)" and features a toggle for "Enable Site-wide Anonymization Script?" which is currently "Enabled". Below this is a text area for the "Edit Anonymization Script" containing a JavaScript script. The script performs various DICOM header manipulations, such as removing patient birth date, other patient IDs, names, birth names, weight, and address, and adding a hashUID and a study instance UID. A note explains that this script is applied to all incoming and archiving DICOM resources and must conform to the DicomEdit format. At the bottom of this section are "Discard Changes" and "Save" buttons. The second section, "Series Import Filter (Site Wide)", has a toggle for "Enable Site-wide Series Import Filter?" which is "Disabled". It includes a "Filter Mode" dropdown set to "Blacklist" and a descriptive note about how whitelists and blacklists function in the context of XNAT import tools. A "Save" button is also present at the bottom of this section.

Advanced XNAT Settings

- Pipeline Settings
- DICOM SCP Receivers
- File System
- Async Operations

Other

- Manage Investigators
- Miscellaneous

Anonymization Script (Site Wide)

Enable Site-wide Anonymization Script? Enabled

Edit Anonymization Script

```
(0010,0010) := subject
(0010,0020) := session
- (0010,0030) //Remove Patient Birth Date
- (0010,1000) //Remove Other Patient IDs
- (0010,1001) //Remove Other Patient Names
- (0010,1005) //Remove Patient Birth Name
- (0010,1030) //Remove Patient Weight
- (0010,1040) //Remove Patient Address
(0020,000D) := hashUID[(0020,000D)] // Study Instance UID
(0012,0063) := "XNAT site level anon script v01"
```

This is the site-wide anonymization script applied to all incoming and archiving DICOM resources. This script can also be supplemented by anonymization operations specified at the project level. The script must conform to [DicomEdit](#) format.

Note that if the site-wide anonymization is enabled, even with an empty script, it will add a deidentification method status entry to DICOM headers. To allow DICOM files to be imported without any changes, disable site-wide anonymization.

Discard Changes Save

Series Import Filter (Site Wide)

Enable Site-wide Series Import Filter? Disabled

Filter Mode: Blacklist

Creating a **whitelist** means that *only* DICOM series with a series description that matches one of series filter patterns will be considered by XNAT import tools such as the upload assistant. Creating a **blacklist** means that all DICOM series will be considered *except* for series that have one of the specified series filter patterns. A **modality map** lets you specify boolean expressions in JavaScript that can use DICOM header values from incoming DICOM objects to decide the appropriate modality for the destination session.

Edit Series Import Filter

Flywheel's approach: intelligent risk-based anonymization

The screenshot shows the 'PHI Review' interface. On the left, there are summary statistics: 790 Flags, 587 Flagged Tags, 203 Flagged Images, 11,256 Files, 18 Sessions, and 9 Subjects. The 'Time to Complete' is 24 min. The main area displays a table of tags with columns for Tag Name, Count, and Value. The table lists various tags such as 'Record In-use Flag', 'Directory Record Type', and 'Language Code Sequence'. On the right, a 'Rules' panel shows a list of rules that can be applied, including 'Cross subject Analysis', 'Most private tags', 'Any Date', 'Referring Physician', 'Ignore', 'UIDs', and 'Dr Little'. A 'Start PHI Review' button is visible at the bottom right.

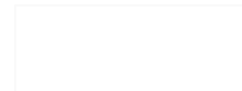
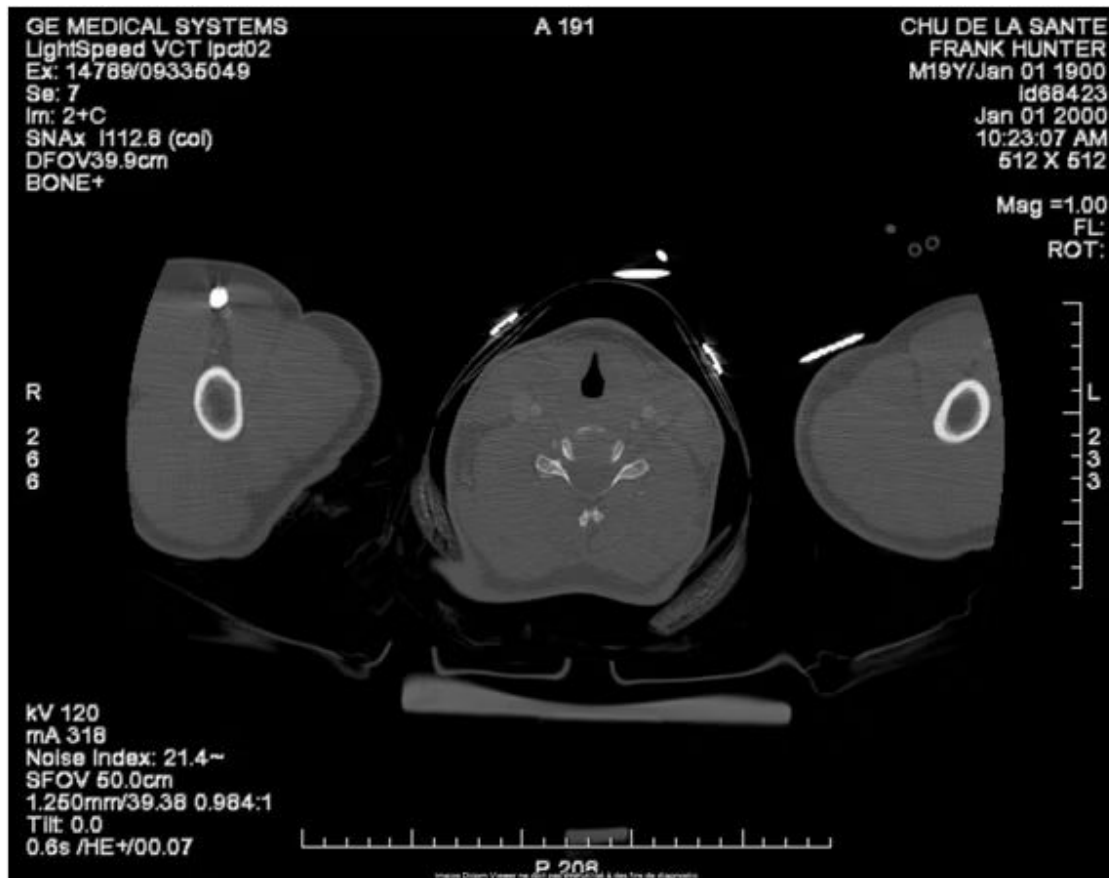
The screenshot shows the 'PHI Review' interface with a grid of medical images. The left sidebar shows the same summary statistics as the previous screenshot. The main area displays a grid of 20 axial MRI brain scans, numbered 1 through 20. The top right corner shows '258/5,238' images. The right side of the interface has a vertical toolbar with various icons for image manipulation and review. A 'Start PHI Review' button is visible at the bottom right.

How to anonymize an enormous number of images:

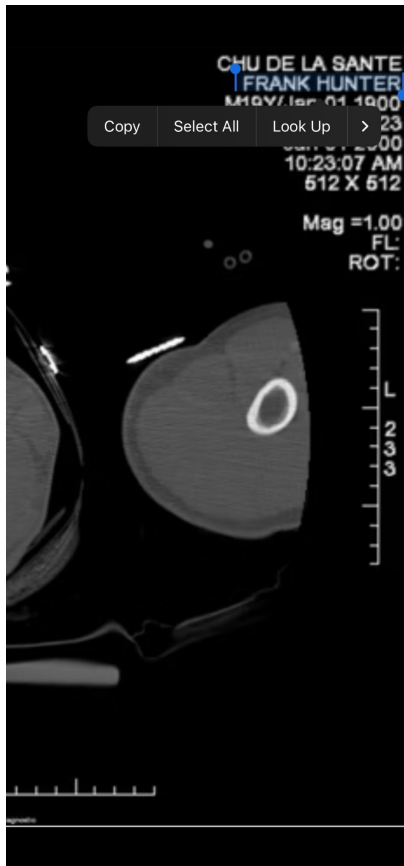
- Well known attributes
- Heuristics (e.g. if an attribute has the same value across a large # of patients... it's not PHI).
- Optical character recognition (a solved problem at this point).
- Natural language processing (maybe solved?)



Flywheel's approach: intelligent risk-based anonymization



Flywheel's approach: intelligent risk-based anonymization



https://chat.openai.com/?model=...
chat.openai.com/?model=text-davinci-002-render-sha

+ New chat

Is Frank Hunter a person's name?

Yes, Frank Hunter is a name that could belong to a person. "Frank" is a common given name, while "Hunter" is a fairly common surname. It's possible for someone to have the full name "Frank Hunter." However, without any specific context or further information, it is challenging to determine if a particular individual named Frank Hunter exists or if you are referring to a specific person.

Regenerate response

Send a message.

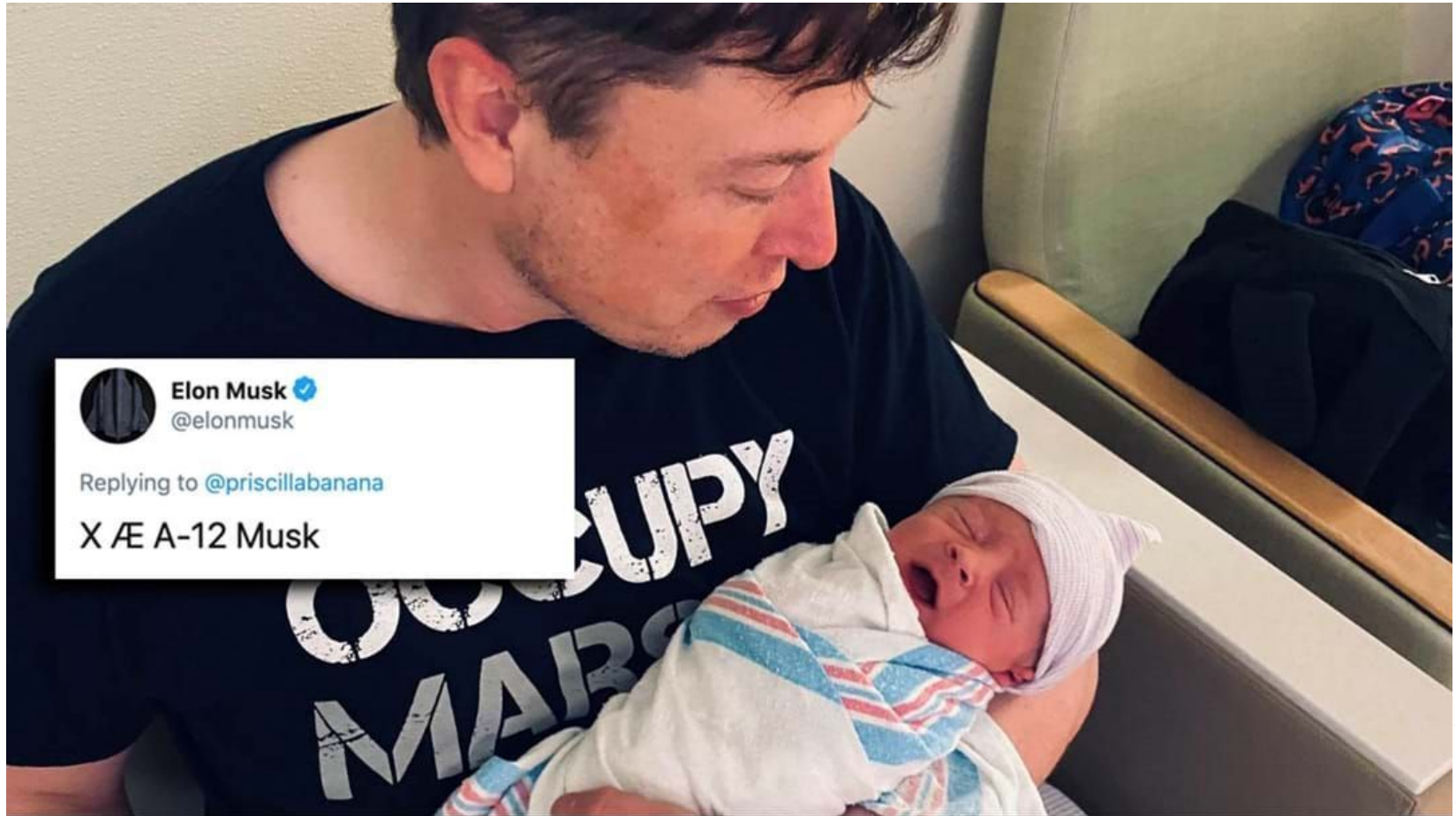
Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 12 Version](#)

Upgrade to Plus NEW

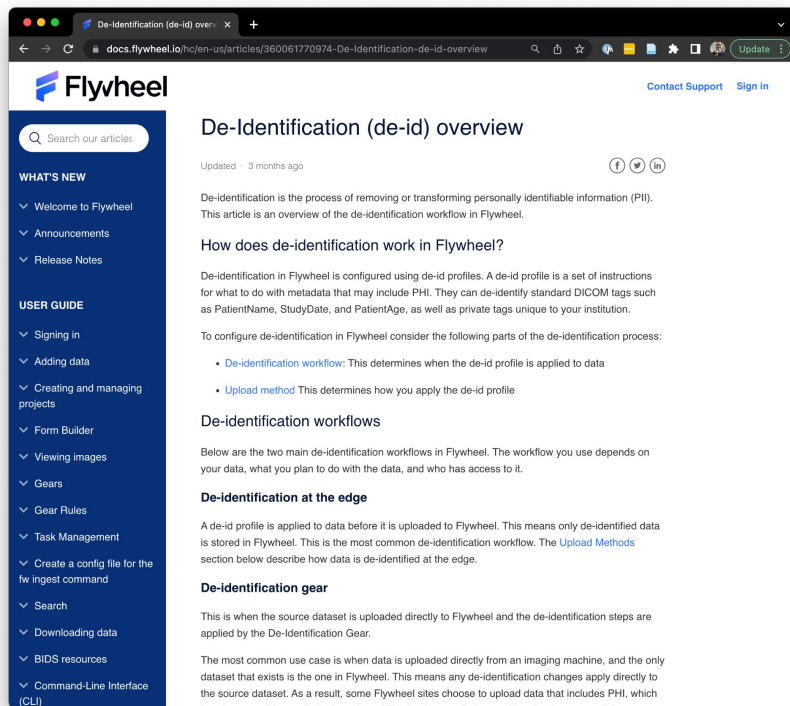
Dan Marcus

This screenshot shows a chat interface on a web browser. The user asks, "Is Frank Hunter a person's name?". The AI responds with a detailed answer: "Yes, Frank Hunter is a name that could belong to a person. 'Frank' is a common given name, while 'Hunter' is a fairly common surname. It's possible for someone to have the full name 'Frank Hunter.' However, without any specific context or further information, it is challenging to determine if a particular individual named Frank Hunter exists or if you are referring to a specific person." The interface includes a "Regenerate response" button, a "Send a message" input field, and a footer with a disclaimer: "Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 12 Version". The user's name "Dan Marcus" is visible at the bottom left.

Flywheel's approach: intelligent risk-based anonymization



Flywheel's approach: intelligent risk-based anonymization



The screenshot shows a web browser displaying the "De-Identification (de-id) overview" article on the Flywheel website. The URL is docs.flywheel.io. The page features a dark blue sidebar with navigation links under "WHAT'S NEW" and "USER GUIDE". The main content area has a white background with a blue header. The article title is "De-Identification (de-id) overview" and it was updated 3 months ago. The text explains that de-identification is the process of removing or transforming personally identifiable information (PII) and describes how it works in Flywheel using de-id profiles. It lists two main workflows: "De-identification at the edge" and "De-identification gear".

De-Identification (de-id) overview

Updated · 3 months ago

De-identification is the process of removing or transforming personally identifiable information (PII). This article is an overview of the de-identification workflow in Flywheel.

How does de-identification work in Flywheel?

De-identification in Flywheel is configured using de-id profiles. A de-id profile is a set of instructions for what to do with metadata that may include PHI. They can de-identify standard DICOM tags such as PatientName, StudyDate, and PatientAge, as well as private tags unique to your institution.

To configure de-identification in Flywheel consider the following parts of the de-identification process:

- **De-identification workflow:** This determines when the de-id profile is applied to data
- **Upload method** This determines how you apply the de-id profile

De-identification workflows

Below are the two main de-identification workflows in Flywheel. The workflow you use depends on your data, what you plan to do with the data, and who has access to it.

De-identification at the edge

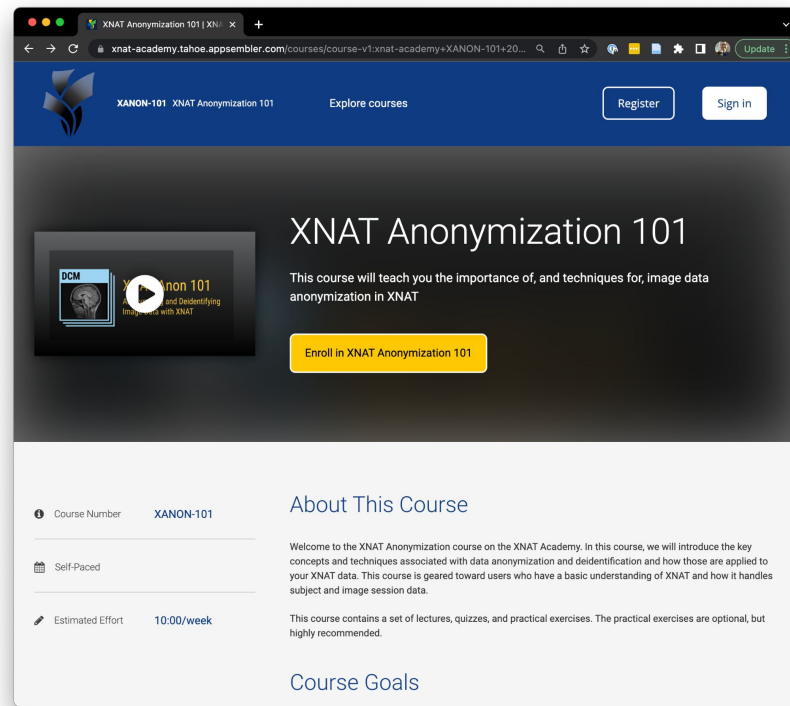
A de-id profile is applied to data before it is uploaded to Flywheel. This means only de-identified data is stored in Flywheel. This is the most common de-identification workflow. The [Upload Methods](#) section below describe how data is de-identified at the edge.

De-identification gear

This is when the source dataset is uploaded directly to Flywheel and the de-identification steps are applied by the De-Identification Gear.

The most common use case is when data is uploaded directly from an imaging machine, and the only dataset that exists is the one in Flywheel. This means any de-identification changes apply directly to the source dataset. As a result, some Flywheel sites choose to upload data that includes PHI, which is only accessible to a limited number of Flywheel users. If you prefer the data accessible to users

Flywheel Enterprise De-Identification Overview



The screenshot shows a web browser displaying the "XNAT Anonymization 101" course page on the XNAT Academy website. The URL is xnac-academy.tahoe.appsembler.com. The page features a dark blue header with the course title and navigation links. The main content area has a white background with a blue header. The course title is "XNAT Anonymization 101" and it includes a video thumbnail. The text describes the course content and provides a yellow "Enroll in XNAT Anonymization 101" button. Below the main content, there is a section titled "About This Course" with details about the course number, self-paced nature, and estimated effort.

XNAT Anonymization 101

This course will teach you the importance of, and techniques for, image data anonymization in XNAT

[Enroll in XNAT Anonymization 101](#)

About This Course

Course Number	XANON-101
Self-Paced	
Estimated Effort	10:00/week

Welcome to the XNAT Anonymization course on the XNAT Academy. In this course, we will introduce the key concepts and techniques associated with data anonymization and deidentification and how those are applied to your XNAT data. This course is geared toward users who have a basic understanding of XNAT and how it handles subject and image session data.

This course contains a set of lectures, quizzes, and practical exercises. The practical exercises are optional, but highly recommended.

Course Goals

XNAT De-Identification Overview