Storage Evolution

Context

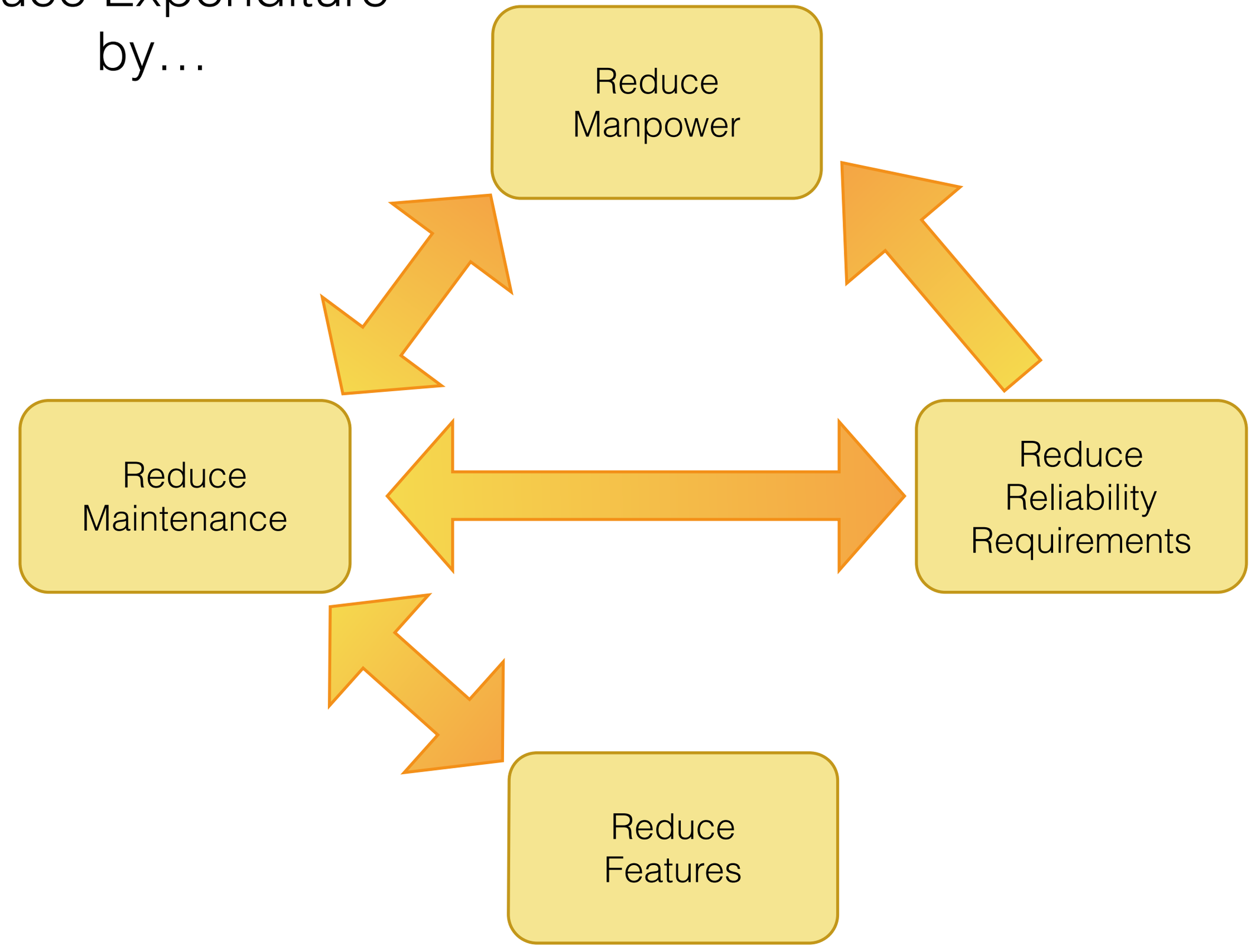
- Reduction in funded manpower
- Sites need to run with less people.
- (And need to transition with less people, too, as decline has already started)
- Sites also need to work well with non-LHC assumptions, techniques.

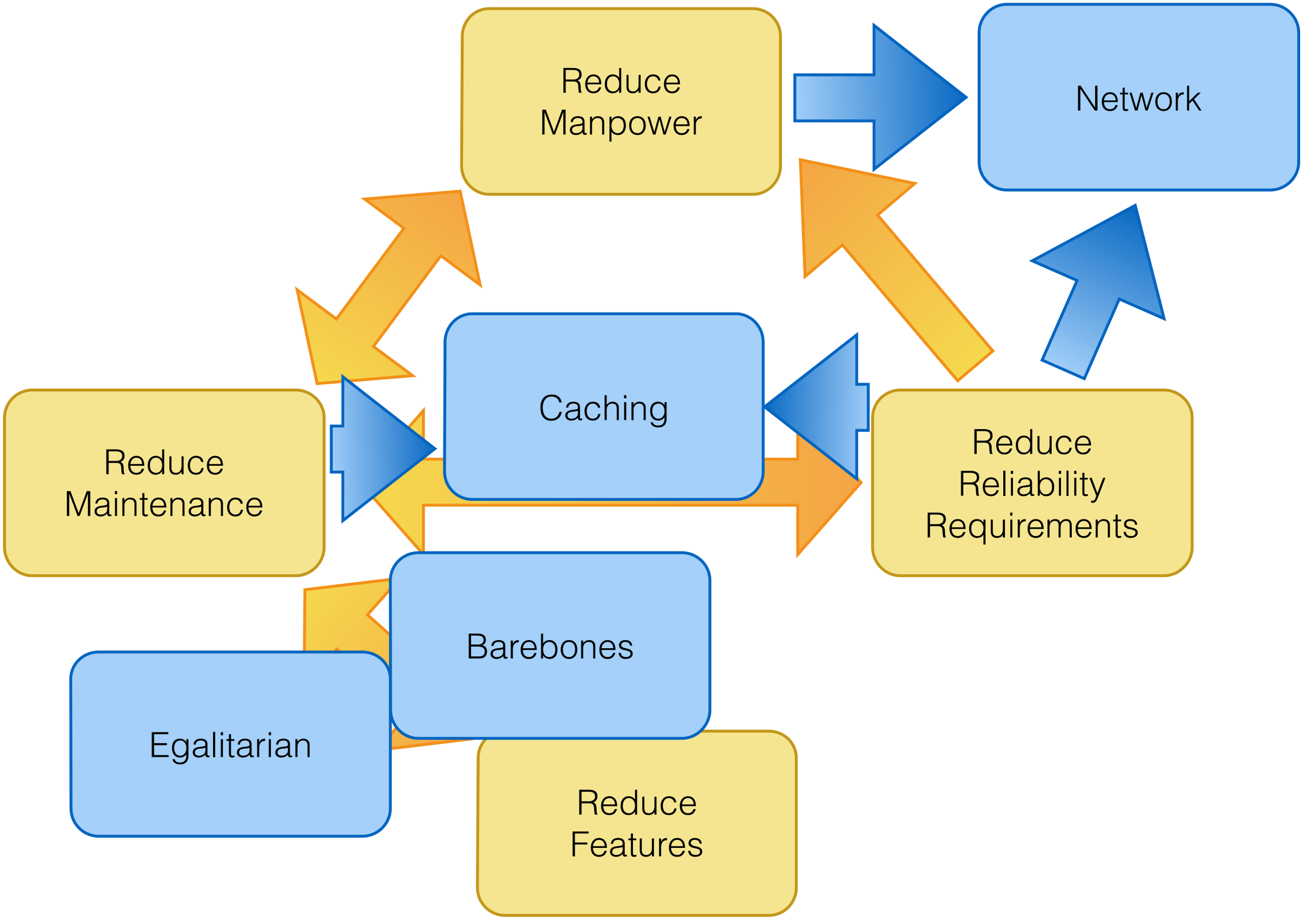
- Characteristics of T2 Storage
  - Churn (cache like?)
    - but low reuse rate
  - Large Size variation - 3PB down to <500TB
  - Access methods
    - SRM, Xroot, GridFTP, HTTP…
  - X509+VOMS authentication

Total capacity

- Unlikely that total demand for T2 storage will fall.
- What do we do with existing storage?
  - Allow to "slowly degrade" at T2Cs?
  - "Physically consolidate" at T2Ds?

Reduce Expenditure by…




Reduce
Manpower
Network
Caching
Reduce
Maintenance
Reduce
Reliability
Requirements
Barebones
Egalitarian
Reduce
Features

The dangers of specialisation


Vaporeon
Water
Jolteon
Electric
Flareon
Fire
Espeon
Psychic
Eevee
Normal
Umbreon
Dark
Glaceon
Ice
Sylveon
Fairy
Leafeon
Grass

Summer Olympics Rio 2016

	Country				
1	United States	46	37	38	**121**
2	**Great Britain**	**27**	**23**	**17**	**67**
3	China	26	18	26	**70**
4	Russia	19	18	19	**56**
5	Germany	17	10	15	**42**
6	Japan	12	8	21	**41**
7	France	10	18	14	**42**
8	South Korea	9	3	9	**21**
9	Italy	8	12	8	**28**
10	Australia	8	11	10	**29**
11	Netherlands	8	7	4	**19**
12	Hungary	8	3	4	**15**
13	Brazil	7	6	6	**19**
14	Spain	7	4	6	**17**
15	Kenya	6	6	1	**13**
16	Jamaica	6	3	2	**11**
17	Croatia	5	3	2	**10**
18	Cuba	5	2	4	**11**
19	New Zealand	4	9	5	**18**

Source: Google

- The salient lesson of the UK Olympics strategy:
    - Targeted funding is great for focussing specialisation.
    - *But harms anything outside that.*
        - UK Cycling *only* wins at the Olympics (not any other sporting events)
        - Sports that don't win at the Olympics get no money (and hence *continue* to not win anywhere).

- Monetary efficiency

Rank	Country	Weighted Medals[2]	🥇	🥈	🥉	GDP[1]	GDP per Weighted Medal[1]
1	Grenada	2	0	1	0	0.82	0.41
2	Jamaica	32	6	3	2	15.07	0.47
3	Kenya	37	6	6	1	33.62	0.91
4	Fiji	4	1	0	0	3.81	0.95
5	Armenia	10	1	3	0	10.25	1.02
6	Georgia	14	2	1	4	14.37	1.03
7	Burundi	2	0	1	0	2.33	1.16
8	North Korea	16	2	3	2	22.00	1.38
9	Bahamas	5	1	0	1	7.79	1.56
10	Tajikistan	4	1	0	0	6.52	1.63
11	Uzbekistan	27	4	2	7	45.36	1.68
12	Cuba	28	5	2	4	60.81	2.17
13	Azerbaijan	28	1	7	10	63.40	2.26
14	Croatia	28	5	3	2	63.85	2.28
15	Ethiopia	13	1	2	5	31.71	2.44
16	Serbia	18	2	4	2	45.04	2.50
17	Mongolia	3	0	1	1	8.56	2.85
18	Niger	2	0	1	0	6.02	3.01
19	Hungary	42	8	3	4	140.03	3.33
20	New Zealand	39	4	9	5	130.68	3.35
21	Belarus	16	1	4	4	55.14	3.45
22	Bahrain	6	1	1	0	21.90	3.65
23	Kosovo	4	1	0	0	18.84	4.71
24	Iran	18	3	1	4	86.53	4.81
25	CÃ´te d'Ivoire	5	1	0	1	24.07	4.81
26	Slovenia	9	1	2	1	49.54	5.50
27	Kazakhstan	31	3	5	9	186.20	6.01
28	Moldova	1	0	0	1	7.00	7.00
29	Jordan	4	1	0	0	28.84	7.21
30	Ukraine	22	2	5	4	165.25	7.51
31	Slovakia	12	2	2	0	95.99	8.00
32	Lithuania	5	0	1	3	42.73	8.55
33	Denmark	27	2	6	7	332.68	12.32
34	Bulgaria	4	0	1	2	53.51	13.38
35	Russian Federation	131	19	18	19	1857.77	14.18
36	Great Britain	171	27	23	17	2431.59	14.22
37	Czech Republic	15	1	2	7	215.22	14.35
38	Tunisia	3	0	0	3	45.86	15.29
39	Netherlands	50	8	7	4	836.26	16.73
40	Colombia	19	3	2	3	331.65	17.46
41	South Africa	22	2	6	2	408.24	18.56
42	Greece	16	3	1	2	298.73	18.67
43	Romania	9	1	1	3	179.79	19.98
44	Vietnam	6	1	1	0	123.96	20.66
45	Australia	64	8	11	10	1371.76	21.43
46	South Korea	51	9	3	9	1116.25	21.89
47	Estonia	1	0	0	1	22.18	22.18
48	Trinidad and Tobago	1	0	0	1	22.48	22.48
49	Puerto Rico	4	1	0	0	92.61	23.15
50	Sweden	23	2	6	3	538.13	23.40
51	Thailand	14	2	2	2	345.65	24.69
52	Poland	20	2	3	6	514.50	25.72
53	France	90	10	18	14	2773.03	30.81
54	Malaysia	9	0	4	1	278.67	30.96
55	Argentina	14	3	1	0	445.99	31.86
56	Italy	64	8	12	8	2194.75	34.29
57	Germany	103	17	10	15	3570.56	34.67
58	Switzerland	18	3	2	2	635.65	35.31
59	Spain	42	7	4	6	1490.81	35.50
60	Belgium	14	2	2	2	511.53	36.54
61	China	166	26	18	26	7298.10	43.96
62	Canada	37	4	3	15	1736.05	46.92
63	Algeria	4	0	2	0	188.68	47.17
64	United States	296	46	37	38	15094.00	50.99
65	Brazil	46	7	6	6	2476.65	53.84

Source: www.medalspercapita.com

- Monetary efficiency

Rank	Country	Weighted Medals[2]	🥇	🥈	🥉	GDP[1]	GDP per Weighted Medal[1]
1	Grenada	2	0	1	0	0.82	0.41
2	Jamaica	32	6	3	2	15.07	0.47
3	Kenya	37	6	6	1	33.62	0.91
4	Fiji	4	1	0	0	3.81	0.95
5	Armenia	10	1	3	0	10.25	1.02
6	Georgia	14	2	1	4	14.37	1.03
7	Burundi	2	0	1	0	2.33	1.16
8	North Korea	16	2	3	2	22.00	1.38
9	Bahamas	5	1	0	1	7.79	1.56
10	Tajikistan	4	1	0	0	6.52	1.63
11	Uzbekistan	27	4	2	7	45.36	1.68
12	Cuba	28	5	2	4	60.81	2.17
13	Azerbaijan	28	1	7	10	63.40	2.26
14	Croatia	28	5	3	2	63.85	2.28
15	Ethiopia	13	1	2	5	31.71	2.44
16	Serbia	18	2	4	2	45.04	2.50
17	Mongolia	3	0	1	1	8.56	2.85
18	Niger	2	0	1	0	6.02	3.01
19	Hungary	42	8	3	4	140.03	3.33
20	New Zealand	39	4	9	5	130.68	3.35
21	Belarus	16	1	4	4	55.14	3.45
22	Bahrain	6	1	1	0	21.90	3.65
23	Kosovo	4	1	0	0	18.84	4.71
24	Iran	18	3	1	4	86.53	4.81
25	CÃ´te d'Ivoire	5	1	0	1	24.07	4.81
26	Slovenia	9	1	2	1	49.54	5.50
27	Kazakhstan	31	3	5	9	186.20	6.01
28	Moldova	1	0	0	1	7.00	7.00
29	Jordan	4	1	0	0	28.84	7.21
30	Ukraine	22	2	5	4	165.25	7.51
31	Slovakia	12	2	2	0	95.99	8.00
32	Lithuania	5	0	1	3	42.73	8.55
33	Denmark	27	2	6	7	332.68	12.32
34	Bulgaria	4	0	1	2	53.51	13.38
35	Russian Federation	131	19	18	19	1857.77	14.18
36	Great Britain	171	27	23	17	2431.59	14.22
37	Czech Republic	15	1	2	7	215.22	14.35
38	Tunisia	3	0	0	3	45.86	15.29
39	Netherlands	50	8	7	4	836.26	16.73
40	Colombia	19	3	2	3	331.65	17.46
41	South Africa	22	2	6	2	408.24	18.56
42	Greece	16	3	1	2	298.73	18.67
43	Romania	9	1	1	3	179.79	19.98
44	Vietnam	6	1	1	0	123.96	20.66
45	Australia	64	8	11	10	1371.76	21.43
46	South Korea	51	9	3	9	1116.25	21.89
47	Estonia	1	0	0	1	22.18	22.18
48	Trinidad and Tobago	1	0	0	1	22.48	22.48
49	Puerto Rico	4	1	0	0	92.61	23.15
50	Sweden	23	2	6	3	538.13	23.40
51	Thailand	14	2	2	2	345.65	24.69
52	Poland	20	2	3	6	514.50	25.72
53	France	90	10	18	14	2773.03	30.81
54	Malaysia	9	0	4	1	278.67	30.96
55	Argentina	14	3	1	0	445.99	31.86
56	Italy	64	8	12	8	2194.75	34.29
57	Germany	103	17	10	15	3570.56	34.67
58	Switzerland	18	3	2	2	635.65	35.31
59	Spain	42	7	4	6	1490.81	35.50
60	Belgium	14	2	2	2	511.53	36.54
61	China	166	26	18	26	7298.10	43.96
62	Canada	37	4	3	15	1736.05	46.92
63	Algeria	4	0	2	0	188.68	47.17
64	United States	296	46	37	38	15094.00	50.99
65	Brazil	46	7	6	6	2476.65	53.84

Source: www.medalspercapita.com

- Kenya: existing culture of excellence at long distance running.

- Fiji: existing culture of excellence at rugby.

- Jamaica: nucleus around single "superstar"

- Don't need to "push" performance by additional investment.

- Take advantage of existing skills, tools etc.

- Some existing solutions..

Caching /
Read only

Data CVMFS

- Read only data provision, optimised for small, slowly changing datasets.
- CVMFS + (Xrootd backhaul) + [future cache tiers]
- Deployed for LIGO US by OSG - tested in production.
- Some exploration at RAL for nonLHC VOs
- Not suitable for rapidly changing, large datasets.
- Easy transition (for all sites with CVMFS up to date…)

Caching /
Read only

T2C Caching

- WLCG Data Caching Group loosely "coordinating".
- Two solutions:
  - Xrootd - mature, not widely used?, compatible with xrootd federation
  - "DPM" - conceptual, in development, protocol agnostic, but DPM specific.
- Easy transition (for sites with DPM or Xrootd services)

Caching /
Read only

ARC Cache

- ARC "prefetch" cache for jobs (see previous talks)
  - Working happily at Durham for some months now.
  - No movement visible on Rucio integration testing (next phase to) there.
- Easy transition (for sites with ARC CEs)

Questions on caching

- None of the mentioned solutions solve job *output* issues.
- Is caching efficient?
  - Small scale studies suggest 90% of data is read maybe twice, at best.
  - Caching as an proxy for acceptance of low reliability?
  - How is this different from redirection, in practice?

Pure network model

- Tests by Alastair Dewhurst against Oxford T2.
- Slowed by departure of Ewan & by issues with HammerCloud's assumptions re data locality.
- Obviously, only test ATLAS workloads…

ATLAS T2C Setup

- Can we run a significant range of ATLAS workflows at a CPU only sites?
  - Theoretically yes, if it has good network connectivity.
- Want to test FAX to see if we can read from many sites?

Action	Requirement
Read from specific site	AGIS configuration: '**se**', '**copyprefix**', '**ddm**' (example: UCL)
Read from many sites	FAX. AGIS configuration: '**wansinklimit**', '**wansourcelimit**'
Write to specific site	AGIS configuration: '**se**', '**copyprefix**', '**ddm**'
Write to different sites	Alternative stage out as part of pilot 2.0 development.

Test Setup

- Configure ANALY_OX with:
  - Wansinklimit: 3Gb/s
- Submit Derivation jobs (Relatively I/O intensive)
- Functional test have been run:
  - To Oxford, data at Oxford
  - To QMUL, data at QMUL
  - To QMUL, data at QMUL
- Brokerage (mostly) works.
  - Working with Hammer Cloud developers to be able to scale up.
- New monitoring being developed:
  - http://dashb-wdt-xrootd.cern.ch/ui/#vo=(atlas)

Simpler storage

- Current T2 storage is specialist - big RAID-6 arrays.
- Work by Marcus on ZFS, software 'RAID':
  - faster, more flexible, more extensible than HW RAID
  - small efficiency gains via compression (~4%)
- Small step (server local changes).
- (Also heads off RAID6 scaling limit @ 8TB disks)

Barebones
Egalitarian

SRM retirement?

- Slow movement on this, due to existing dependancies.
- Agreement "in principle" that not necessary for T2 workflows
  - but spacetokens still an issue.
- DPM "DOME" work makes this more pressing (DOME accounting orthogonal to SRM accounting)
- Still no good, general-purpose alternative for protecting/guaranteeing VO space.

Barebones
Egalitarian

Single protocol?

- HTTP(S)/WebDAV?
  - HTTP Deployment taskforce signed off on features being complete…
- XROOT? [or xrootd]
  - "unusual", but xrootd also supports HTTP(S)
- GridFTP?
  - Advantage(?) of "Globus Connect" compatibility

Barebones
Egalitarian

SE "retirement"

- Without SRM, what is the need for a traditional SE?
- Xrootd, HTTP, GridFTP endpoints in front of:
  - ceph/hdfs/other distributed storage [~posix]
- Object storage via [xrootd/S3/…]
- Transition is *hard*.
  - (Unless we can dump all of the data at a site…)

Barebones
Egalitarian

Interoperability

- GridPP perspective
  - DIRAC support?
  - Authentication/Authorisation
    - all three protocols have pluggable support
- WLCG level
  - WLCG Experiment support
  - FTS…

Wider projects

- Indigo Datacloud
  - release #1 - MidnightBlue
- dCache is only "LHC Grid Storage" member
- Token Translation Service is interesting
- General Cloud Storage focus

Funding Signals

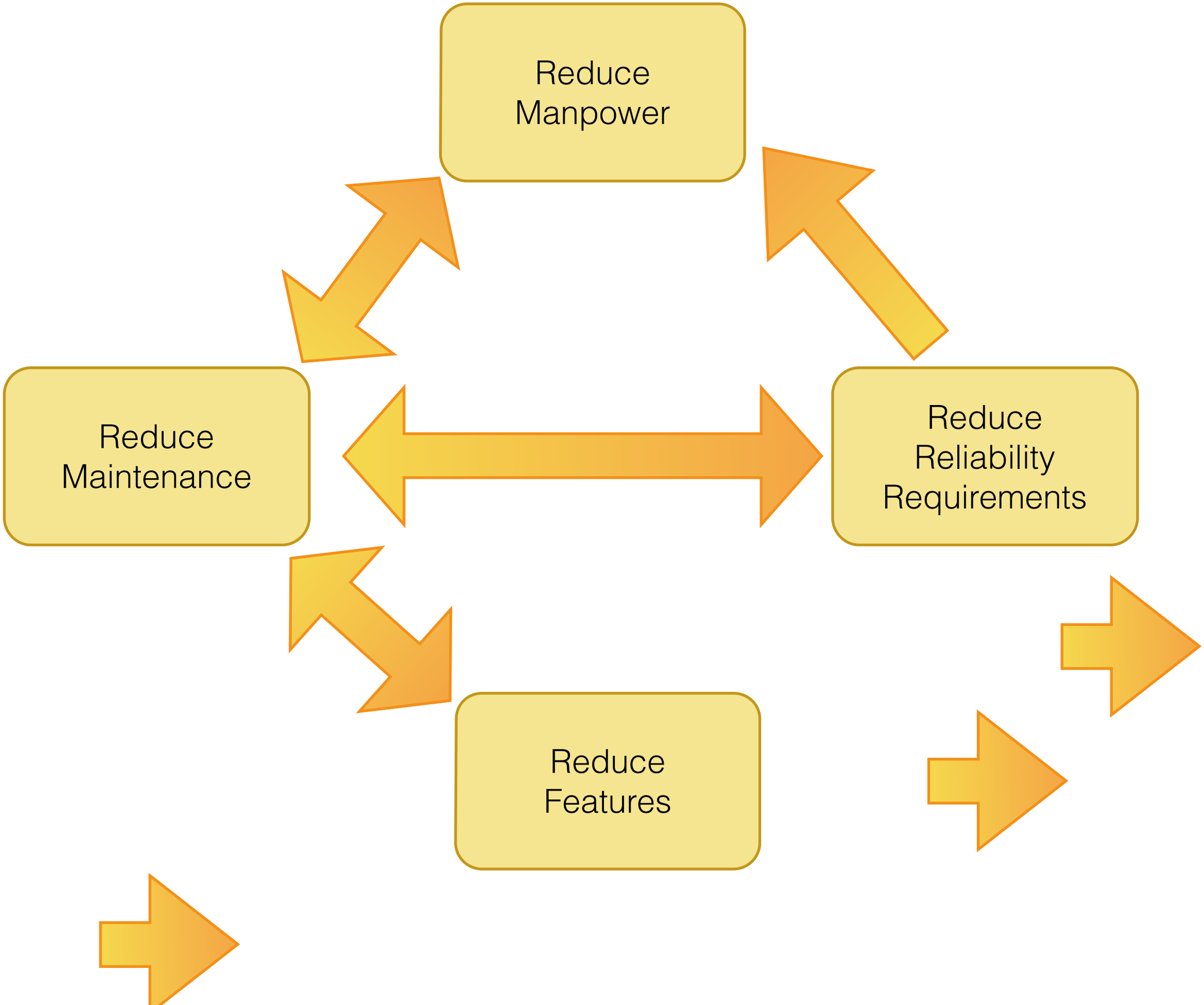
- Returning to the UK Olympics:
  - Funding is the most important signal in project direction.
  - Specialisation will adapt to improve access to funding.

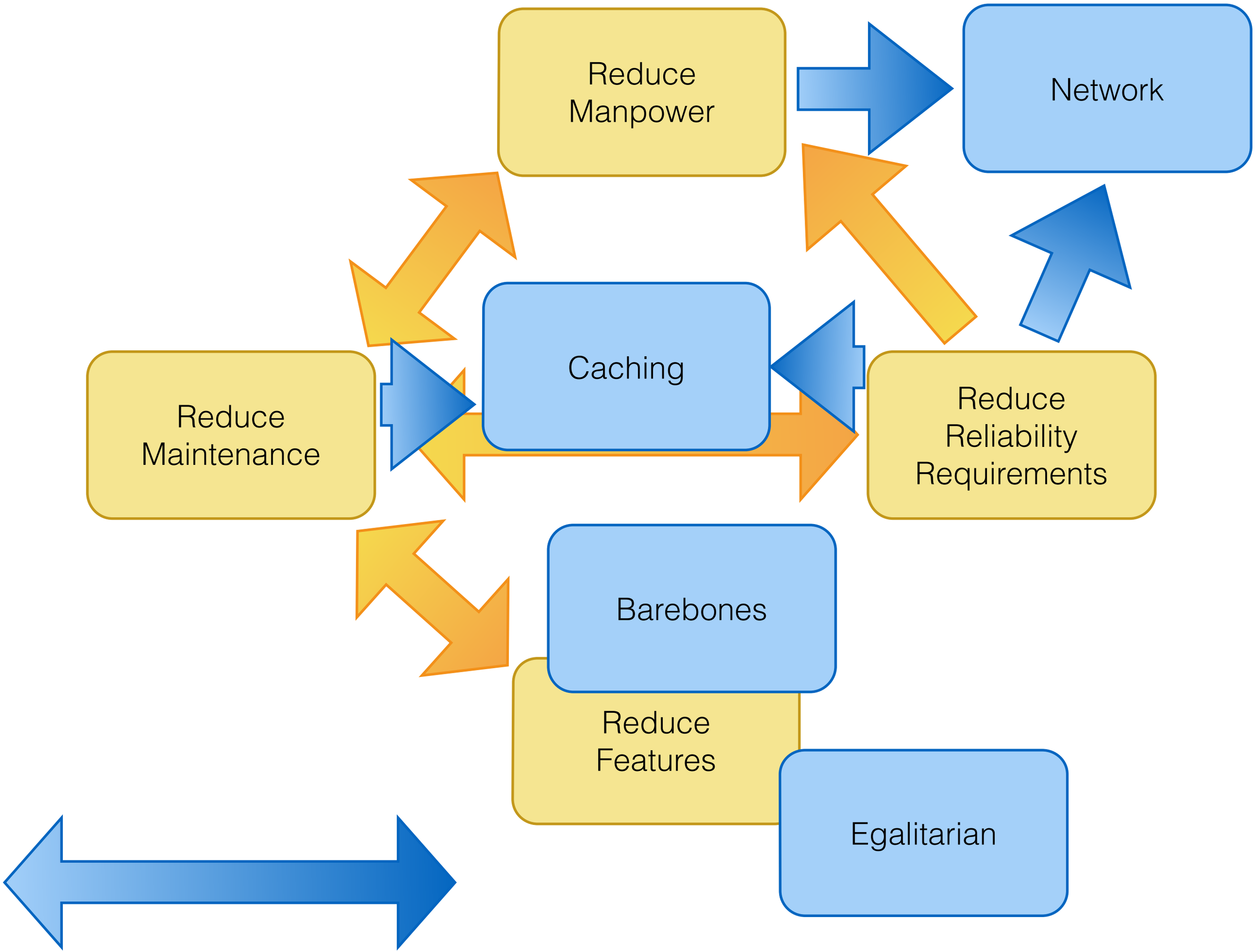
Funding Signals

- Questions for the PMB:
  - Will cache/non-reliable storage count towards pledge?
  - Will cache/non-reliable storage count towards funding?
    - (Remote access to disk is to be guaranteed by policy?)

- Points for discussion:
  - "UK" Xrootd Federated Pilot [+Caching?]
  - Provision for non-WLCG VOs
  - Simpler sites - strategies
    - low-cost (transition to zfs)
    - Consolidation?
- Funding model
- **Path of transition from here to there.**

- Backup Slides


Reduce
Manpower
Reduce
Maintenance
Reduce
Reliability
Requirements
Reduce
Features


Reduce
Manpower
Network
Caching
Reduce
Maintenance
Reduce
Reliability
Requirements
Barebones
Reduce
Features
Egalitarian