RESEARCH DATA ALLIANCE

Components/Services/Ideas/etc. What needs to be taken up?

Peter Wittenburg

research data sharing without barriers

rd-alliance.org

Data Science is Changing

- Accepting/converging on trends
  - Increasing volumes & complexity
  - Increasing anonymous relationship
  - Increasing sharing & re-use
  - Increasingly fed
  - and logical systems

Senior people:
It is time to change, our data work is too inefficient.
But need guidance and clarity for investments.
Need to accelerate solution finding.

- Drivers: scientific projects require advanced infrastructures
- **Is there a glue for all like with TCP/IP? – RDA a Chance**

Funders want changes

- let's use the G8 formulations – data should be
  - searchable -> create useful metadata
  - accessible -> deposit in trusted repository a
  - interpretable -> ics

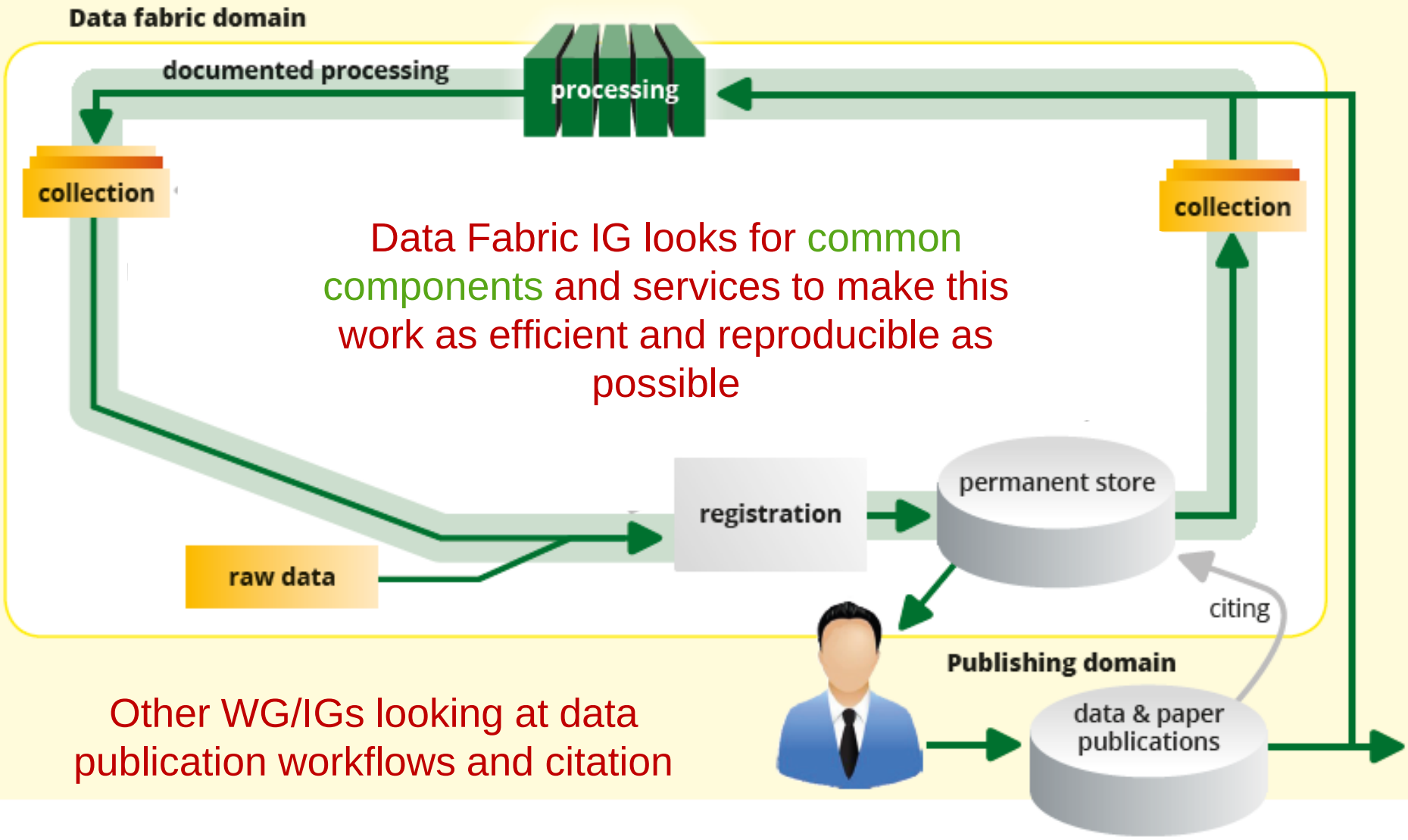
Funders:
There is a need and will to change globally and cross-disciplinary.
RDA is a chance – let's take it!

anagement Plans?

What are the consequences of these principles?

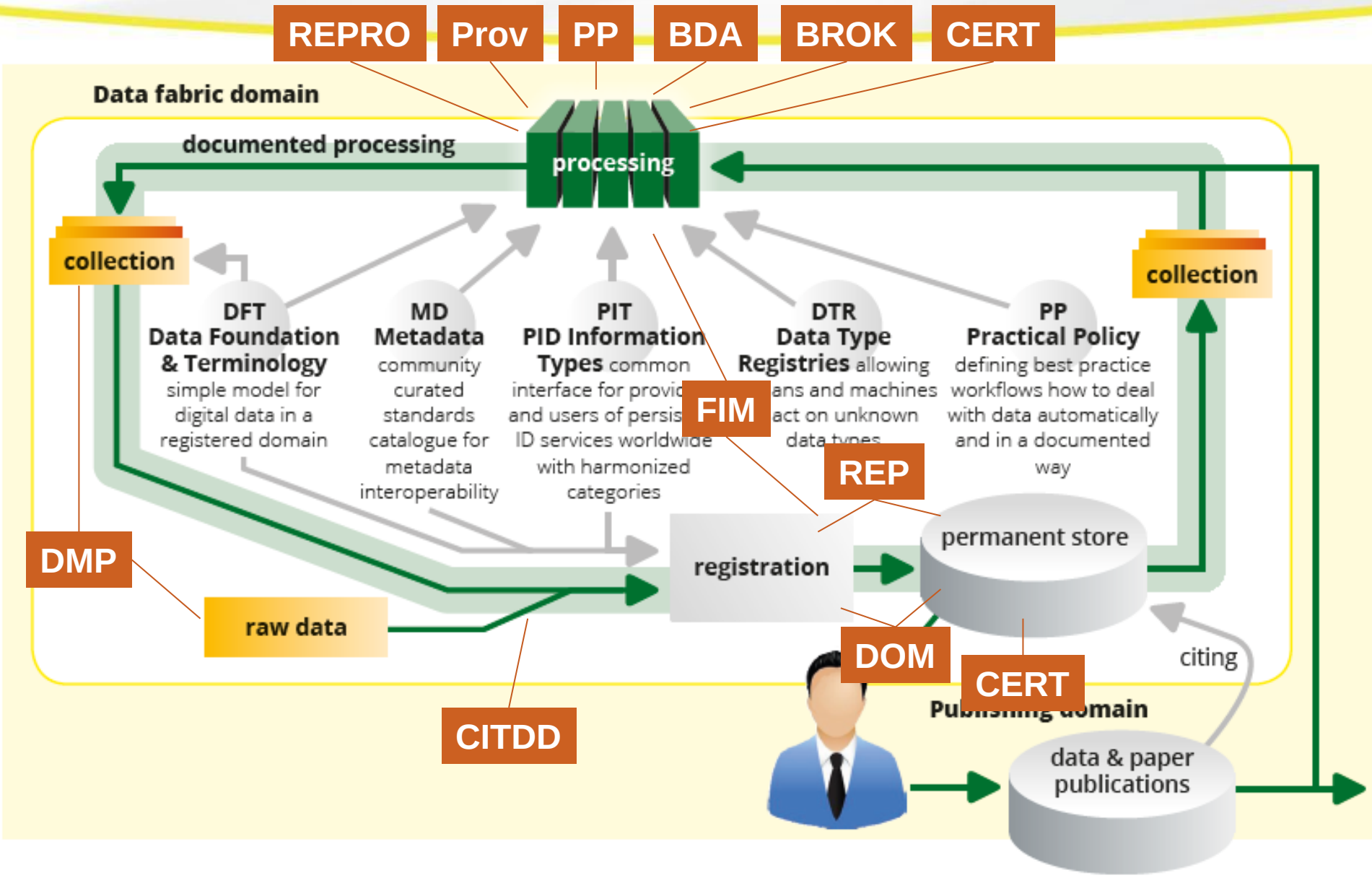
- How to design the necessary infrastructure?

Data Fabric Interest Group



Other WG/IGs looking at data publication workflows and citation

Positioning of WG/IGs



Use Cases are the basis!

	Name	Institute	state
1	Language Archive	Max Planck Institute NL	in operation
2	Geodata Sharing Platform	Academy of China	In operation
3	Datanet Federation Concortium	RENCI US	In operation
4	ADCIRC Storm Forcasting	RENCI US	In operation
5	EPOS Plate Observation	INGV/CINECA Italy	tion
6	ENVRI Environment Observation		n
7			in design
8		EPFL Switzerland	in testing
9	Climate Modeling	DKRZ Germany	In operation
10	LIGO Gravitation Physics	NCSA US	In operation
11	ECRIN Medical Trial Interoperation	U Düsseldorf Germany	In testing
12	VPH Physiology Simulation	U London UK	In operation
13	Species Archive	Nature Museum Germany	In operation
14	International NeuroI Facility	INCF Sweden	In operation
15	Molecular Genetics	MPI Germany	In operation

Participate by submitting Use Cases via Data Fabric Wiki

Position Paper "Paris.doc"

- Recently paper a number of colleagues engaged in RDA Data Management Trends, Principles and Components – What Needs to be Done?
- Co-authors don't claim to own any ideas – but kick-off a broad discussion
- Need to accelerate solution finding and convergence process

8 Common Trends	Partly stable, some still in debate
G8+ Principles	Widely agreeed
Consequences of Principles	Not really thought through
19 Components	To be discussed now
Organizational Approaches	To be discussed now

Doc: http://hdl.handle.net/11304/992fe6a0-fe34-11e4-8a18-f31aa6f4d448

Data Fabric Wiki: https://rd-alliance.org/node/44520/all-wiki-index-by-group

Components - Position Paper

1. PID System
2. Actor ID System
3. Registry S for Trusted Repositories
4. Metadata S
5.
6.
7.
8.
9. Prefabricated PP Modules
10. Distributed Authentication S
11. Authorisation Record Registry S

- OAI-PMH, ResourceSync, SRU/CQL
- Workflow Engine & Environment
- Conversion T
- try
- Training Modules

Much already out there
Why is it not used broadly?
Why is there a hesitation to broadly invest?
Need a systemic approach!
Need to start from Use Cases!

Partcipant Issues I

Issue	Comment	Action (sugg)

Let‘s include everyone willing, let‘s base on rough consensus, let‘s not repeat errors!
If we exclude groups, there will be noise!

Partcipant Issues II

Issue	Comment	Action

Partcipant Issues II

Issue	Comment	Action

Necessary Actions

- Need a kind of short shared report on the outcomes and upload this as a contribution to the DFIG discussion (kind of first response to the paper) — RDAE KO
- Need a person to synch all activities in infrastructures — RDAE KO
  - At least two levels: senior/policy level – junior/detail level
- Should form short live Interest Groups that inspire process (such as for core PID Info Types)