Drivers for the growing focus on research data

➢ The rise of datasets as research objects of interest
➢ The drive by funders to make data publicly available
➢ The move by publishers to accommodate improved association of research data with the publication
➢ The development of new, more data-focussed, publishing platforms
  ➢ Science Matters
  ➢ EMBO Source Data
  ➢ Wellcome Open Research

Initial responses have involved repositories

➢ Repositories
  ➢ Older repositories moving to incorporate deposit of data in addition to publications
  ➢ New repositories focused solely or primarily on data Addition of new capabilities for accessing and making use of data after deposit
➢ Institutions
  ➢ Interest in getting more data into repositories
  ➢ Adopting/adapting repositories for deposit of datasets
➢ Publishers
  ➢ Requirement to provide greater quantity and quality of data associated with a publication
  ➢ Repositories as 'parking place' for data
  ➢ List of acceptable repositories

So far, little focus on 'data to repository' or 'data to publication'

That's a problem, because

- **People aren't used to depositing data**
- **Preparing data for deposit is extra work**
- **Data is often imprisoned in tools from which deposit is impossible or difficult**

RSpace

What's needed:

A vehicle or mechanism to stimulate
and streamline the flow of
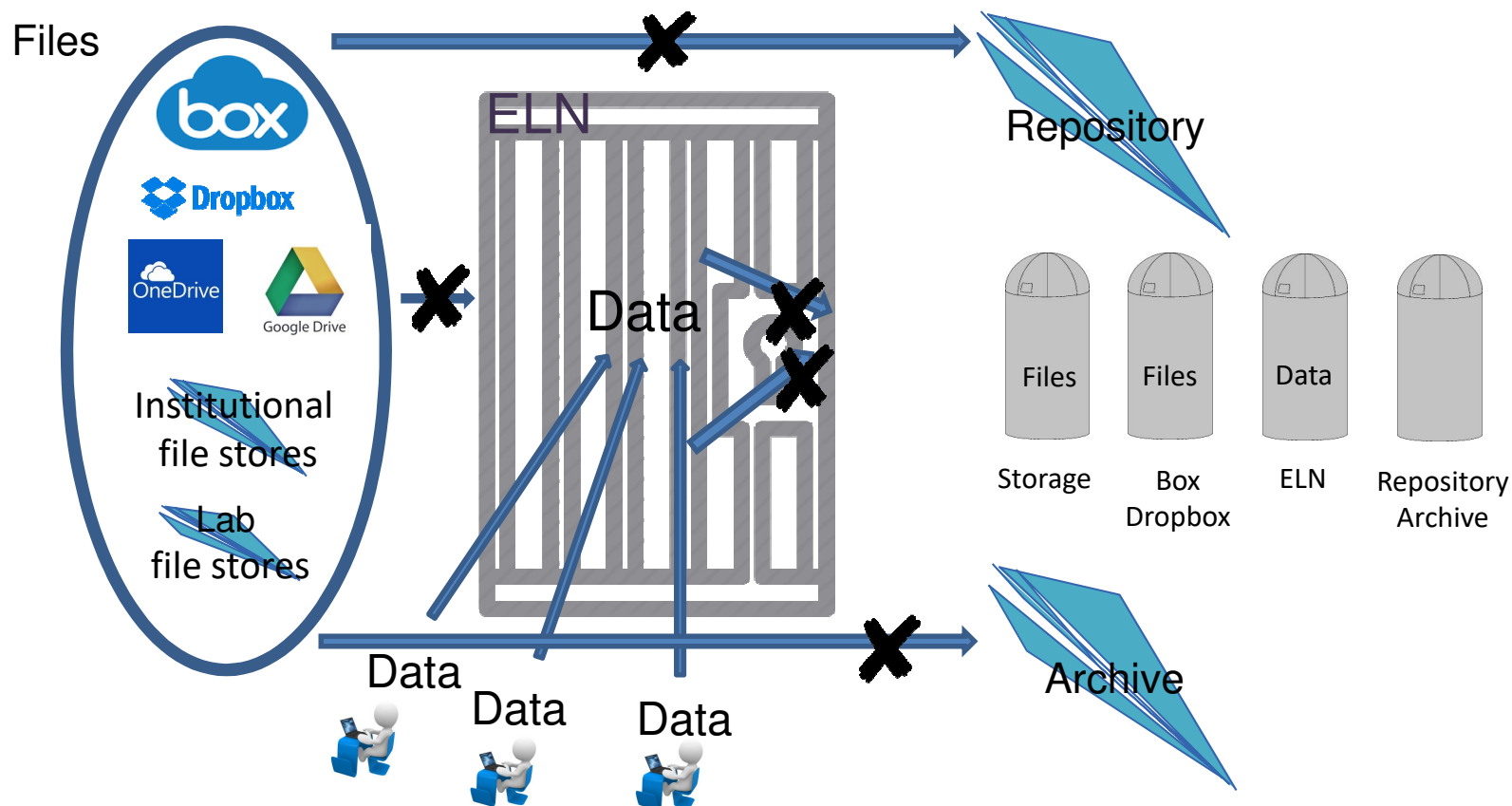data into repositories

RSpace

**Electronic lab notebooks
are ideally suited to perform this function**

- **They solve the 'data collection' problem**
- **Data is already structured in the ELN**

**But only if they are properly
constituted**

RSpace

Currently, ELNs liberate research data from paper, but don't allow links to files or data export, creating yet another silo

Files

ELN

Repository

Data

Files | Files | Data

Storage | Box Dropbox | ELN | Repository Archive

box

Dropbox

OneDrive | Google Drive

Institutional file stores

Lab file stores

Data Data Data

Archive

And leaving data and files separated and researchers and institutions unable to capture research data

RSpace is a 'connected' ELN that acts a conduit between research and publishing

- Connect to files
  - One Drive, Google Drive, Dropbox, Box
  - Institutional file stores
  - Lab file stores
- Export data
  - In a variety of formats: Word, pdf, html, xml
  - To archives, e.g. Dataverse, Figshare, DSpace when publishing a paper and/or for preservation
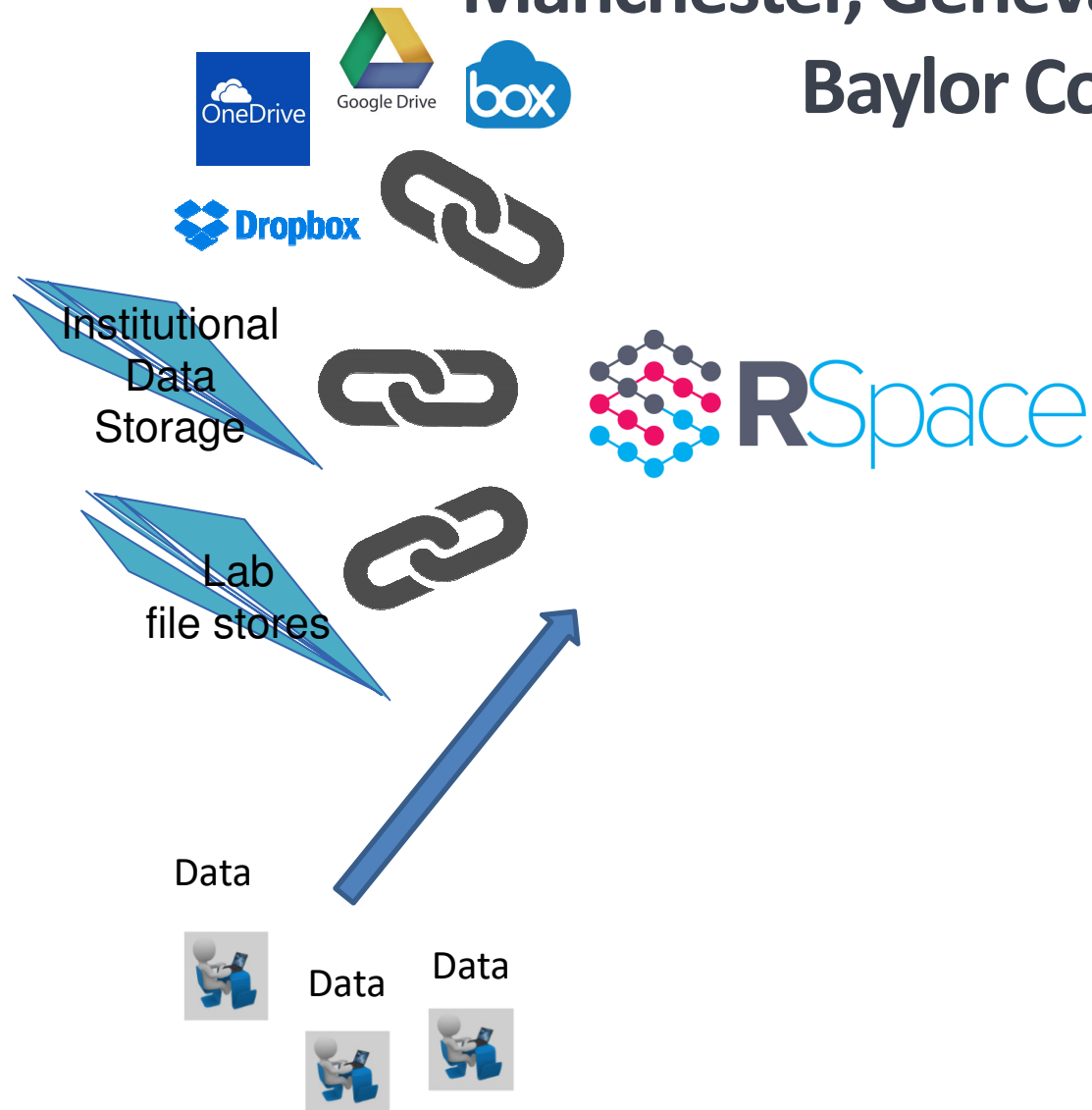
Connectedness 1

Linking to files

Why files are important

- Your data is already on them!
  - Images
  - Sequencing data
- The data is already structured
- It's not practical or desirable to move this data into the ELN
- But it's critical that it can become part of the research record documented in the ELN
- This is possible only if the ELN supports linking to files
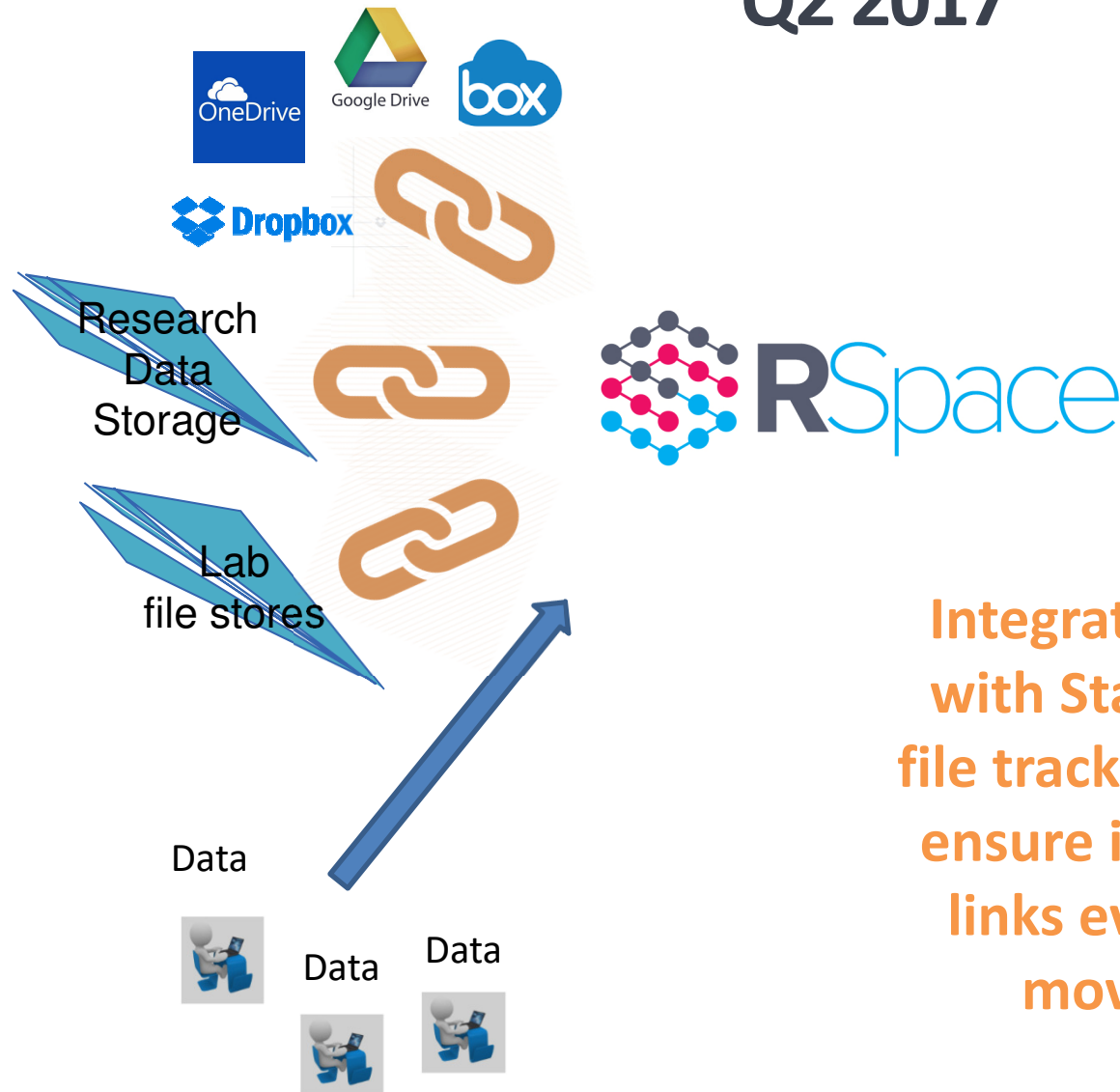- Most ELNs don't support this – it requires substantial back end and front end development

RSpace @ Goettingen, Edinburgh, Manchester, Geneva, Freiburg, Julich, Baylor College of Medicine

RSpace @ Harvard Medical School
Q2 2017

Research Data Storage

Lab file stores

Data

Data   Data

Integration of RSpace with Starfish Storage file tracking system will ensure integrity of file links even after files move location

Connectedness 2

Exporting data to archives and repositories

RSpace acts as a conduit rather than a silo

➢Powerful, configurable data export
>➢Document, folder, project, lab, system etc. level
>➢Individual user, PI, admin control
>➢Multiple formats using open standards: Word, PDF, HTML, XML
>➢Include ORCID IDs
➢Integrations with archives
>➢Easy for researchers to deposit
>➢Maintain data structure
➢Support publishing workflow
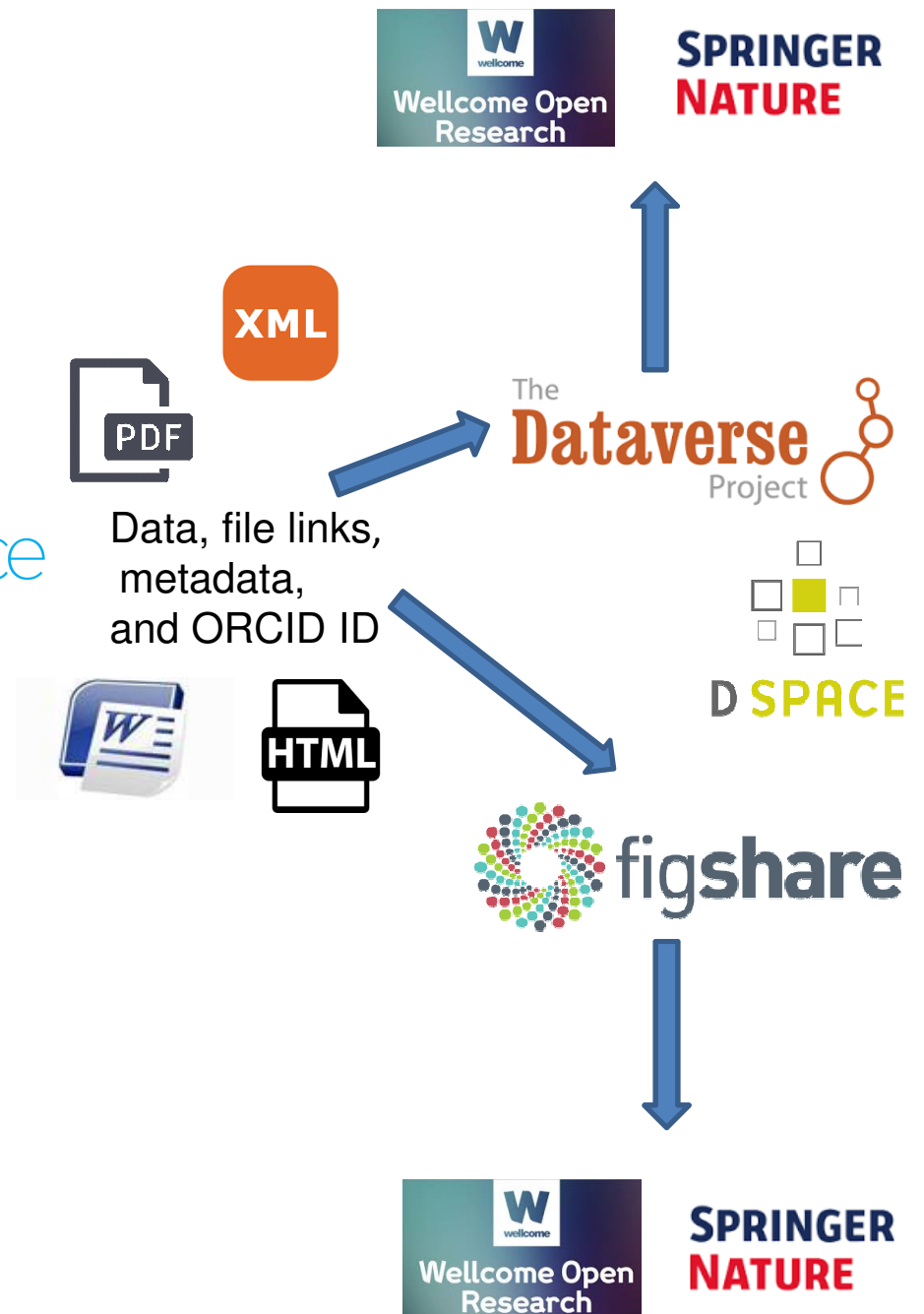
Thank you!

ngoddard@researchspace.com

www.researchspace.com

RSpace-Figshare video

RSpace-Datashare video