RSpace – DMPTool Integration Demo

Rory Macneil

Research Space

Maria Praetzellis

California Digital Library/UC3

Open Science Fair
September 2021

University of California
CDL
California Digital Library

RSpace

Agenda

Overview of machine-actionable DMPs/DMPTool

Electronic Lab Notebooks/RSpace

The new integrated workflow with the RSpace/DMPTool

University of California
CDL
California Digital Library

RSpace

California Digital Library (CDL)/UC3

CDL founded by the University of California in 1996

University of California Curation Center (UC3) is CDL's program concerned with maintaining, preserving, and adding value to digital research data throughout its lifecycle **UC3 areas of focus:**

- Research data management
- Data publication and data metrics
- Persistent identifiers
- Digital preservation
- Data/software skills training

University of California
CDL
California Digital Library

RSpace

DMPTool

Free, open-source community supported tool

+60K users, 310 participating institutions

Create next-generation, machine-actionable DMPs

University of California
CDL
California Digital Library

What's in a DMP?

UC3

University of California
CDL
California Digital Library

DMP Basics

A document that addresses how you will **manage** and **secure** your data throughout the lifecycle of a research project

Can be both a **required document** for grants and a **living document** for research planning purposes

UC3

University of California
CDL
California Digital Library

Standard Components of a DMP

- Data Collection
  - What data will you collect or create?
- Documentation and Metadata
  - What documentation and metadata will accompany the data?
- Ethics and Legal Compliance
  - How will you manage any ethical issues?
- Storage and Backup
  - How will the data be stored and backed up during the research?
- Selection and Preservation
  - Which data are of long-term value and should be shared, and/or preserved?
- Data Sharing
- Responsibilities and Resources

UC3

University of California
CDL
California Digital Library

Machine-actionable DMPs

The principle goal of a machine actionable DMP is to support the creation and stewardship of FAIR data

- Allow data and information about research to be communicated and shared across stakeholders
- Facilitating
  - notifications and verification
  - real-time reporting
  - automated compliance
- maDMPs should lessen the administrative burden on researchers and grant administrators.

- Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support. https://doi.org/10.29242/report.effectivedatapractices2020

UC3

University of California CDL California Digital Library

NSF EAGER research

Active DMPs Grant 2018-2021

- Developinging and implementing the metadata structure and technical features needed to facilitate machine-actionability of DMPs
- Testing the hypothesis that we can connect DMPs to PID graph

FAIR Island Grant 2021-2023

- Building interoperability between pieces of critical research infrastructure -- Data Management Plans (DMPs), open data policy, DOIs, and publications.
- Piloting and continuing to develop the technical infrastructure built in our prior NSF grant.

UC3

University of California CDL California Digital Library

Networked DMPs and the PID Graph

UC3

University of California
CDL
California Digital Library

PIDs for DMPs

Generating identifiers for DMPs creates an unbreakable link between a data plan to the project outputs and facilitates connections in support of FAIR data.

UC3

University of California
CDL
California Digital Library

Thanks to Erin Robinson, Metadata Game Changers for the use of this graphic.

Identifiers connect research activities

DMPTool supports PIDs within a DMP:

- **DMP-IDs**
- **RORs** for research organizations
- **Funder Registry IDs** for funders
- **ORCIDs** for DMP creators and collaborators
- **Registry of Research Data Repositories** (re3data)
- **Licenses** (spdx)
- **RDA Metadata Standards Directory**

UC3

University of California
CDL
California Digital Library

Landing page for a DMP-ID

https://dmphub.cdlib.org/dmps/doi:10.48321/D1F88S

A data management plan for this project was submitted to the Gordon and Betty Moore Foundation (GBMF) with the information below.

This page represents key information from a data management plan.

Moorea Biocode

Contributors to this project

Brent Mishler: Investigation, University of California, Berkeley (UCB), https://orcid.org/0000-0001-5727-4916
Chris Meyer: Investigation, Smithsonian Institution (SI), https://orcid.org/0000-0003-2501-7952
Claude Payri: Investigation, University of French Polynesia (upf.pf), https://orcid.org/0000-0002-0393-6811
Craig Moritz: Investigation, University of California, Berkeley (UCB), https://orcid.org/0000-0001-5313-7279
George Roderick: Investigation, University of California, Berkeley (UCB), https://orcid.org/0000-0001-7557-2415
Gustav Paulay: Investigation, University of Florida (ufl.edu), https://orcid.org/0000-0003-4118-9797
Jean-Yves Meyer: Investigation, Délégation à la Recherche, Government of French Polynesia, https://orcid.org/0000-0001-7968-9944
John Deck: Data-curation, University of California, Berkeley (UCB), https://orcid.org/0000-0002-5905-1617
Joyce Gross: Data-curation, University of California, Berkeley (UCB)
Len Hirsch: Investigation, Smithsonian Institution (SI)
Matteo Garbelotto: Investigation, University of California, Berkeley (UCB), https://orcid.org/0000-0001-9737-0128
Neil Davies: Data-curation, Investigation, Gump South Pacific Research Station (moorea.berkeley.edu), https://orcid.org/0000-0001-8085-5014
Olivier Gargominy: Investigation, National Museum of Natural History (mnhn.fr), https://orcid.org/0000-0001-7807-944X
Richard Pyle: Investigation, Bernice P. Bishop Museum (bishopmuseum.org), https://orcid.org/0000-0003-0768-1286
Serge Planes: Investigation, French National Centre for Scientific Research (cnrs.fr), https://orcid.org/0000-0002-5689-5371
Sylvain Charlat: Investigation, University of Lyon System (lyon-university.org), https://orcid.org/0000-0003-0760-0087

Project details

Project Start: November 22, 2007
Project End: March 31, 2010
Created: August 02, 2021
Modified: August 02, 2021
Ethical issues related to data that this DMP describes?: no

Citation

When citing this DMP use:
Neil Davies. (2021). "Moorea Biocode" [Data Management Plan]. DMPHub. https://doi.org/10.48321/D1F88S

When connecting to this DMP to related project outputs (such as datasets) use the ID:
https://doi.org/10.48321/D1F88S

Funding status and sources for this project

Status: Granted
Funder: Gordon and Betty Moore Foundation (GBMF)
Funding opportunity number: GBMF1619
Grant: https://www.moore.org/grant-detail?grantId=GBMF1619

Project description

The core goal of the Moorea Biocode Project (MBP) is to genetically barcode every non-microbial species on Moorea. We will thus create a professionally produced, verifiable (vouchered) All Taxa Biotic Inventory (ATBI) of Moorea (Output 1), together with the informatics services needed for ATBI and biocode-enabled research in Model Ecosystems (Outcome 2). We will also design a long-term plan (business model) to ensure the financial self-sustainability of biocode services, and carry out proof of concept studies to encourage broad-based support from funding agencies (Output 3).

Planned outputs

Dataset for: Moorea Biocode

Format: Dataset
Anticipated volume: unspecified
Release timeline: Unknown
Intended repository:
License for reuse: unspecified

Other works associated with this research project

Articles

Nitta, J. (2008). "Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea" [Article-journal]. In Wiley. https://doi.org/10.1002/tax.573006

ANKER, A. (2018). "Metabetaeus Borradaile, 1899 revisited, with description of a new marine species from French Polynesia (Crustacea: Decapoda: Alpheidae)" [Article-journal]. In Magnolia Press. https://doi.org/10.11646/zootaxa.2552.1.2

ANKER, A. (2018). "New findings of rare or little-known alpheid shrimp genera (Crustacea, Decapoda) in Moorea, French Polynesia" [Article-journal]. In Magnolia Press. https://doi.org/10.11646/zootaxa.2403.1.3

ANKER, A. (2018). "On two snapping shrimps, Alpheus baccheti n. sp. and A. coetivensis Coutière from the Tuamotu Islands, French Polynesia (Crustacea, Decapoda)" [Article-journal]. In Magnolia Press. https://doi.org/10.11646/zootaxa.2492.1.3

ANKER, A. (2018). "The mud-shrimp genus Axianassa Schmitt, 1924 (Crustacea, Decapoda, Axianassidae) in the Indo-West Pacific, with description of a new species from French Polynesia" [Article-journal]. In Magnolia Press. https://doi.org/10.11646/zootaxa.2557.1.5

KOMAI, T., ANKER, A. (2018). "Two new species of the laomediid genus Naushonia Kingsley, 1897 (Crustacea: Decapoda: Gebiidea) from the Indo-West Pacific" [Article-journal]. In Magnolia Press. https://doi.org/10.11646/zootaxa.2504.1.3

Malay, M., Paulay, G. (2010). "PERIPATRIC SPECIATION DRIVES DIVERSIFICATION AND DISTRIBUTIONAL PATTERN OF REEF HERMIT CRABS (DECAPODA: DIOGENIDAE:CALCINUS)" [Article-journal]. In Wiley. https://doi.org/10.1111/j.1558-5646.2009.00848.x

Plaisance, L., Knowlton, N., Paulay, G., Meyer, C. (2009). "Reef-associated crustacean fauna: biodiversity estimates using semi-quantitative sampling and DNA barcoding" [Article-journal]. In Springer Science and Business Media LLC. https://doi.org/10.1007/s00338-009-0543-3

Hubert, N., Delrieu-Trottin, E., Irisson, J., Meyer, C., Planes, S. (2010). "Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae" [Article-journal]. In Elsevier BV. https://doi.org/10.1016/j.ympev.2010.02.023

COHEN, N., DEEDS, J., WONG, E., HANNER, R., YANCY, H., WHITE, K., THOMPSON, T., WAHL, I., PHAM, T., GUICHARD, F., HUH, I., AUSTIN, C., DIZIKES, G., GERBER, S. (2009). "Public Health Response to Puffer Fish (Tetrodotoxin) Poisoning from Mislabeled Product" [Article-journal]. In International Association for Food Protection. https://doi.org/10.4315/0362-028X-72.4.810

Check, E. (2006). "Treasure island: pinning down a model ecosystem" [Article-journal]. In Springer Science and Business Media LLC. https://doi.org/10.1038/439378a

Leray, M., Boehm, J., Mills, S., Meyer, C. (2012). "Moorea BIOCODE barcode library as a tool for understanding predator–prey interactions: insights into the diet of common predatory coral reef fishes" [Article-journal]. In Springer Science and Business Media LLC. https://doi.org/10.1007/s00338-011-0845-0

Deck, J., Gaither, M., Ewing, R., Bird, C., Davies, N., Meyer, C., Riginos, C., Toonen, R., Crandall, E. (2018). "The Genomic Observatories Metadatabase (GeOMe): A new repository for field and sampling event metadata associated with genetic samples" [Article-journal]. In Public Library of Science (PLoS). https://doi.org/10.1371/journal.pbio.2002925

Hubert, N., Delrieu-Trottin, E., Irisson, J., Meyer, C., Planes, S. (2010). "Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae" [Article-journal]. In Elsevier BV. https://doi.org/10.1016/j.ympev.2010.02.023

RSpace

UC3

University of California
CDL
California Digital Library

RSpace

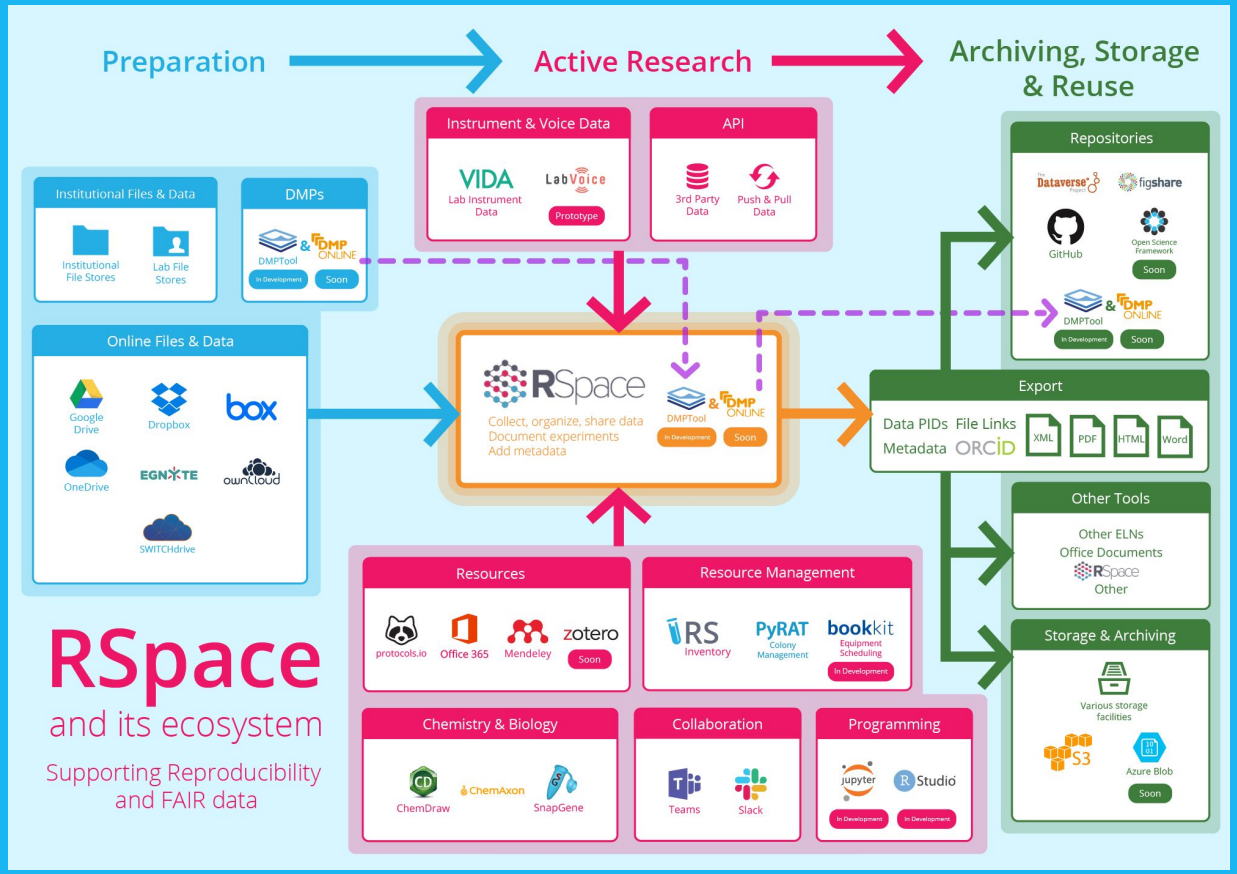
Electronic lab notebook

Digital hub for collecting, organizing managing and sharing data in the active research phase and passing data on for storage, archiving and wider access

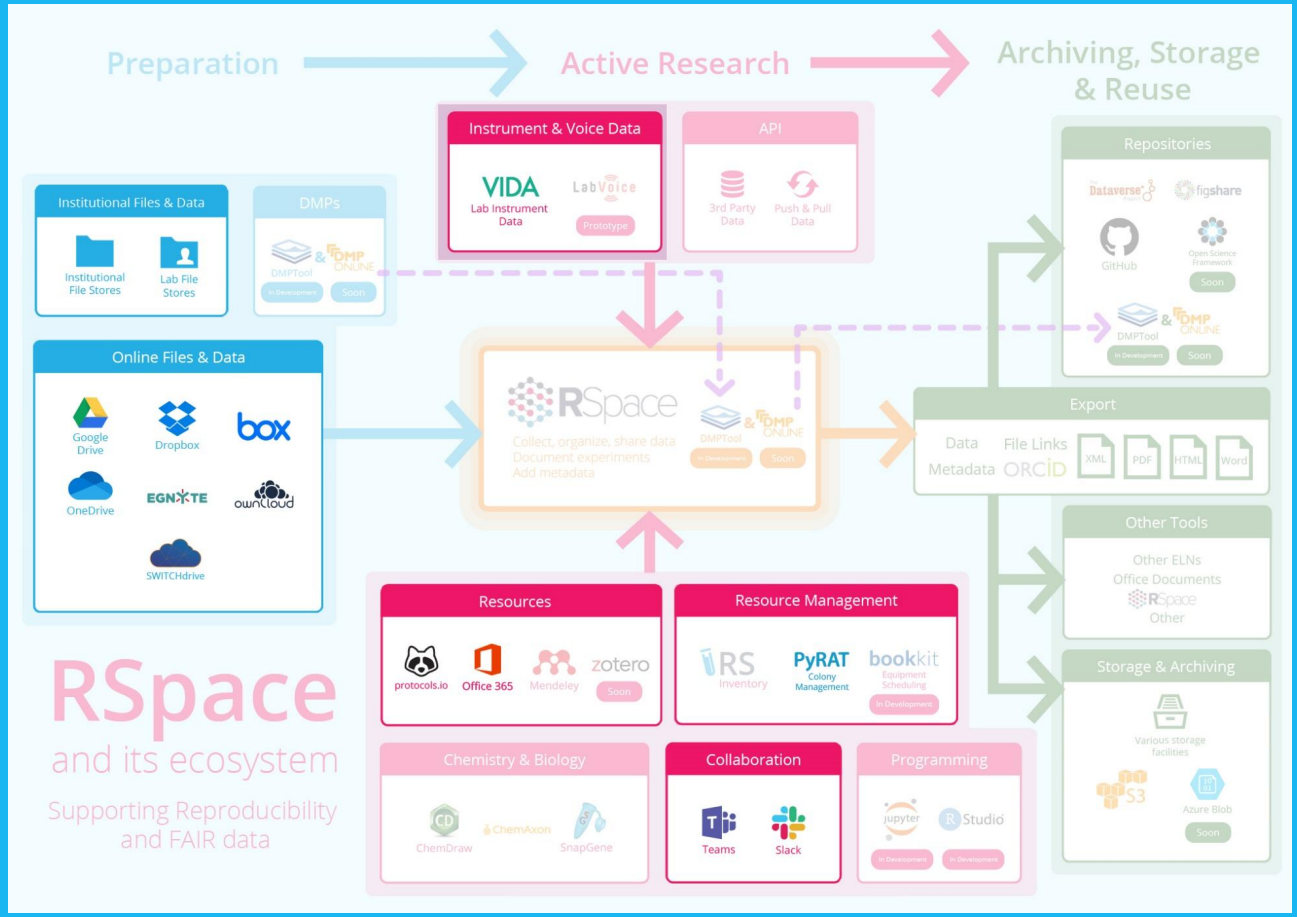
RSpace

- University of Wisconsin origins
- Built around academic workflows
- Provides a centralised hub for data
- Integrates with tools researchers already use
- A conduit, not a silo
- Connects data management planning, active research, and data archiving and storage
- Enhances reproducibility and FAIRification of data

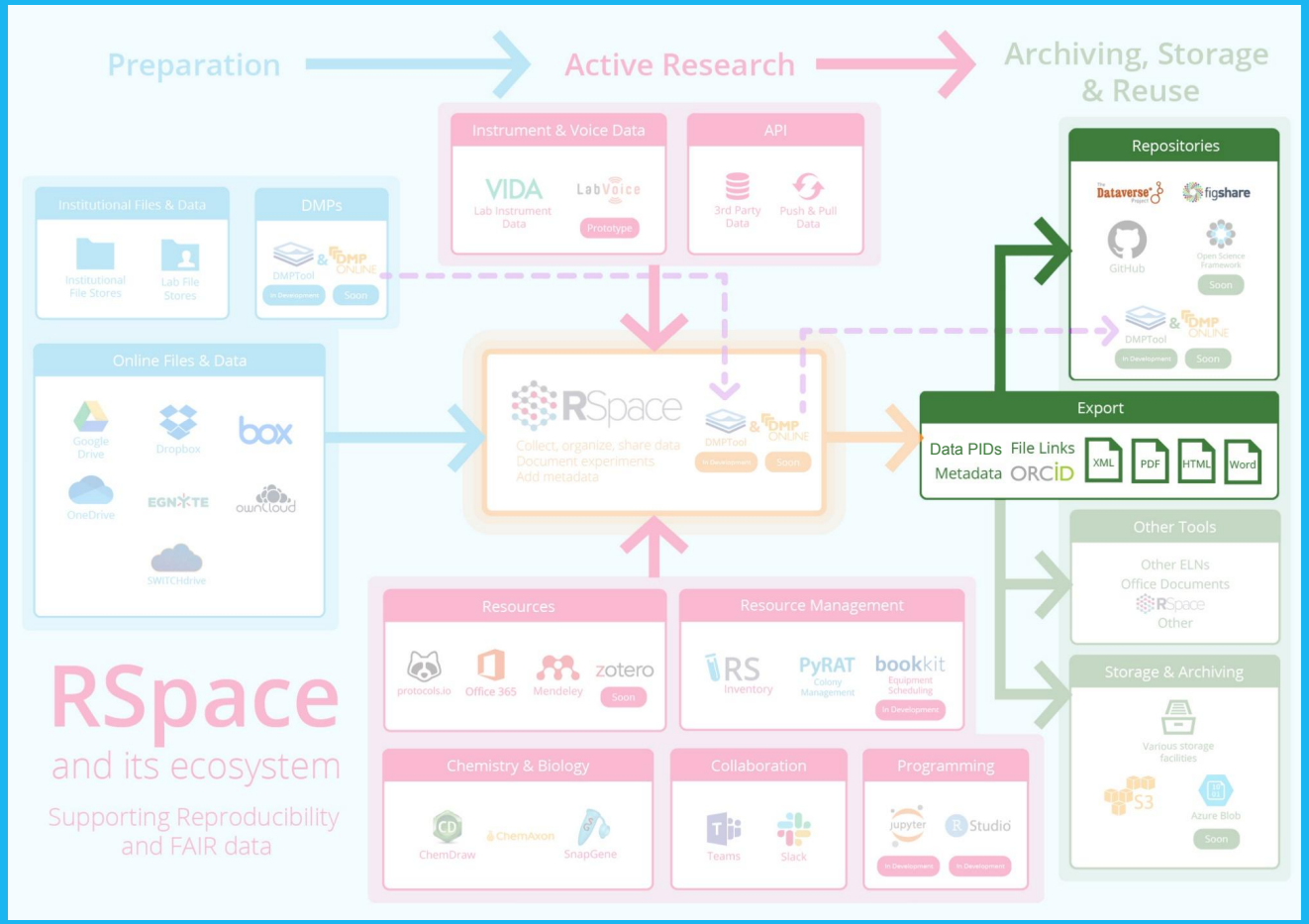
From tool to ecosystem



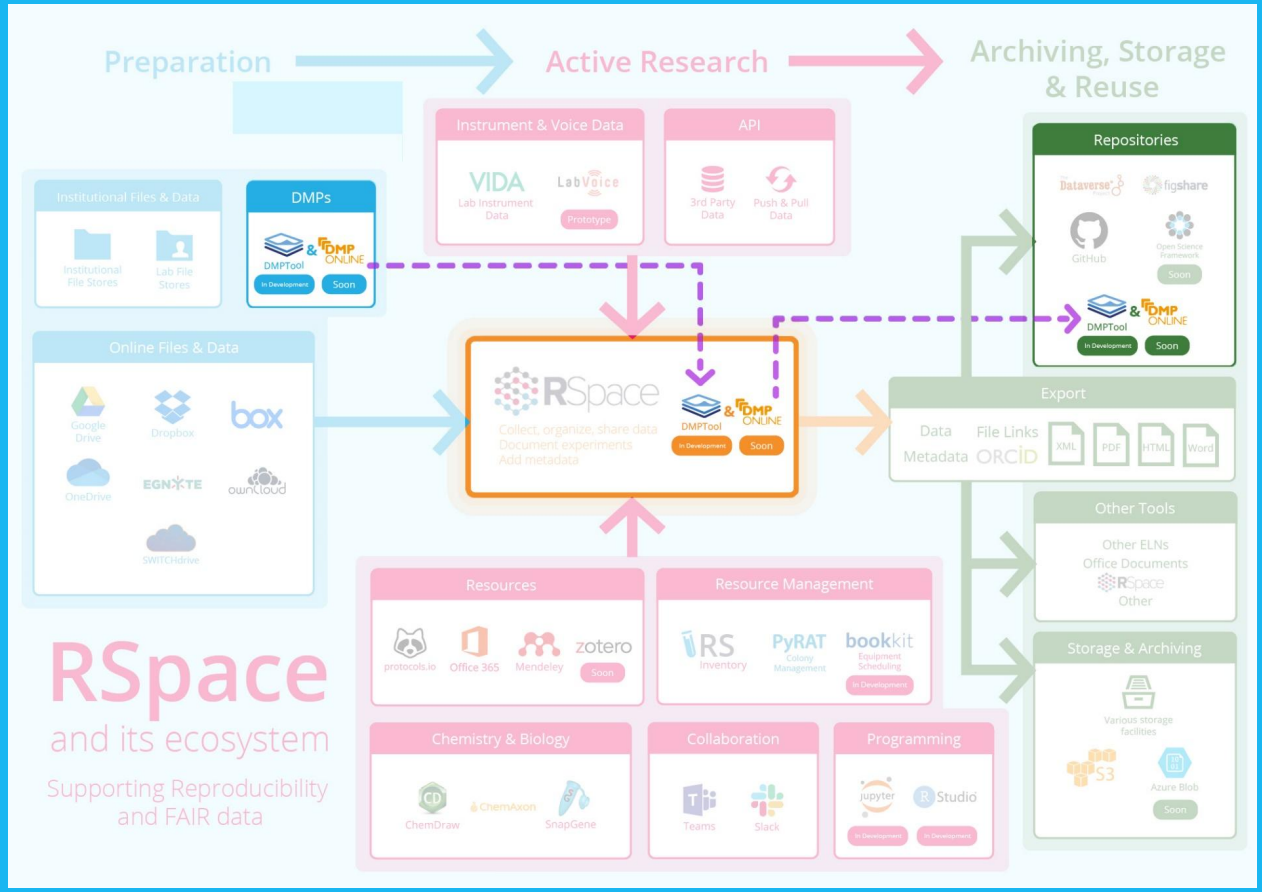
Incorporate data from external resources



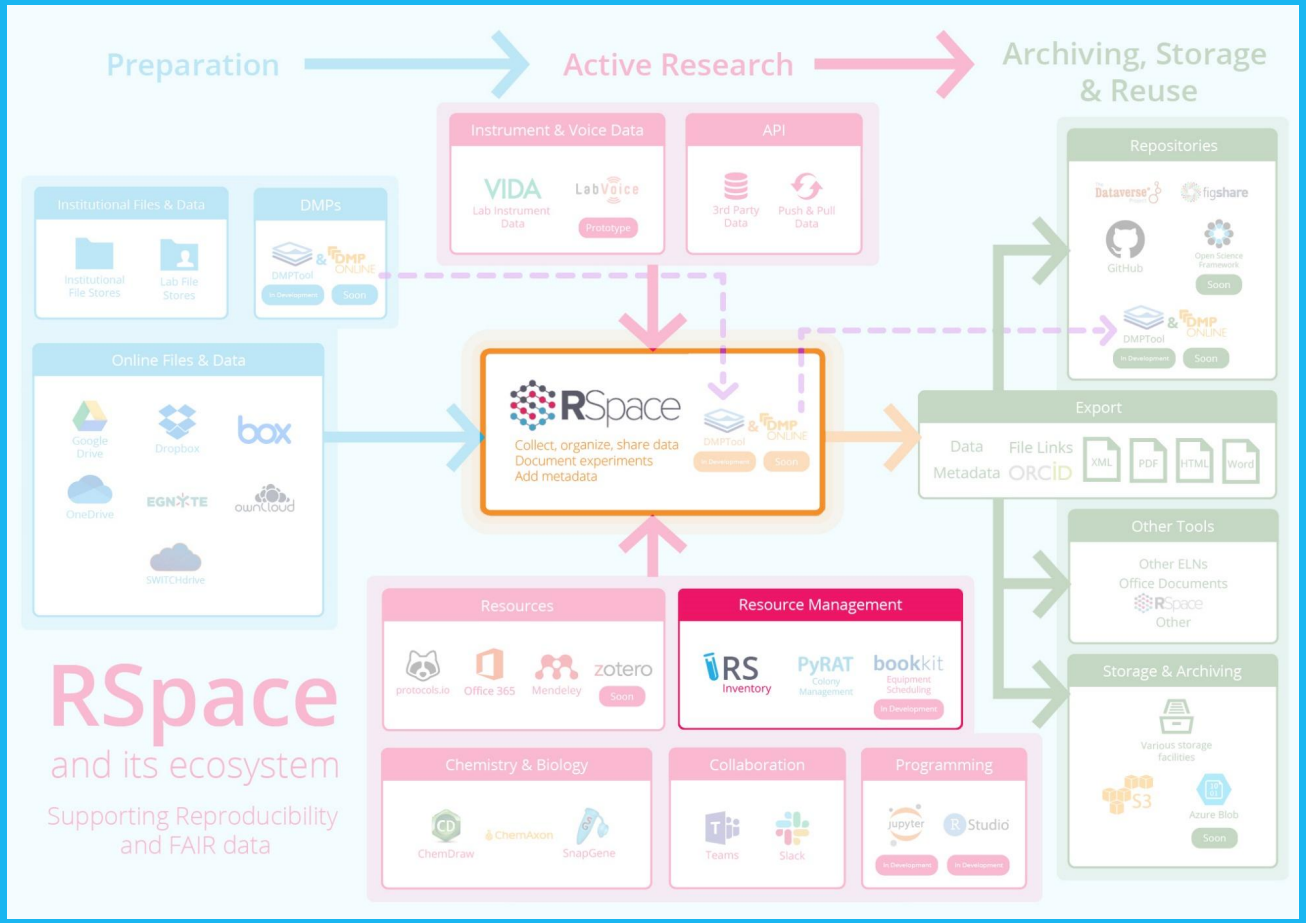
Export data to repositories



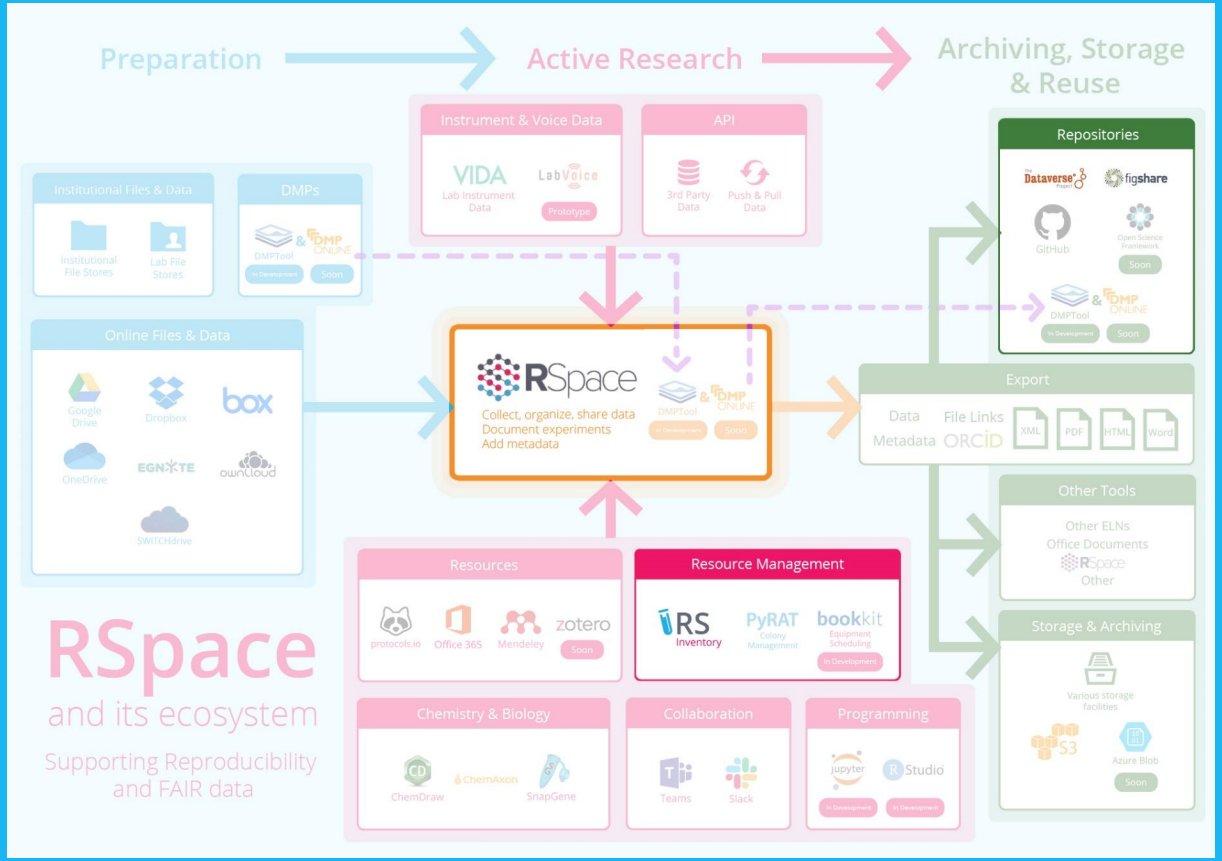
Integrate with maDMPs and export to repositories



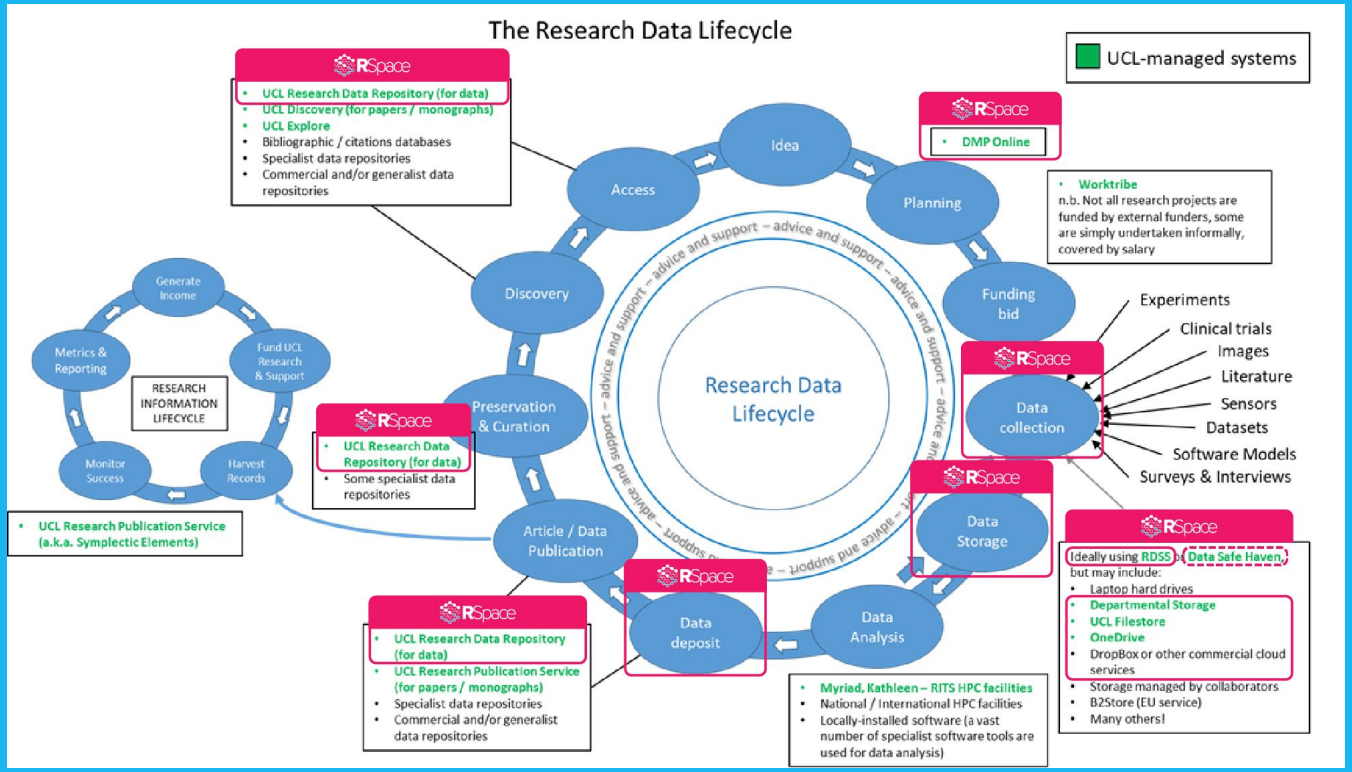
Incorporate sample data into experimental record



Sample data in experimental context captured for export to repository



RSpace in an institutional environment



Looking ahead: Automating metadata proliferation throughout the research lifecycle

- Tools and resources
  - maDMPS
  - ORCID enhancements
  - PIDS
    - Domain specific: IGSNS, RRIDS
    - From DMPS
      - DMP-IDs
      - RORs for research organizations
      - Funder Registry IDs for funders
      - ORCIDs for DMP creators and collaborators
      - Registry of Research Data Repositories (re3data)
      - Licenses (spdx)
      - RDA Metadata Standards Directory
- People and process
  - ERN (RSpace) as data curation tool for data stewards

RSpace

DMPTool - RSpace - Dataverse

Demo video

https://www.youtube.com/watch?v=dIsYLcx6qNc

University of California
CDL
California Digital Library

RSpace

Next steps

University of California
CDL
California Digital Library

RSpace

FAIR Island Vision

*Develop the **optimal data policies and technical infrastructure** necessary to create an environment where all data and knowledge collected **at a place** is curated and made openly available as quickly as possible.*

METADATA GAME CHANGERS

BERKELEY Institute for Data Science

University of California CDL California Digital Library

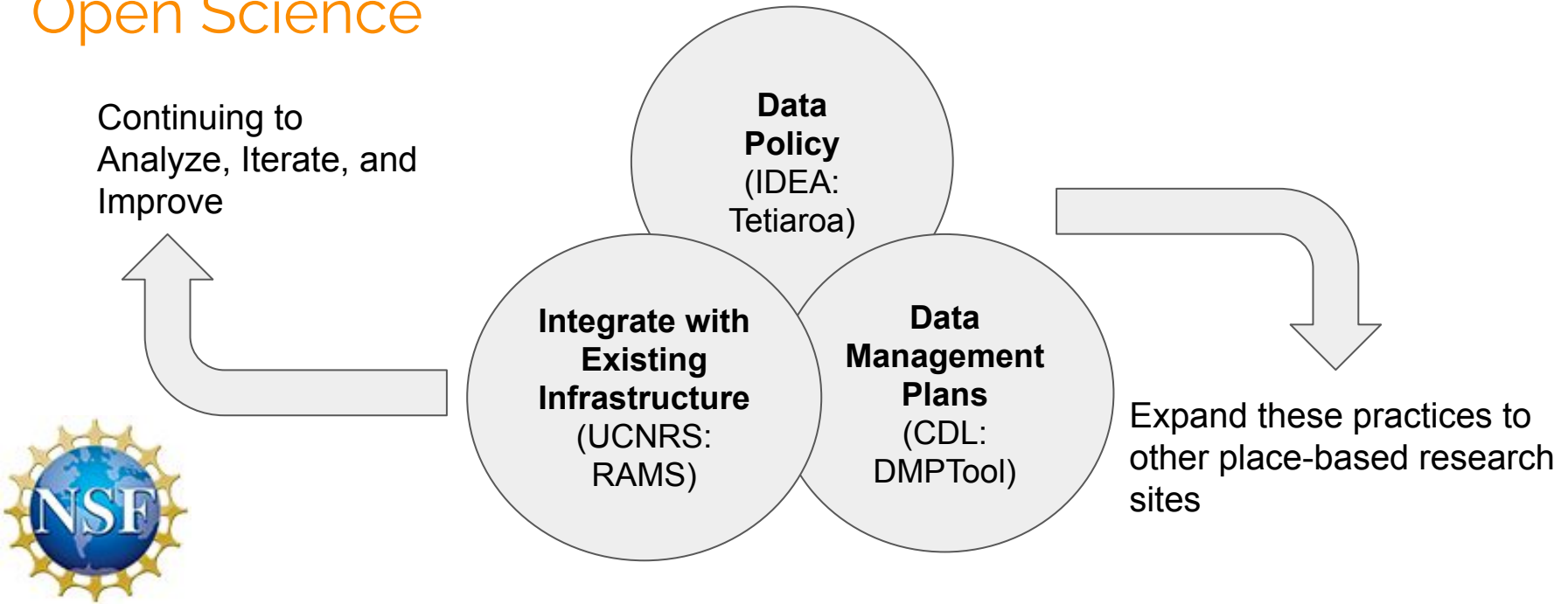
UC3 UC Curation Center

DataCite FIND, ACCESS, AND REUSE DATA

gump station University of California

TETIAROA SOCIETY

NSF EAGER: The FAIR Island Project for Place-based Open Science



Maria Praetzellis, John Chodacki, Neil Davies, Erin Robinson, Matthew Buys, & Catherine Nancarrow. (2021). EAGER: The FAIR Island Project for Place-based Open Science. Zenodo.
https://doi.org/10.5281/zenodo.5117892

University of California CDL California Digital Library

Support for DMPs as living documents: Active research phase

PHASE 1 (Fall 2021): Researchers can keep DMPs up to date in RSpace by connecting outputs to a project documented in the original DMP

PHASE 2 (Winter 2021/2): Researchers can ingest their DMP into RSpace and edit the DMP as research progresses

University of California
CDL
California Digital Library

RSpace

Support for DMPs as living documents: Post research phase

RSpace users can generate an inventory of outputs associated with a project and tied to a DMP for use in for grant reporting and compliance

RSpace can pull other kinds of PIDs, e.g. RORs, from the DMP, to associate with RSpace datasets

RSpace users can create/publish DMP metadata and acquire a DOI for a text-based DMP written outside the DMP management tools

University of California
CDL
California Digital Library

RSpace

Extend integration to other DMP services

University of California
CDL
California Digital Library

RSpace

Thank you!

Maria Praetzellis
@thedmptool
maria.praetzellis@ucop.edu

Rory Macneil
@TheConnectedLab
rmacneil@researchspace.com

University of California
CDL
California Digital Library

RSpace