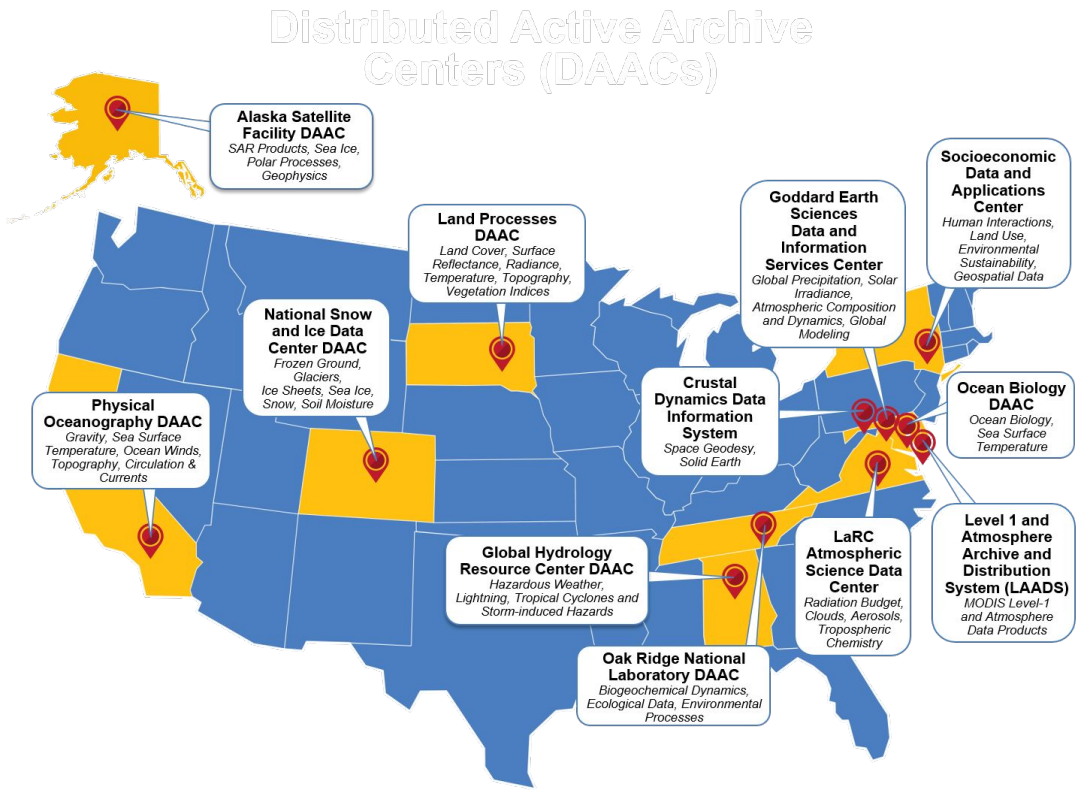
Data Product Strengths & Limitations

Outcomes from the NSIDC DAAC FY21 Data Quality Project

Shannon Leslie, Project Lead

NSIDC DAAC Organization

NASA Snow & Ice Distributed Active Archive Center (DAAC) at the
National Snow and Ice Data Center (NSIDC)



nsidc.org/daac

Project Motivation

2020 DAAC User Working Group (UWG) meeting: Presenting Limitations

- What is DAAC role in documenting data quality/limitations?
- How are data quality results reported back to DAAC and users?

Data quality content curation lifecycle

Data accession & publication

- Data quality is part of DAAC accession evaluation
- User Guides and Mission/Project documents (e.g., ATBDs) document data quality

Data in action

- Scientific research reveals more info about data quality
- Publications (e.g., journal articles) note data product:
  - Use
  - Limitations

Data product updates

- Data quality documentation updated
- Add publications (e.g., journal articles) citing data

DAAC Strengths

- High quality, detailed documentation
- Prompt, superb user support
- Broad expertise (staff and UWG)
- Good working relationships with data producers

Project Motivation

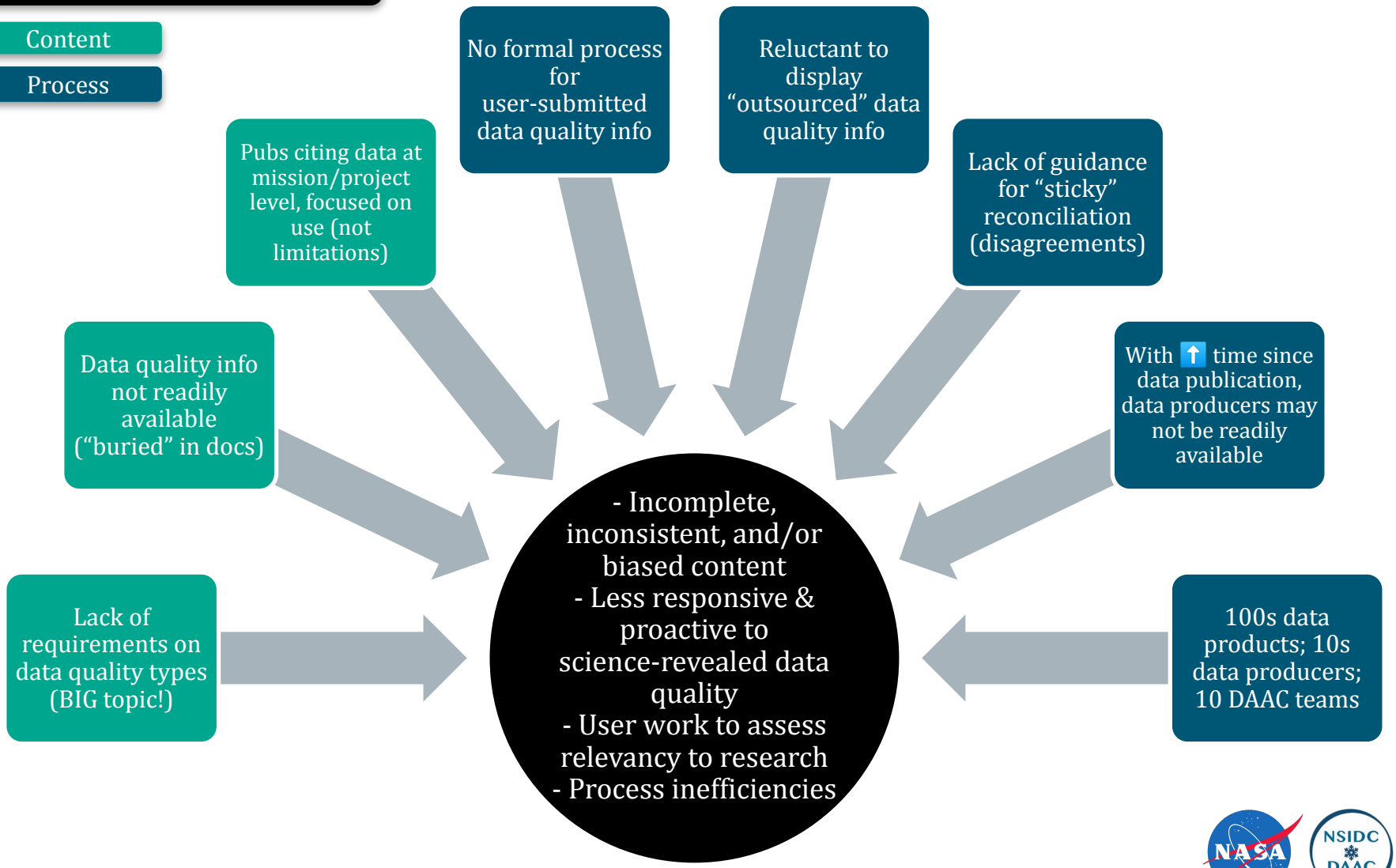
Challenges & Gaps

Content

Process

2020 UWG meeting: Presenting Limitations

- What is DAAC role in documenting data quality/limitations?
- How are data quality results reported back to DAAC and users?



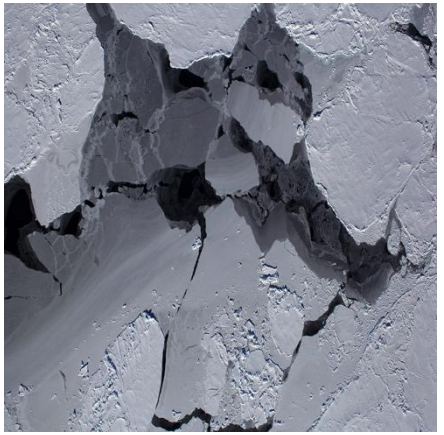
Project goals

- Develop model for improving content and display of specific, high priority scientific data quality info:
  - **how** content is to be collected, documented & displayed
  - **what** content is to be documented & displayed
  - **where** content is to be displayed
- Implement model on high priority data products (**proof of concept**)
- Get all of this  done in 3 months

Roadmap for today

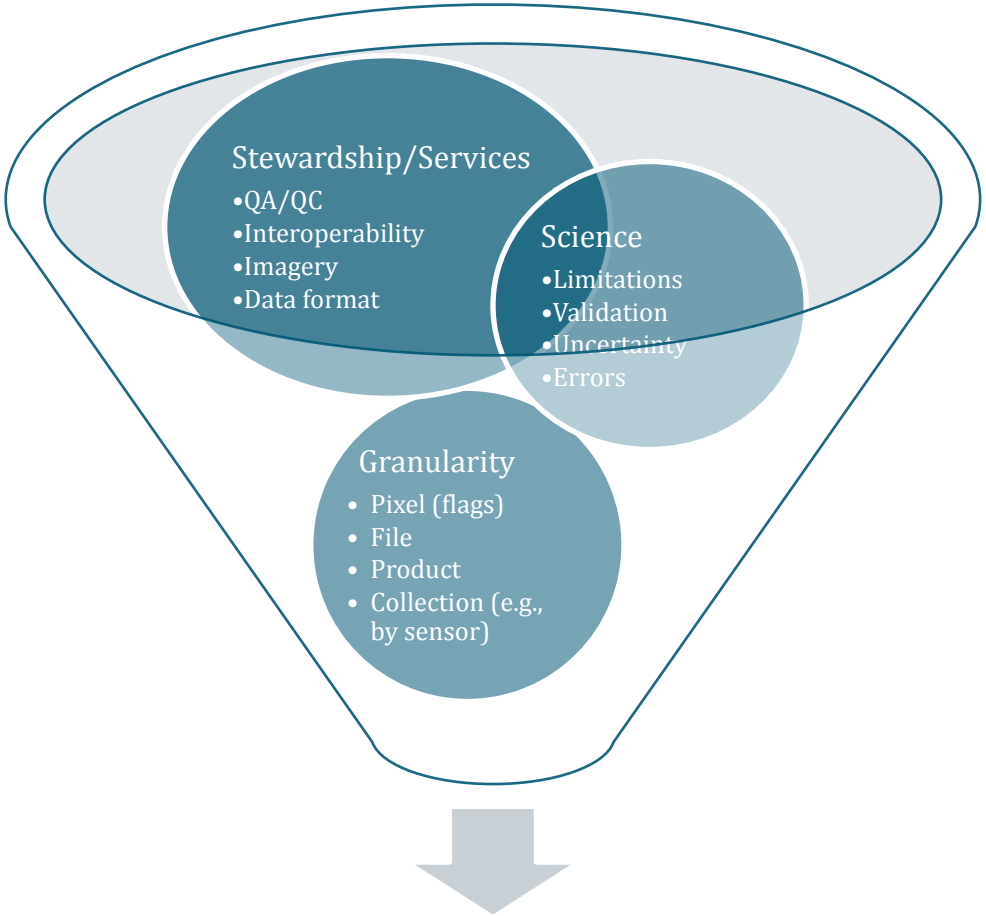
- Project scope
- The HOW
  - Processes & guidelines for how content is collected, documented & displayed
- The WHAT
  - What types of documentation
  - Content examples
- The WHERE
  - Where content is displayed
- WRAP-UP
  - Summary and what's next

Project scope

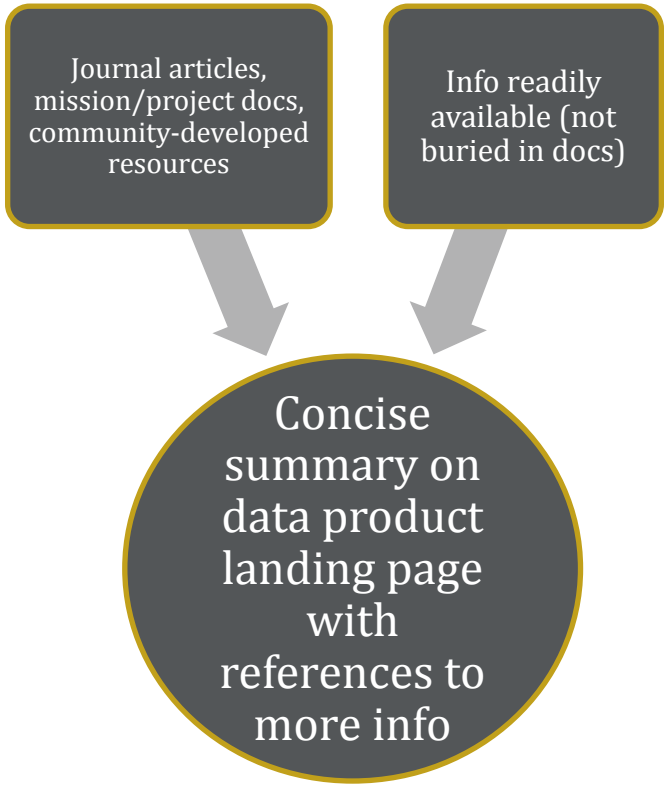


Scope: the WHAT & WHERE

TYPES of Data Quality



SOURCES of Data Quality Info



Data Product Scientific Strengths & Limitations

Scope: Proof-of-Concept data products

Data product(s)
AMSR-E SWE products (AE_DySno, AE_5DSno, AE_MoSno, NSIDC-0271)
ATLAS/ICESat-2 L3A Land Ice Height (ATL06)
IceBridge BedMachine Greenland (IDBMG4)
SMMR/SSM/I-SSMIS Sea Ice products (NSIDC-0051, -0079)
Near-Real-Time SSM/I-SSMIS EASE-Grid Daily Global Ice Concentration and Snow Extent (NISE)
SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture (SPL2SMP)
MEaSUREs GIMP products (8 products)

Why selected

- ✔ Highly visible, popular
- ✔ Known data quality issues
- ✔ Personnel expertise (UWG, DAAC Scientists)
- ✔ Leverage existing publications or documents (e.g., SnowPEx publication, SMAP Assessment Report)
- ✔ Engaged data producers
- ✔ Good “range” to test scalability of model

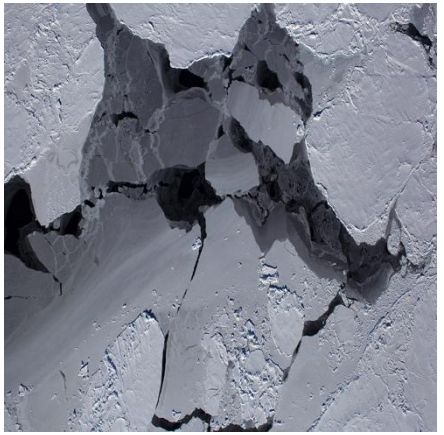
Scope: the HOW

As we curate content for proof-of-concept data products…

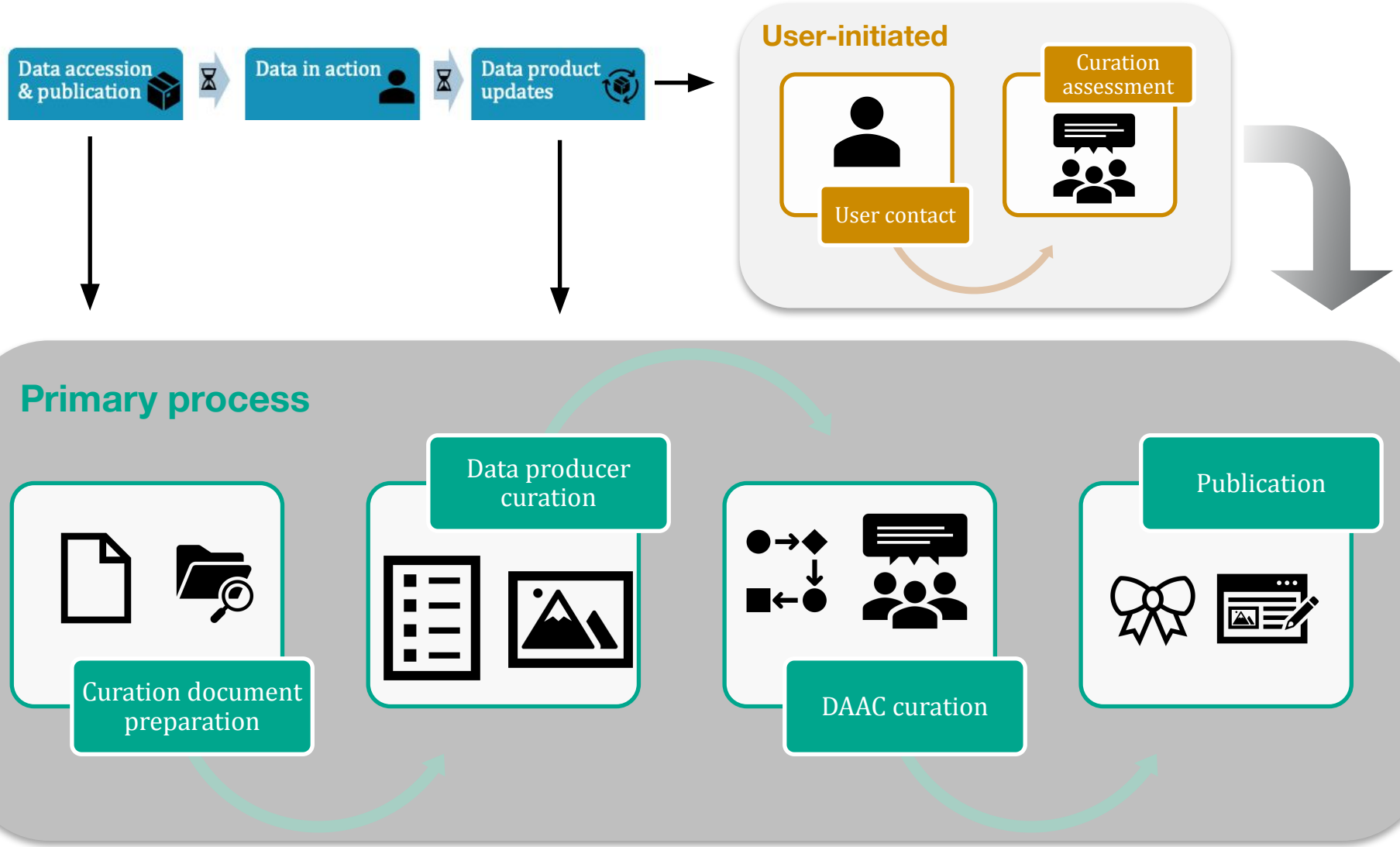
- Develop and iterate on processes and guidelines:
  - What exactly this new content is
  - Incorporating new content into existing data publication workflows
  - User-submitted content and reconciliation
- Gather feedback from DAAC, Data Producers, UWG

The HOW

how content is collected, documented & displayed

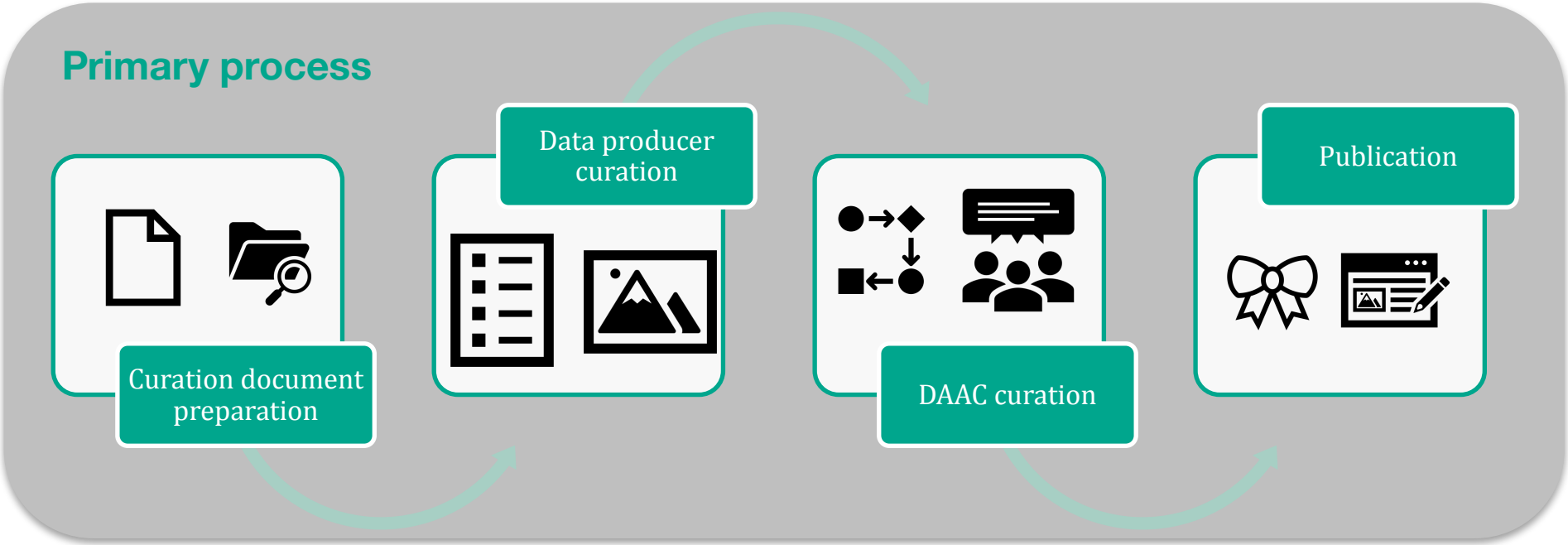


The HOW: Processes & Guidelines

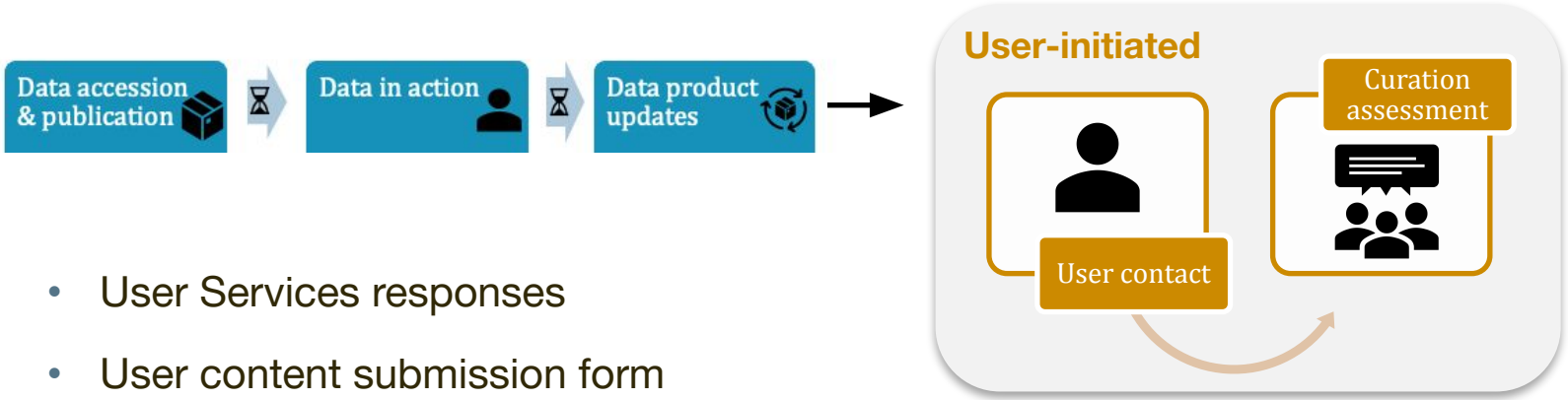


The HOW: Processes & Guidelines

- Clear guidelines for data producer content submission
- Collaborative curation process
- Internal (staff roles, system/doc management)



The HOW: Processes & Guidelines

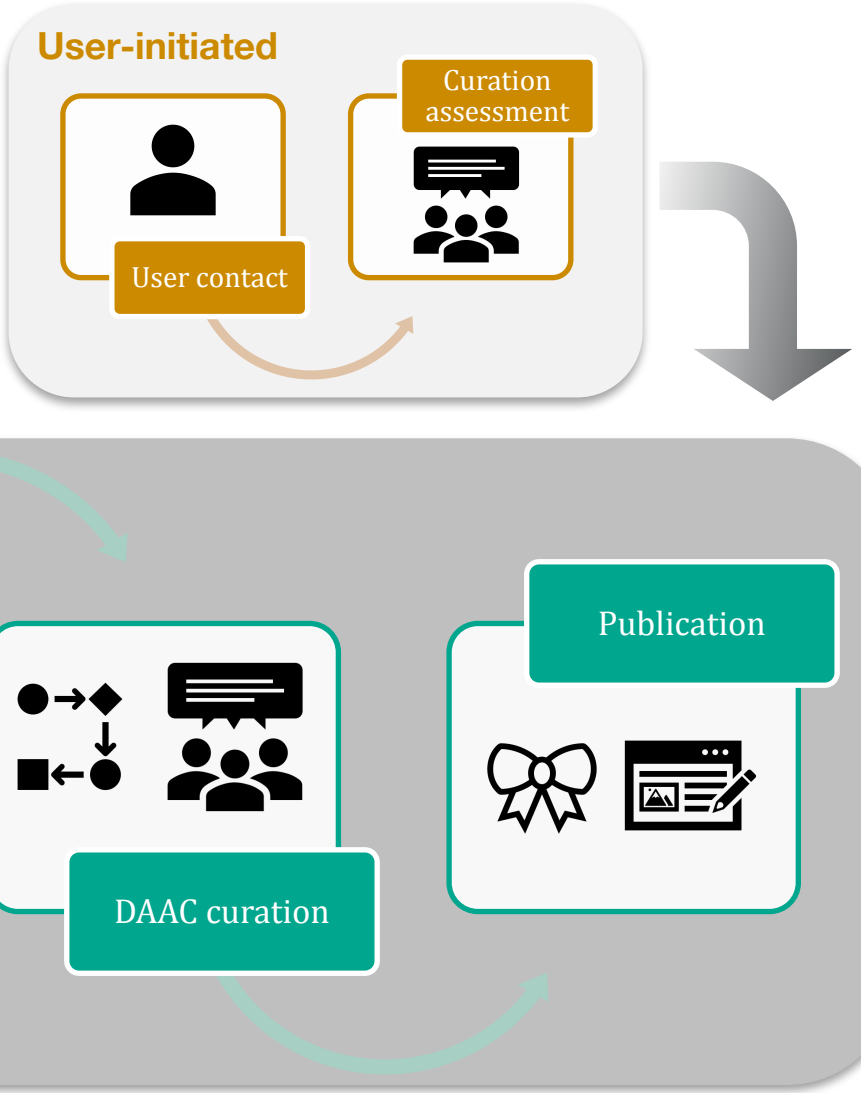


- User Services responses
- User content submission form
- Internal DAAC preliminary assessment

The HOW: Processes & Guidelines

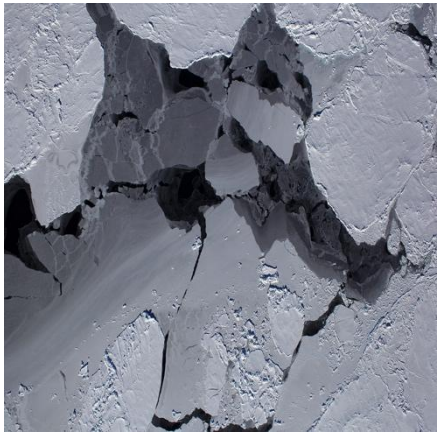
Reconciliation (rare, “sticky” scenarios)

- Present both “sides” in summary content backed by peer-reviewed publications
- Consult with DAAC Scientist, DAAC Management; possibly UWG, Project/Mission



The WHAT

what content is documented & displayed



The WHAT: Documentation types

>>> Curated, concise summary <<<

- Includes key scientific strengths and limitations in order to help users to relatively quickly distinguish this data product from others and determine suitability for research or application
- Does not include file-level information (e.g., data quality flags) or Level of Service-type information (e.g., imagery, documentation, data format)
- Bulleted form - a phrase or 1-2 sentences

>>> References <<<

- Hyperlinked citations to references (DOI URL) for more detailed info (1-2 per bullet)
- May include peer-reviewed journal articles, mission/project documents (e.g., ATBDs), or other community resources (e.g., NCAR Climate Data Guide)

The WHAT: Content examples

Strengths

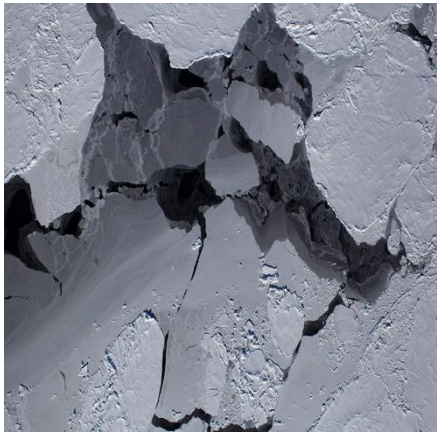
- *AMSR-E SWE*: Microwave observations provide surface snow and ice coverage during cloudy and night-time (including polar night) conditions (Cavalieri et al., 1999; Comiso and Nishio, 2008).
- *OIB*: Errors in bed elevations are minimized near flight lines and, where the source is mass conservation, in areas with fast ice flow speed (Morlighem et al., 2011).
- *SMAP*: Both 6 AM and 6 PM soil moisture retrievals have similar accuracy, and taken together, can provide data roughly every 1.5 days for a majority of locations (depending on latitude).
- *NISE*: Uses data from up to 5 days previous to fill in all spatial gaps (particularly important for lower latitude snow cover) (Armstrong and Brodzik, 2001).

Limitations

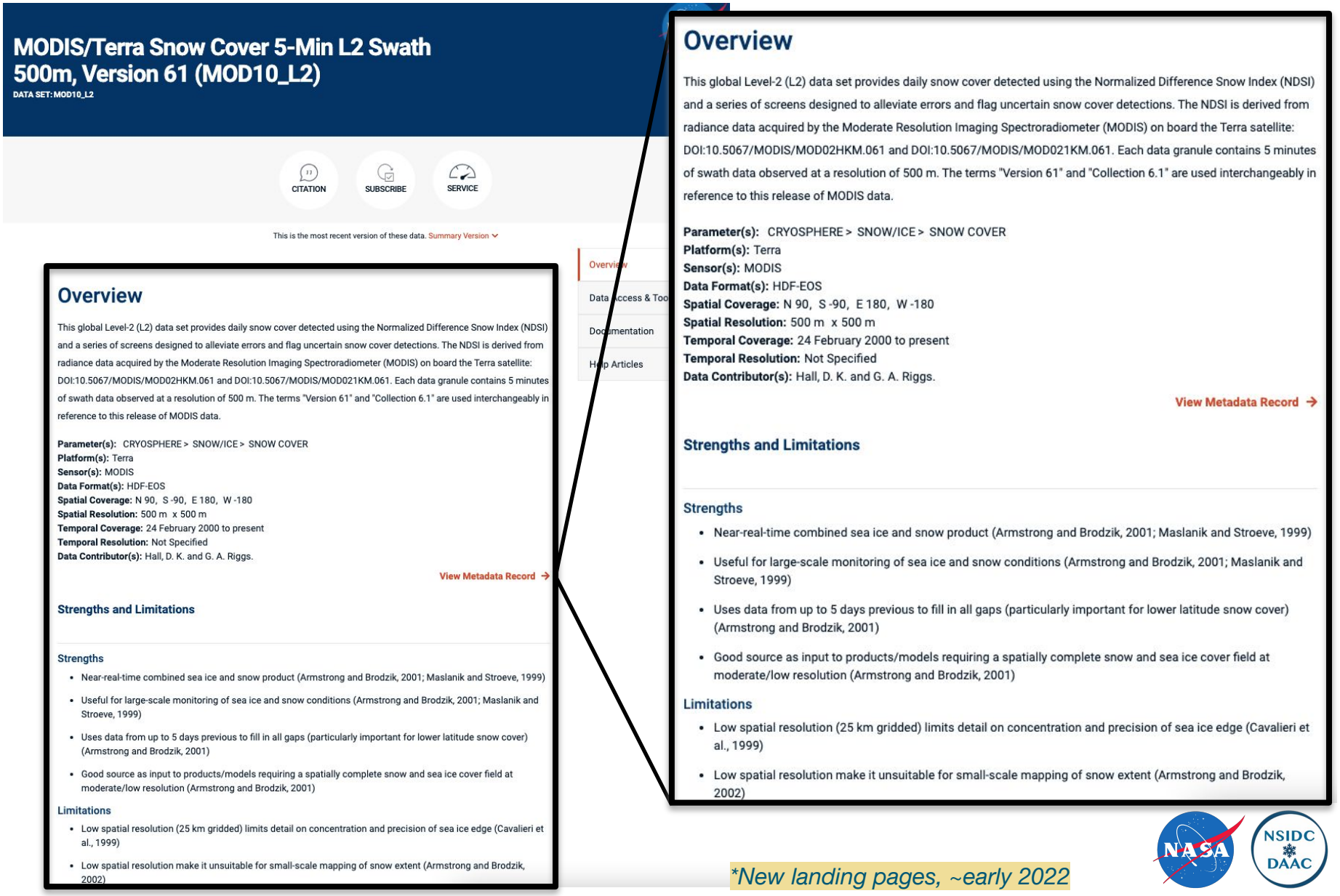
- *AMSR-E SWE*: Comparison with in situ data and other gridded SWE products under general snow and terrain conditions reveals substantial performance issues throughout the snow season (Mortimer et al, 2020).
- *PM sea ice*: Underestimates sea ice concentration during melt season (Kern et al., 2020) and/or when the ice is thin (Ivanova et al., 2015).
- *ATL06*: For the assessment of land ice mass change, height measurements provided by this product must be corrected for changes in the density of the firn (firn air content) through time. Estimates of firn air content are currently associated with large uncertainties (Smith et al., 2020).

The **WHERE**

where content is displayed

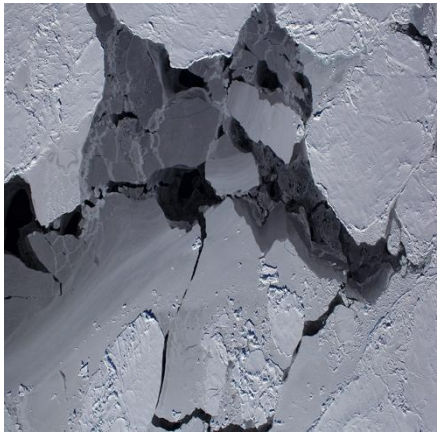


The WHERE: Landing page display*



*New landing pages, ~early 2022

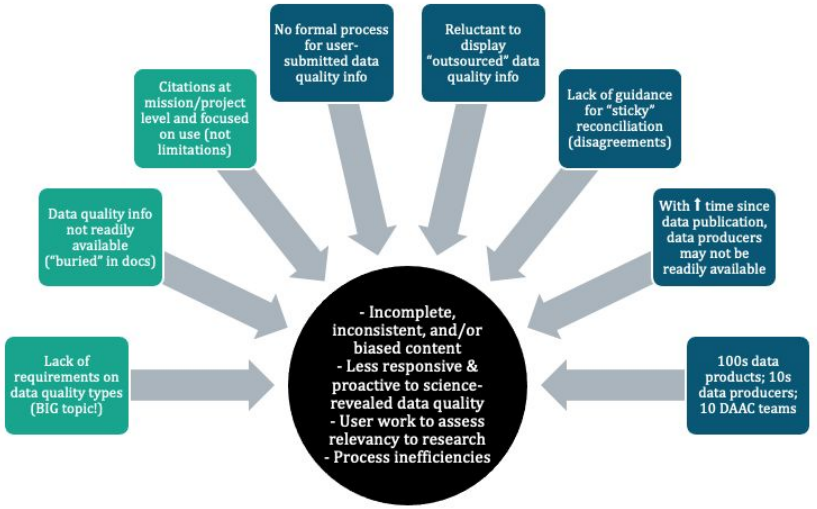
WRAP-UP
summary & moving forward



In summary

The WHAT & WHERE: Landing page summary with citations to detailed info

➔ Decrease work/time required for users to assess relevancy to research … and more time to do science!



The HOW: Processes & Guidelines

➔ Improve consistency, reduce bias in content

➔ Enable reactive and proactive curation of science-revealed data quality info

➔ Facilitate science community engagement, enable effective collaboration and reconciliation of content, and reduce reluctancy to display "outsourced" data quality info

➔ Scale and incorporate processes into existing data publication workflows

Moving forward

What's next?

- ❏ Implementation (beyond the model): All new and priority existing data products

When?

- ❏ Early 2022 (new nsidc.org website launch)

THANK YOU!

Feel free to contact me with questions/comments (shannon.leslie@colorado.edu)