

The Astronaut and Spaceflight Participant Archive (ASPA)

Technical Documentation

Version 1.2

Last Updated: 21 October 2025

Lakshmi Kanikkannan, MS

Katie Hogan, MD PhD

Robert Reynolds, MS MPH PhD PStat®

Database and documentation hosted at:

<https://github.com/osmed-admin/aspa>

Table of Contents

Table of Contents	2
Introduction	3
Components and Features of the Database	3
1. Tables	3
2. Relationships	4
3. Data Types	4
4. Logical Design	5
5. Data Integrity and Standards	5
6. Access and Querying	6
Data Dictionary	7
1. Primary Tables	7
tblPeople	7
tblMissions	8
tblOrganizations	8
tblProgram	9
tblCraft	9
2. Lookup Tables	10
LU_COD	10
LU_CraftType	10
LU_Education	10
LU_MilitaryBranch	11
LU_Rank	11
LU_Race	12
3. Junction Tables	14
jnc_PeopleOrg	14
jnc_OrgProg	14
jnc_CraftProgCraftType	14
Cardinality Statements	15
Relationships	16
Sample Code	17
Logic Checks	18
Future Release Updates	18
Maintenance and Guidelines	19
Sources	19

Introduction

The Astronaut and Spaceflight Participant Archive (ASPA) is a database designed to consolidate biographical, demographic, and vital statistics data for both international governmental astronauts and commercial spaceflight participants (CSPs). Its primary goal is to address gaps in accessible data and provide researchers with a unified resource to contextualize biomedical findings and support spaceflight health and performance studies.

ASPA brings together data from international governmental space programs—including data on astronauts from the United States, Canada, Soviet Union and Russian Federation, European Union, Japan, and China, as well as participants from commercial spaceflight companies. The database hopes to serve as a complement to existing resources by offering demographic and career history information not easily found elsewhere.

This resource has been structured to support a range of research efforts, including but not limited to:

- Epidemiological studies on astronaut health and mortality.
- Assessments of health risks and countermeasures for long-duration missions.
- Analyses of demographic trends among space travelers, including comparisons between governmental astronauts and CSPs.

By centralizing these data, ASPA seeks to reduce barriers to access for critical information and enable high-quality, multidisciplinary research on the long-term health effects of space exploration.

Components and Features of the Database

The ASPA database consists of several core components, designed to organize and interconnect data on spaceflight participants. These components are available as one .db file and several flat Excel files. These components include:

1. Tables

The database is organized into a set of relational tables, including primary tables, lookup tables, and junction tables. In line with standard relational database design:

- **Primary Tables** store the core records, each with a unique identifier.
- **Lookup Tables** contain standardized code values used to label or categorize entries in the primary tables.
- **Junction Tables** manage many-to-many relationships by linking primary keys from Primary and Lookup tables.

Table 1 below lists the current set of tables in the ASPA and a brief description of their contents.

Table 1. Current tables in the ASPA and their descriptions.

Table Name	Description
tblMissions	Documents information about space missions, including mission names, associated spacecraft, and dates.
tblOrganizations	Lists the organizations involved in spaceflight activities, including governmental and private entities.
tblCraft	Details spacecraft used in missions, including their associated programs and capsule types.
LU_CraftType	Defines standardized spacecraft type classifications.
LU_Race	Stores standardized race categories.
LU_Rank	Store predefined military ranks for various branches of service.
jnc_CraftProgCraftType	Junction table connecting spacecraft, space programs, and craft types.
jnc_OrgProg	Junction table linking organizations to their associated programs.

2. Relationships

The database follows relational design principles to establish clear, structured relationships between tables:

- **Primary Keys:** Unique identifiers for each record (e.g., PersonID in tblPeople), ensuring each row can be distinctly referenced.
- **Foreign Keys:** Fields that reference primary keys in other tables, enabling links between related records.
- **Many-to-Many Relationships:** Represented using junction tables—for example, linking astronauts to multiple organizations or associating spacecraft with multiple programs.

3. Data Types

Each table defines specific data types for its fields, selected to support accurate querying, validation, and integration:

- **Identifiers / Boolean Variables:** Typically stored as integers (e.g., MissionID, PersonID).
- **Strings:** Used for names, descriptions, and nominal categorical values (e.g., ProgramName, CraftTypeName).
- **Dates:** Applied to mission schedules and other time-based fields (e.g., MissionDate).
- **Numeric Values:** Used for durations, counts, or other quantitative measurements.

Universal Coding:

- **Dates:**
 - Unknown/Missing = 99/99/9999
 - If Year is known but not month or day: 07/02/YYYY
 - If Year and Month are known but not day: MM/15/YYYY
 - Not applicable = 88/88/8888
- **Integers:**
 - Unknown/Missing = -9
 - Not Applicable = -8
- **Strings:** filled in with “Unknown” or “NA” as needed
- **Boolean values in numeric fields:** 1 if TRUE/YES, 0 if FALSE/NO (examples include Pilot and MedicalDoctor in tblPeople)
 - Unknown/Missing = -9
 - Not Applicable = -8

4. Logical Design

The database is structured for flexibility and scalability, supporting future growth and minimizing redundancy:

- **Normalization:** Tables are normalized to reduce data duplication and maintain consistency across records.
- **Referential Integrity:** Enforced through primary and foreign key constraints to ensure valid, consistent relationships between entities.

5. Data Integrity and Standards

The database employs multiple strategies to ensure data accuracy and consistency:

- **Standardized Coding:** Lookup tables enforce consistent coding schemes for categorical variables (e.g., race, rank, military branch).
- **Referential Integrity:** All foreign keys are constrained to valid entries in their corresponding primary tables, preventing orphaned or invalid references.
- **Data Type Enforcement:** Ensures each field adheres to its defined type (e.g., integer, date, string).

- **Domain Constraints:** Limits the permissible values to those described in this documentation (e.g., values contained in lookup tables) to prevent input errors.
- **Unique Constraints:** Guarantees that fields intended to be unique (e.g., primary keys) do not allow duplicates.

6. Access and Querying

The database can be accessed programmatically or via direct SQL queries to support a variety of use cases, including:

- **Retrieving participant demographics** (e.g., age, race, rank, military affiliation).
- **Analyzing mission details** (e.g., dates, duration, associated spacecraft).
- **Cross-referencing spacecraft, programs, and organizations** to study program affiliations and asset deployment.

Data Dictionary

1. Primary Tables

tblPeople

Column Name	Data Type	Description
PersonID	Integer	Primary key. A unique numeric identifier for each person.
PersonLName	String	Indicates the person's last name.
PersonFName	String	Indicates the person's first name.
PersonMI	String	Indicates the person's middle initial, if applicable.
Title	String	Indicates the person's suffix, if applicable. (Ex: Jr., II)
Race	Integer	Indicates the person's race. Please reference LU_Race for coding.
Ethnicity	Integer	Indicates the person's ethnicity if Race is coded 0 (White) or 1 (Black). Coding: (1=Hispanic/Latino; 0=Not Hispanic/Latino; -8=NA)
Sex	Integer	Indicates the person's sex (1 for male, 0 for female).
Military	Integer	Indicates if the person has served in the military (1 if yes, 0 if no).
Branch	Integer	Indicates which military branch the person was affiliated with. Please reference LU_MilitaryBranch for the coding.
Rank	Integer	Indicates which military rank the person held. Please reference LU_Rank for coding.
Pilot	Integer	Indicates if the person has ever been a pilot; includes obtaining a pilot's license for any reason (1 if yes, 0 if no).
NationalityBirth	String	Indicates the person's primary nationality, usually their country of birth or their country of citizenship, if not the same.
Nationality2	String	Indicates the person's secondary nationality, if applicable.
Nationality3	String	Indicates the person's tertiary nationality, if applicable.
HighestEd	String	Indicates the highest degree the person has earned.
Educat	Integer	Indicates the highest level of education the person has reached. Please reference LU_Education for coding.
MedicalDoctor	Integer	Indicates if the person is a medical doctor, either MD or DO (1 if yes, 0 if no).
DOB		Indicates the person's date of birth (MM/DD/YYYY)
Died	Integer	Indicates the person's vital status (0=alive; 1=dead).
DOD		Indicates the person's date of death, if deceased.
COD	String	Indicates the person's cause of death, if deceased. (Data were sourced from media reports, so not present for all deaths.) None Not dead Exact unknown We have some info, but exact COD unknown Unknown COD completely unknown
COD_Group	Integer	Indicates the Cause of Death grouping for natural causes.
COD_Group2	Integer	Indicates the Cause of Death grouping for accidents.

tblMissions

Column Name	Data Type	Description
MissionID	Integer	Primary Key. Unique identifier for each mission record in the table.
CraftID	Integer	Indicates the unique database identifier for the spacecraft used on the mission.
CallSign	String	Call sign associated with the craft, if any.
MissionName	String	Indicates the name of each mission.
ProgramID	Integer	Indicates the unique database identifier of the program that the mission to which the mission belongs.
MissionStart		Indicates the date that the mission began.
MissionEnd		Indicates the date that the mission ended.
MissionType	String	Denotes if the mission reached suborbital levels, low earth orbit (LEO), or the lunar surface.
Inclination	Double	The angle of tilt of a spacecraft's orbit around the Earth
Altitude	Integer	Altitude of the orbit above the surface of the Earth
SpaceLimitLine	String	Denotes the maximum border of space the craft in the mission reached: Karman line (internationally recognized boundary of space, 100km) or Edge of Space (80km)

tblOrganizations

- A table tracking organizations with the intent, scope, and history to go to space.

Column Name	Data Type	Description
OrganizationID	Integer	Primary Key. Unique identifier for each organization. 1 US Air Force 2 NASA 3 Roscosmos 4 ESA 5 JAXA 6 NASDA 7 CSA 8 CNSA 9 USSR 10 Axiom Space 11 Space Adventures 12 Virgin Galactic 13 Blue Origin 14 SpaceX
OrganizationName	String	Full name of the organization (e.g., NASA, US Air Force).
Country	String	Country that the organization is based in.

tblProgram

Column Name	Data Type	Description
ProgramID	Integer	Primary Key . Unique identifier for the space program. This is used as a label in several places, notably: tblMissions, tblCraft, jnc_OrgProg, and jnc_CraftProgCraftType.
ProgramName	String	Name of the program.

tblCraft

Column Name	Data Type	Description
CraftID	Integer	Primary Key . Unique identifier for a specific craft.
CraftName	String	Name of the spacecraft.
CraftTypeID	Integer	Foreign Key . Encoded type of spacecraft. References LU_CraftType.
ProgramID	Integer	Program under which the craft was used. References tblProgram.

2. Lookup Tables

LU_COD

- Table used to code deaths at various levels
- Referenced in tblPeople.

Column Name	Data Type	Description
COD_Group	Integer	Categorizations for causes of death
COD_Description	String	Categorization for the natural causes of death. -8 NA (An accidental/unnatural cause of death) 0 Not dead 1 Cancer 2 Cardiovascular 3 External causes 4 Other natural
COD2_Description	String	Categorization for accidental/unnatural causes of death. -8 NA (A natural cause of death) 0 Not dead 1 Plane crash 2 Spacecraft accident 3 Training accident 4 Other accident 5 Non-accidental external

LU_CraftType

Column Name	Data Type	Description
CraftTypeID	Integer	Primary Key. Encoded type of spacecraft.
CraftType	String	Name of the craft type.

LU_Education

- Table used to code highest level of education
- Referenced in tblPeople

Column Name	Data Type	Description
Educat	Integer	Encoded education level: 0 Less than Bachelor 1 Bachelor 2 Master 3 Doctorate
Degree	String	Categorization for the integer levels of education (e.g., 0 = Less than Bachelor, 1 = Bachelor).

LU_MilitaryBranch

- Table used to code military service affiliations
- Referenced in tblPeople.
- Civilians are coded “-8” (NA) as they have no branch of service

Column Name	Data Type	Description
BranchID	Integer	Unique identifier for military branch: -8 N/A (Civilian) 1 US Air Force 2 US Army 3 US Coast Guard 4 US Navy 5 US Marines 6 U.K. Royal Air Force 7 French Air and Space Force 8 Royal Saudi Air Force
BranchName	String	Names of the branches of the military.

LU_Rank

- Referenced in tblPeople.
- Civilians are coded “-8” for all branches of service as they have no rank.

Column Name	Data Type	Description
RankID	Integer	Unique identifier for a rank category.
ArmyRank	String	Names of the ranks of the US Army, US Air Force, and the US Marine Corps: 0 Enlisted 1 2nd Lieutenant 2 1st Lieutenant 3 Captain 4 Major 5 Lt. Colonel 6 Colonel 7 Brigadier General 8 Major General 9 Lieutenant General 10 General
NavyRank	String	Names of the ranks of the US Navy and US Coast Guard: 0 Enlisted 1 Ensign 2 Lieutenant Junior Grade 3 Lieutenant 4 Lt. Commander 5 Commander 6 Captain 7 Rear Admiral (lower half)

		8 Rear Admiral (upper half) 9 Vice Admiral 10 Admiral Chief of Naval Op
RussianAirForce	String	Names of the ranks of the Russian Air Force: 0 Jr. Lieutenant 1 Lieutenant 2 Senior Lieutenant 3 Captain 4 Major 5 Lt. Colonel 6 Colonel 7 Major General 8 Lieutenant General 9 Colonel General 10 General of the Army

LU_Race

- Referenced in tblPeople.

Column Name	Data Type	Description
RaceID	Integer	Primary Key. Unique identifier for a race category: 0 White 1 Black 2 Arab 3 Asian 4 Native American 5 South Asian 6 Persian 7 Turkish
RaceText	String	Description of race.

- Legacy race coding

The following table indicates how legacy Race coding categories (used in prior publications) were migrated to the current coding (which is new to the original release of ASPA).

Original Race Coding	New Race Coding	Notes
0 White	0 White	
1 Black	1 Black	
2 Hispanic	[NONE]	Now reflected by ethnicity variable, for White and Black only
3 Asian	3 Asian	Now inclusive of China, Japan, Korea, Philippines, Indonesia, Vietnam, and Kazakhstan.

4 Native American	4 Native American	
5 East Indian	5 South Asian	Now inclusive of India and Pakistan.

(Note: this table only specifies how prior codes were transformed, and does not include all the codes in the current version. See the LU_Race table above for full current coding.)

3. Junction Tables

jnc_PeopleOrg

- Joins potentially multiple organizations to people.
- There can be multiple entries per personID if they have affiliations with multiple spaceflight organizations.

Column Name	Data Type	Description
PeopleOrgID	Integer	Primary Key.
PersonID	Integer	Unique numeric identifier for each person.
OrganizationID	Integer	ID linking to the tblOrganizations table.
StartDate	Date	Start Date for each affiliation.
Role	String	Specifies the role that person was fulfilling during their time at the corresponding organization: Trained Astronaut, Payload Specialist, Commercial Civilian, Military, and Private Space Org Employee.
Notes	String	Any notes.
Jacobs_PS	Integer	An indicator value reflecting the payload specialists we believe were used in the Jacobs analysis.

jnc_OrgProg

Column Name	Data Type	Description
OrgProgID	Integer	Primary Key. Unique descriptor ID for each entry.
OrganizationID	Integer	ID linking to the tblOrganizations table.
ProgramID	Integer	ID linking to the tblPrograms table.

jnc_CraftProgCraftType

Column Name	Data Type	Description
CraftDescrID	Integer	Primary Key. Unique descriptor ID for each entry.
CraftID	Integer	Unique identifier for a specific craft. References tblCraft.
ProgramID	Integer	Program under which the craft was used. References tblProgram.

Cardinality Statements

This section details the cardinality relationships between the tables in this database. This adds logical definition to the structure of the database.

tblPeople ↔ jnc_PeopleOrg ↔ tblOrganizations	<ul style="list-style-type: none"> • A person can be affiliated with more than one organization. • An organization can have many people.
tblOrganizations ↔ jnc_OrgProg ↔ tblProgram	<ul style="list-style-type: none"> • An organization can have multiple programs. • A program cannot be run by multiple organizations.
tblCraft ↔ tblProgram	<ul style="list-style-type: none"> • One craft can belong to multiple programs. • Each program does not necessarily include multiple crafts.
tblCraft ↔ LU_CraftType	<ul style="list-style-type: none"> • One craft cannot have more than one type. • One craft type can be used by many crafts.
tblMissions → tblCraft	<ul style="list-style-type: none"> • Multiple missions can use the same craft. • A mission cannot use more than one craft.
tblMissions → tblProgram	<ul style="list-style-type: none"> • Each mission is tied to a single program. • A program can have multiple missions.
tblPeople → LU_Race, LU_COD, LU_Rank, LU_MilitaryBranch, LU_Education	<ul style="list-style-type: none"> • Each person can only be coded with one of each of these in the appropriate fields in tblPeople.
jnc_CraftProgCraftType ↔ tblCraft, LU_CraftType, tblProgram	<ul style="list-style-type: none"> • Each row represents a single craft instance, with its corresponding craft type and program details.

Relationships

This section provides a schema diagram for the database:

tblPeople		
Primary Key	PersonID	INTEGER
	PersonLName	TEXT
	PersonFName	TEXT
	PersonMI	TEXT
	Title	TEXT
	Race	INTEGER
	Ethnicity	INTEGER
	Sex	INTEGER
	Military	INTEGER
	Branch	INTEGER
	Rank	INTEGER
	Pilot	INTEGER
	PrimaryNationality_BirthCitizen	TEXT
	SecondNationality	TEXT
	ThirdNationality	TEXT
	HighestEd	TEXT
	Educat	INTEGER
	MedicalDoctor	INTEGER
	DOB	
	Died	INTEGER
	DOD	
	COD	
	COD_Group	INTEGER
	COD_Group2	INTEGER

tblProgram		
Primary Key	ProgramID	INTEGER
	ProgramName	TEXT

tblOrganizations		
Primary Key	OrganizationID	INTEGER
	OrganizationName	TEXT

tblMissions		
Primary Key	MissionID	INTEGER
	CraftID	INTEGER
	MissionName	TEXT
	CallSign	TEXT
	ProgramID	INTEGER
	MissionStart	
	MissionEnd	
	MissionType	TEXT
	Inclination	INTEGER
	Altitude	INTEGER
	SpaceLimitLine	TEXT
	Notes	TEXT

tblCraft		
Primary Key	CraftID	INTEGER
	CraftName	TEXT
	CraftTypeID	INTEGER
	ProgramID	INTEGER

LU_Rank		
Primary Key	RankID	
	ArmyRank	
	NavyRank	
	RussianAirForce	

LU_Race		
Primary Key	RaceID	
	RaceText	

LU_Military Branch		
Primary Key	BranchID	type
	BranchName	

LU_CraftType		
Primary Key	CraftTypeID	
	CraftType	

LU_Education		
	Educat	
	Degree	

LU_COD		
	COD_Group	
	COD_Description	
	COD2_Description	

jnc_PeopleOrg		
Primary Key	PersonID	
	PersonLName	
	Country	
	OrganizationID	
	StartDate	

jnc_OrgProg		
	OrganizationID	
	Organizatoion	
	ProgramID	
	Program	

jnc_CraftProgCraftType		
	CraftDescrID	
	CraftID	
	CraftName	
	CraftTypeID	
	CraftType	
	ProgramID	
	Program	

Sample Code

This section provides sample SQL queries to assemble commonly used datasets from source tables in the database.

1. Code to assemble a cohort dataset of NASA astronauts:

```
SELECT * FROM tblPeople  
INNER JOIN jnc_PeopleOrg ON tblPeople.PersonID = jnc_PeopleOrg.PersonID  
WHERE jnc_PeopleOrg.OrganizationID = 2;
```

Logic Checks

Military → Branch → Rank	When Military is 0, Branch and Rank should both be -8.
Died → DOD → COD → CODGroup → COD_Group2	When Died is 0, DOD should be 88/88/8888, COD could be None, and COD_Group and COD_Group2 should both be 0.
COD_Group → COD_Group2	When COD_Group is 3, COD_Group2 should be a number between 1-5. There shouldn't be anything in COD_Group2 when COD_Group is 1,2,4, or 5.
Race → Ethnicity	When Race is 0 or 1, Ethnicity can be either 0 or 1; when Race takes any other value, Ethnicity must be -8

Future Release Updates

- A table for each mission segment (dates, times, vehicles, etc.)
- Additional details on international astronauts and CSPs
- A junction table joining missions and astronauts
- Additional tables incorporating known historical medical data

Maintenance and Guidelines

This database is currently maintained by the authors, but we welcome contributions via pull requests.

If you would like to suggest changes or additions, please fork the repository, make your edits, and submit a pull request. To ensure your contribution can be considered for inclusion, please provide a clear description of the changes, the rationale behind them, and any relevant documentation or references.

This database will soon be maintained by the Organization for Space Medicine, Engineering, and Design ([OSMED.org](https://osmed.org)), a nonprofit dedicated to accelerating human spaceflight readiness.

We invite you to join OSMED and contribute to this and other open-access projects advancing space medicine, spaceflight engineering, and spaceflight architecture and design. Together, we can build the infrastructure for humanity's next giant leap.

Sources

The following sources were used to collect these data:

- NASA Astronaut Fact Book:
(https://www.nasa.gov/wp-content/uploads/2015/05/astronaut_fact_book_june_2022.pdf?emrc=ce5cb4)
- NASA astronaut biographies:
(<https://www.nasa.gov/headquarters/library/find/bibliographies/astronaut-biographies/>)
- Space Facts: spacefacts.de
- Astronaut biographies and mission profiles on [Wikipedia.org](https://en.wikipedia.org)
- Obituaries from various popular press websites