# OUTLINE

1. Common challenges when using data

2. Making data processing *reusable*

# COMMON CHALLENGES WHEN USING DATA

# COMMON CHALLENGES WHEN USING DATA

- Data can be **BIG**

- **Locally hosted data is fragile** – but downloading takes sooo long!

- **Manual downloads are also fragile** – *where did I put the URL?!*

- Updating or adding **new data can be a pain**

- **Data processing take be challenging and time consuming** – especially spatial data (projection, origin, resolution and alignment)

# COMMON CHALLENGES WHEN USING DATA

"Solutions/helpers"

- Data can be **BIG**

- **Locally hosted data is fragile** – but downloading takes sooo long!

- **Manual downloads are also fragile** – *where did I put the URL?!*

- Updating or adding **new data can be a pain**

- **Data processing take be challenging and time consuming** – especially spatial data (projection, origin, resolution and alignment)

cache downloads

use disk-based solutions (not covered today)

# COMMON CHALLENGES WHEN USING DATA

"Solutions/helpers"

- Data can be **BIG**

- **Locally hosted data is fragile** – but downloading takes sooo long!

- **Manual downloads are also fragile** – *where did I put the URL?!*

- Updating or adding **new data can be a pain**

- **Data processing take be challenging and time consuming** – especially spatial data (projection, origin, resolution and alignment)

cache downloads

use disk-based solutions (not covered today)

**script** downloads

# COMMON CHALLENGES WHEN USING DATA

"Solutions/helpers"

- Data can be **BIG**

- **Locally hosted data is fragile** – but downloading takes sooo long!

- **Manual downloads are also fragile** – *where did I put the URL?!*

- Updating or adding **new data can be a pain**

- **Data processing take be challenging and time consuming** – especially spatial data (projection, origin, resolution and alignment)
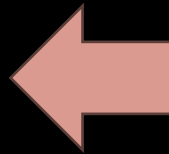
cache downloads

use disk-based solutions (not covered today)

**script** downloads

use functions to process data according to certain expectations, tolerate similar, but not identical, datasets and that have internal checks for data format, etc.

SpaDES workshop - June 11-13 2024

# MAKING DATA PROCESSING REUSABLE

"Solutions/helpers"

shareable/reusable
data processing

| cache downloads |
|---|
| use disk-based solutions (not covered today) |

| **script** downloads |
|---|

| use functions to process data according to certain expectations, tolerate similar, but not identical, datasets and that have internal checks for data format, etc. |
|---|

# MAKING DATA PROCESSING REUSABLE

"Solutions/helpers"

cache downloads

use disk-based solutions (not covered today)

shareable/reusable
data processing

**script** downloads

prepInputs()

use functions to process data according to certain expectations, tolerate similar, but not identical, datasets and that have internal checks for data format, etc.
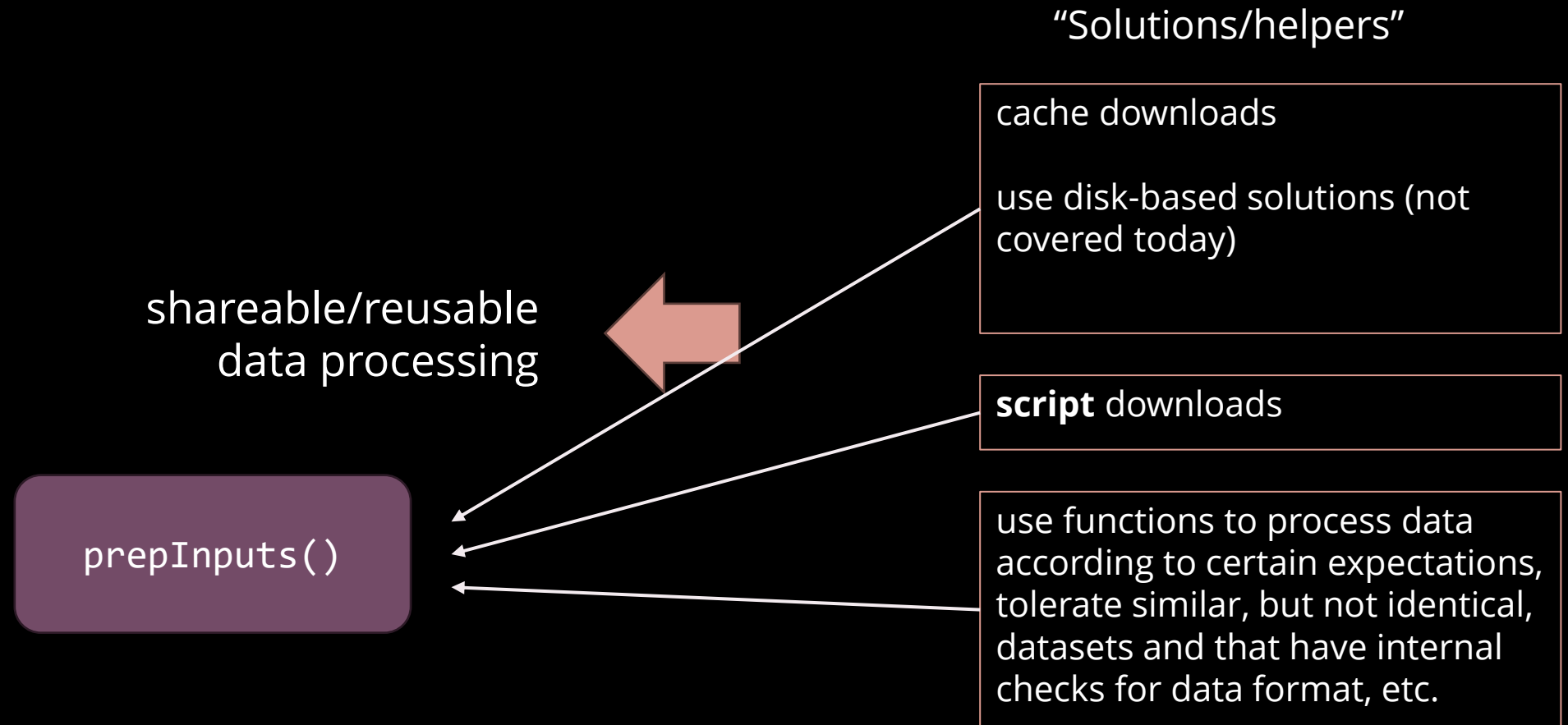
# TIME TO LOOK AT SOME CODE

→ Robust and nimble scientific workflows, using SpaDES

Workshop Agenda

This is a *hands-on* workshop

Please **ask questions**, share your troubles and successes

**Last 15 min** of each WOYO is for discussion

We are always available via **Teams chat**