Triennial Report Generator

Select inputs, then filter by Field to generate a publication-ready DOCX report.

Inputs

Excel (app/local path)

/app/python/source_code/Triennial Data Source_Master File of All Submissio

Style prompt JSON (app/local path)

/app/python/source_code/style_prompt.json

Reference DOCX (app/local path)

/app/python/source_code/reference.docx

Lua pagebreak filter (app/local path, optional)

/app/python/source_code/h2_pagebreak.lua

Lua square-bracket filter (app/local path, optional)

/app/python/source_code/h2_square_bracket_footnotes.lua

☐ Publish a copy to DBFS FileStore (optional)

Working output dir (DBFS path)

dbfs:/FileStore/triennial/out

Final output volume dir (optional)

/Volumes/dpcpsi/gold/triennial_reports

☐ Also copy final DOCX into Volume directory

Tip: Put the input files next to app.py (or under ./assets/) and keep these as relative names (e.g., 'style_prompt.json').

Load Inputs

Field

Cancer

Preflight (review preview before generating)

Optional: style / tone / structure requests (applied to this run)

Example: use shorter sentences; avoid jargon; emphasize collaboration and translational impact.

☐ Also update style_prompt.json in DBFS using the override (optional)

Optional: DOCX primary text color (formatting) — e.g., blue or #0000ff

Build Plan

Preview: 58 filtered rows, 58 unique UIDs for Field='Cancer'.

First UIDs (up to 30):

```
2_CC, 5_CC, 34_CC, 70_ONR, 109_NIAMS, 110_NIAMS, 170_ORWH, 206_NCATS, 211_NCATS, 212_NCATS, 244_NCI, 245_NCI, 246_NCI, 248_N
```

Preview (LLM-generated)

Here is a short, clear preview in numbered steps:

1. Filter the data to prepare it for the report.
2. Organize the data into cards and create an index of unique identifiers.
3. Create paragraphs for each row of data, including a unique identifier at the end of each paragraph.
4. Categorize the data into sections based on the unique identifiers.
5. Generate a summary of the report.
6. Create an introduction to the report.
7. Synthesize the information in each section.
8. Assemble the report content.
9. Replace unique identifier markers with numeric footnotes.
10. Finalize the report document.
11. Optionally, apply a primary color override to the report.
12. Publish the report and make it available for download.

sections that will be included

Section	Activities (UIDs)
Advanced Imaging & AI Tools	34
Combination & Targeted Therapies	1
Data Commons and Computational Resources	1
Epidemiology & Surveillance	5
Genetics, Cell Biology, and -Omics	7
Immunotherapy	2
Nutrition & Symptom Management	2
Preventive Interventions	1
Screening & Early Detection	4
Tumor Microenvironment & Immunology	1

why each UID is in its section

UID	Selected Section	Rationale (LLM)
245_NCI	Preventive Interventions	The activity focuses on lifestyle interventions, such as physical activity a
246_NCI	Advanced Imaging & AI Tools	Selected based on dominant topical/method cues in the activity title/des
248_NCI	Advanced Imaging & AI Tools	Selected based on dominant topical/method cues in the activity title/des
249_NCI	Advanced Imaging & AI Tools	The activity description mentions the use of national trend data to illustr
250_NCI	Genetics, Cell Biology, and -Omics	The activity description mentions cell division, cell preparation to divide
251_NCI	Screening & Early Detection	The activity is focused on improving cervical cancer screening coverage t
252_NCI	Advanced Imaging & AI Tools	Selected based on dominant topical/method cues in the activity title/des
253_NCI	Advanced Imaging & AI Tools	Selected based on dominant topical/method cues in the activity title/des
254_NCI	Genetics, Cell Biology, and -Omics	The activity description mentions 'differences in serum metabolites relat
255_NCI	Advanced Imaging & AI Tools	Selected based on dominant topical/method cues in the activity title/des

Excluded candidate sections (per UID):

- **2_CC** (selected: Advanced Imaging & AI Tools):
- Epidemiology & Surveillance: While the study involves diagnostic performance, which could be related to epidemiology, the primary focus on imaging modalities (PET/CT and MRI) makes Epidemiology & Surveillance a less direct fit compared to Advanced Imaging & AI Tools.
- **5_CC** (selected: Epidemiology & Surveillance):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a less relevant fit.
- Environmental Health and Cancer: While the activity is related to cancer, there is no explicit mention of environmental health factors, which are a key aspect of this section.
- **34_CC** (selected: Nutrition & Symptom Management):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a poor fit.
- Recalcitrant & Hard-to-Treat Cancer Research: While Hereditary Diffuse Gastric Cancer is a rare and potentially challenging condition, the activity's focus on nutrition and symptom management rather than research into the cancer itself makes this section less relevant.
- **70_ONR** (selected: Nutrition & Symptom Management):
- Advanced Imaging & AI Tools: This section is not relevant as the activity description does not mention imaging or AI tools.
- Preventive Interventions: While the activity does discuss prevention of malnutrition, the primary focus is on nutrition and symptom management, making this section a secondary fit.
- Tumor Microenvironment & Immunology: The activity description does not discuss the tumor microenvironment or immunology, making this section unrelated to the activity's focus on nutrition and symptom management.
- **109_NIAMS** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a poor fit.
- Nutrition & Symptom Management: While muscle wasting is a symptom of cancer, the activity's focus on molecular mechanisms and cellular responses makes Genetics, Cell Biology, and -Omics a more relevant section.
- Tumor Microenvironment & Immunology: The activity description does not explicitly discuss the tumor microenvironment or immunology, making this section less relevant than Genetics, Cell Biology, and -Omics.
- **170_ORWH** (selected: Epidemiology & Surveillance):
- Advanced Imaging & AI Tools: This section is not a good fit because the activity description does not mention imaging or AI tools, instead focusing on collaborative efforts to address cervical cancer disparities.
- **212_NCATS** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a poor fit.
- Preventive Interventions: While the activity may ultimately lead to the development of new treatments, the focus is on understanding the genetic basis of liver cancer and developing targeted therapies, rather than preventive interventions.
- Screening & Early Detection: The activity description does not discuss screening or early detection methods, but rather the potential for developing new treatments based on genetic insights.
- **245_NCI** (selected: Preventive Interventions):
- Advanced Imaging & AI Tools: This section is not relevant because the activity does not mention imaging or AI tools.
- Epidemiology & Surveillance: While the activity involves studying the relationship between physical activity and cancer mortality, its primary focus is on preventive interventions rather than epidemiological surveillance.
- **249_NCI** (selected: Advanced Imaging & AI Tools):

- Tumor Microenvironment & Immunology: While the report touches on cancer treatment, it does not specifically focus on the tumor microenvironment or immunology, making this section a less suitable match.
- **250_NCI** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section less relevant.
- Immunotherapy: Although cancer is mentioned, the activity description does not discuss immunotherapy or the immune system, making this section a less suitable fit.
- Tumor Microenvironment & Immunology: While the activity mentions cancer cells, the primary focus is on cell division and its reversal, rather than the tumc microenvironment or immunology, making this section secondary to Genetics, Cell Biology, and -Omics.
- **251_NCI** (selected: Screening & Early Detection):
- Advanced Imaging & AI Tools: This section is not relevant because the activity does not mention imaging or AI tools, but rather a self-collection method for HPV testing.
- Preventive Interventions: While the activity does involve prevention of cervical cancer, the primary focus is on screening and early detection, making Screening & Early Detection a more direct fit.
- **254_NCI** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section less relevant.
- Epidemiology & Surveillance: While the study does involve breast cancer survivors and compares them to the general population, the primary focus is on th biological and genetic aspects of the disease, rather than epidemiological patterns or surveillance.
- **256_NCI** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: This section is not a good fit because the activity text does not mention imaging or AI tools, but rather focuses on molecular biology techniques such as microRNA analysis.
- **258_NCI** (selected: Advanced Imaging & AI Tools):
- Combination & Targeted Therapies: While CYTALUX is a targeted agent, its primary purpose is imaging, not therapy, making Advanced Imaging & AI Tools a better fit.
- Genetics, Cell Biology, and -Omics: There is no mention of genetic, cellular, or omics-related research in the activity description, making this section less relevant.
- **259_NCI** (selected: Screening & Early Detection):
- Advanced Imaging & AI Tools: There is no mention of advanced imaging or AI tools in the activity text, making this section a poor fit.
- Preventive Interventions: While the activity does involve preventive measures, the primary focus is on screening and early detection, making Screening & Early Detection a more specific and relevant section.
- **260_NCI** (selected: Epidemiology & Surveillance):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity text, making this section a poor fit.
- Genetics, Cell Biology, and -Omics: Although the study involves antibodies and stomach cells, the primary focus is on epidemiological trends rather than genetic or cellular mechanisms.
- Immunotherapy: The activity text discusses autoimmune gastritis, but the context is epidemiological rather than therapeutic, and there is no mention of immunotherapy approaches.
- **262_NCI** (selected: Immunotherapy):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity text, making this section a poor fit.

- Genetics, Cell Biology, and -Omics: While the activity involves cell biology, the primary focus is on immunotherapy and engineering immune cells, rather tha genetic or omics research.
- Preventive Interventions: The activity is focused on treating existing tumors, rather than preventing cancer, which is the primary focus of preventive interventions.
- **272_NCI** (selected: Advanced Imaging & AI Tools):
- Tumor Microenvironment & Immunology: While the study is related to cancer care, it does not specifically focus on the tumor microenvironment or immunology, making Advanced Imaging & AI Tools a more direct match.
- **273_NCI** (selected: Epidemiology & Surveillance):
- Advanced Imaging & AI Tools: This section is not relevant as the activity does not mention imaging techniques or AI tools.
- Environmental Health and Cancer: While cancer is a key aspect, the activity's focus is on the epidemiology of meningioma risk after childhood cancer treatment, rather than environmental health factors.
- Tumor Microenvironment & Immunology: The activity does not delve into the tumor microenvironment or immunological aspects, making this section a les suitable fit.
- **274_NCI** (selected: Epidemiology & Surveillance):
- Advanced Imaging & AI Tools: This section is not relevant because the activity does not mention imaging or AI tools, but rather focuses on population-level estimates and trends.
- Tumor Microenvironment & Immunology: This section is not the best fit because the activity does not delve into the biological or immunological aspects of cancer, but rather concentrates on the epidemiological aspects of cancer survivorship.
- **275_NCI** (selected: Combination & Targeted Therapies):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a poor fit.
- Genetics, Cell Biology, and -Omics: While the activity description does mention genetic mutations and cell biology, the primary focus is on the development targeted therapies, making this section a secondary match.
- Preventive Interventions: The activity description focuses on treating existing cancer through targeted therapies, rather than preventing cancer, making this section an unlikely match.
- **280_NCI** (selected: Immunotherapy):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section less relevant.
- Genetics, Cell Biology, and -Omics: Although the activity involves genetic mutations and neoantigens, the primary focus is on immunotherapy, not genetics cell biology.
- Preventive Interventions: The activity is focused on treating existing cancer, rather than preventing it, making this section a less suitable fit.
- **319_NIDCR** (selected: Screening & Early Detection):
- Advanced Imaging & AI Tools: While the activity mentions 'applying molecular, cellular, and multi-omics signatures', it does not specifically focus on imaging or AI tools, making this section a less relevant match.
- Genetics, Cell Biology, and -Omics: Although the activity involves 'multi-omics signatures', the primary focus is on early detection, which is more closely aligned with the Screening & Early Detection section. Genetics, Cell Biology, and -Omics is a related but secondary fit.
- **332_NIDCR** (selected: Screening & Early Detection):
- Advanced Imaging & AI Tools: Although the activity involves imaging technology, its primary purpose is to enhance screening and early detection, rather tha solely developing advanced imaging tools.
- Preventive Interventions: The activity is more focused on detection than prevention, as it aims to identify oral cancers at an early stage rather than preventi their occurrence.

- **495_NIEHS** (selected: Genetics, Cell Biology, and -Omics):
- Immunotherapy: While immunotherapy may be related to cancer treatment, the activity description does not explicitly mention immune system modulatio or therapeutic interventions, making Genetics, Cell Biology, and -Omics a more direct fit.
- **622_OSC** (selected: Genetics, Cell Biology, and -Omics):
- Advanced Imaging & AI Tools: Although the activity mentions the use of computerized tomography (CT) scans and magnetic resonance imaging (MRIs), as w as artificial intelligence, the primary focus is on the metabolites and their role in disease diagnosis, rather than the imaging or AI tools themselves.
- **642_OSC** (selected: Data Commons and Computational Resources):
- Advanced Imaging & AI Tools: There is no mention of imaging or AI tools in the activity description, making this section a less relevant fit.
- Genetics, Cell Biology, and -Omics: Although the activity involves genetic pathways and childhood cancer research, the primary focus is on the data portal a its resources, rather than the specific biological or genetic aspects, making Data Commons a more direct match.
- **933_NIBIB** (selected: Advanced Imaging & AI Tools):
- Data Commons and Computational Resources: While the activity does involve the development of a repository for medical images, its primary focus on medical imaging and AI tools makes Advanced Imaging & AI Tools a better fit. Data Commons and Computational Resources is a secondary match due to the repository aspect, but it is not the primary emphasis of the activity.
- **944_NIBIB** (selected: Tumor Microenvironment & Immunology):
- Advanced Imaging & AI Tools: Although the activity mentions the use of focused ultrasound, which could be considered an advanced imaging tool, the primary focus is on brain tumor diagnosis and the tumor microenvironment, rather than the development of new imaging technologies.
- Combination & Targeted Therapies: The activity does not mention combination or targeted therapies, but rather a diagnostic approach, making this section less relevant fit.

☐ Yes — confirm the sections above will be included in the report

☐ Yes — proceed to generate using this plan

Generate Report

Live narration updates here (replaces in place).