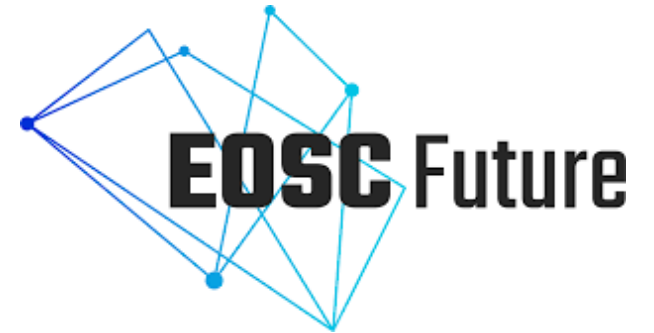




SFI RESEARCH CENTRE
IN APPLIED GEOSCIENCES



IRLDAT GeoData Use Case

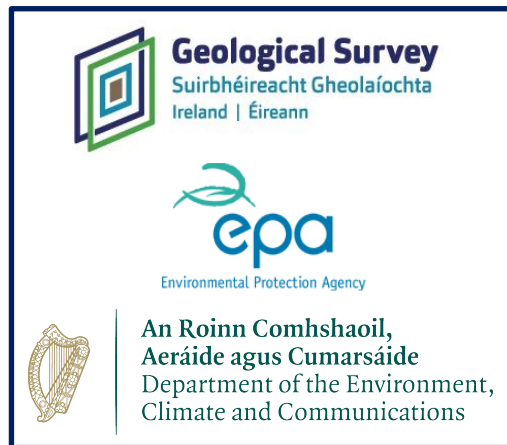
Chris Burbidge

iCRAG, University College Dublin

IRLDAT SAF 2023-12-08

A world-leading SFI research centre

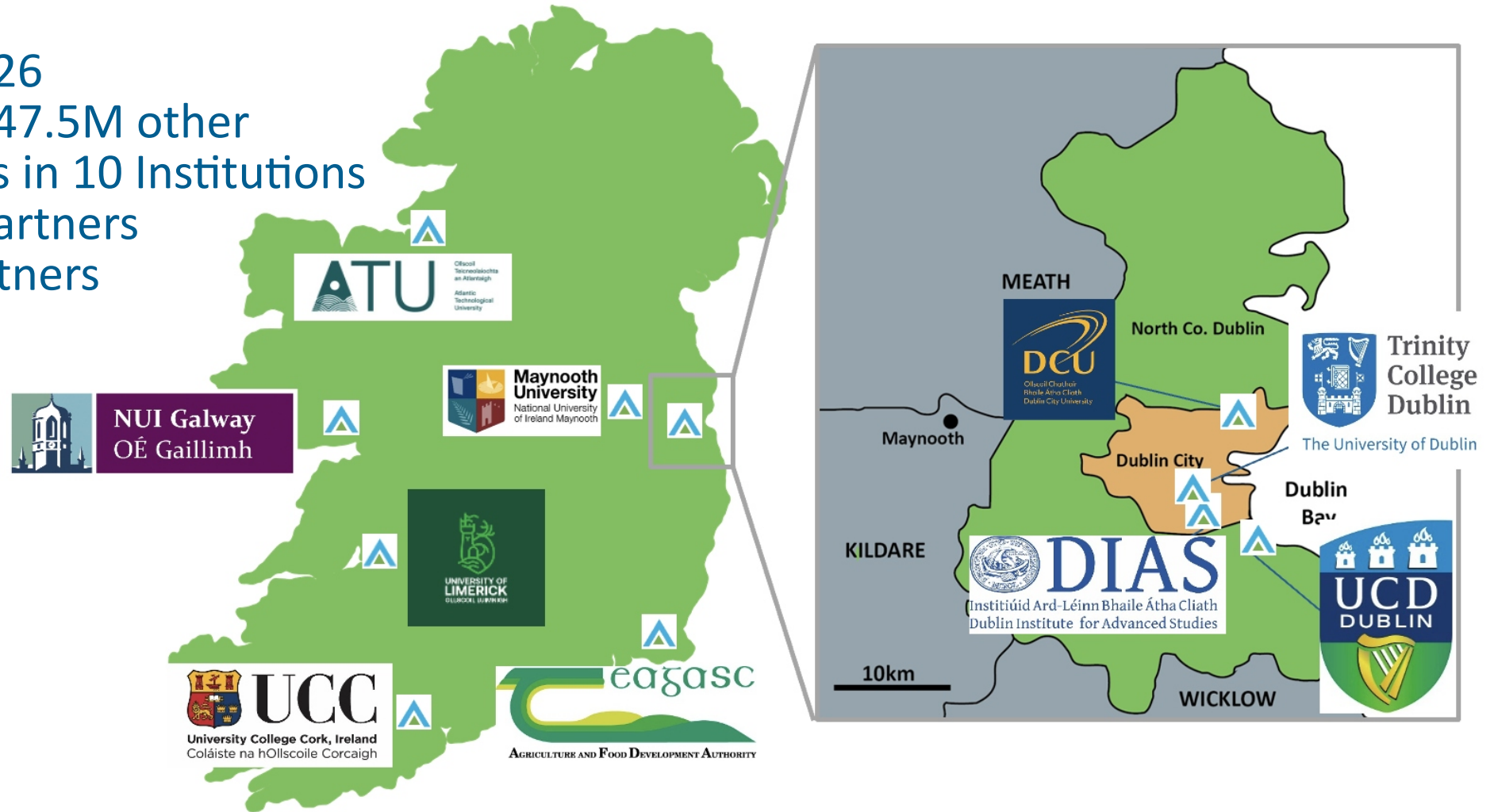
iCRAG2 2021-2026
€22.5M core + €47.5M other
150+ researchers in 10 Institutions
3 Government Partners
20+ Industry Partners
... and growing



Geological Survey
Suirbhéireacht Gheolaíochta
Ireland | Éireann

epa
Environmental Protection Agency

An Roinn Comhshaoil,
Aeráide agus Cumarsáide
Department of the Environment,
Climate and Communications



iCRAG Geodata Platform

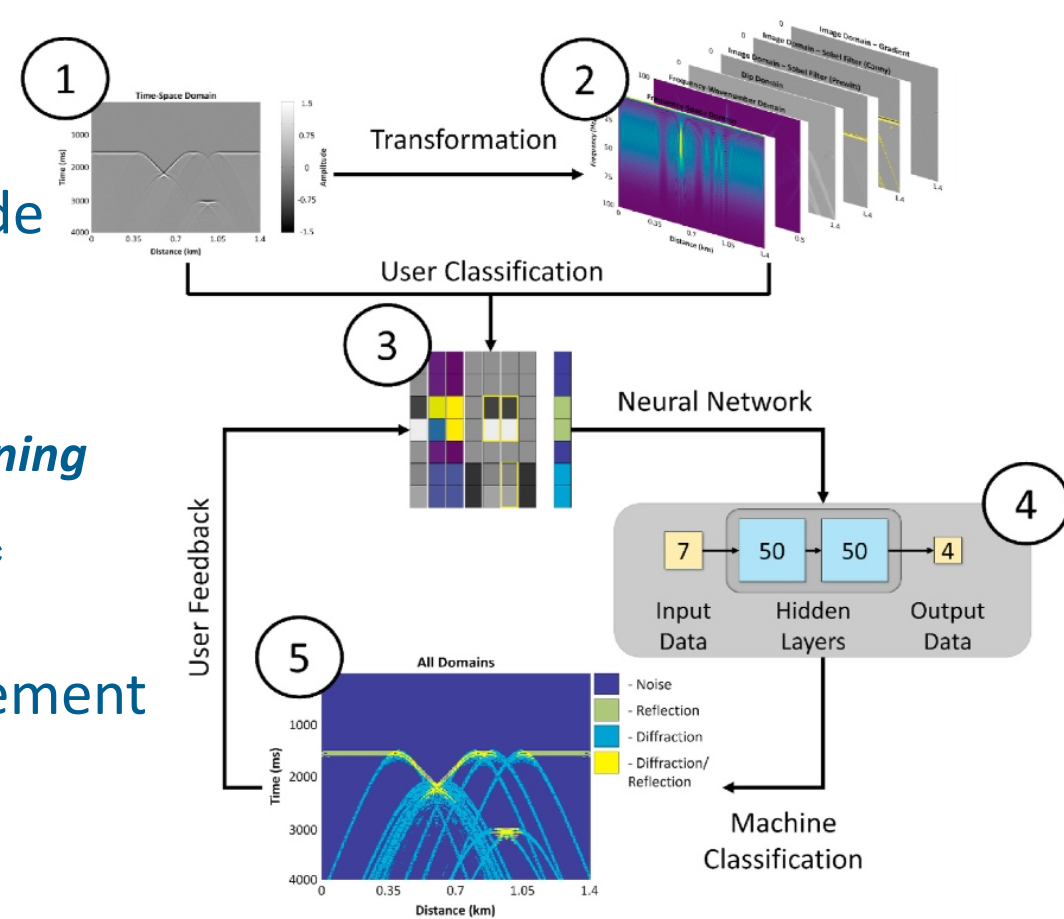
Geodata Platform underpins activities centre-wide

Patterns in data drive understanding of the data:

big data • mathematical modelling • data analytics learning

Improve on traditional 'model-driven' approach*

Increasingly now data infrastructure and management



Lowney et al., Computers and Geoscience
155, 2021, 104845

iCRAG Data Infrastructure Planning

Complex, diverse, needs around the centre
Community is the basis for the system

Larger 'simpler' projects can self-organise
Make the system scalable for the future

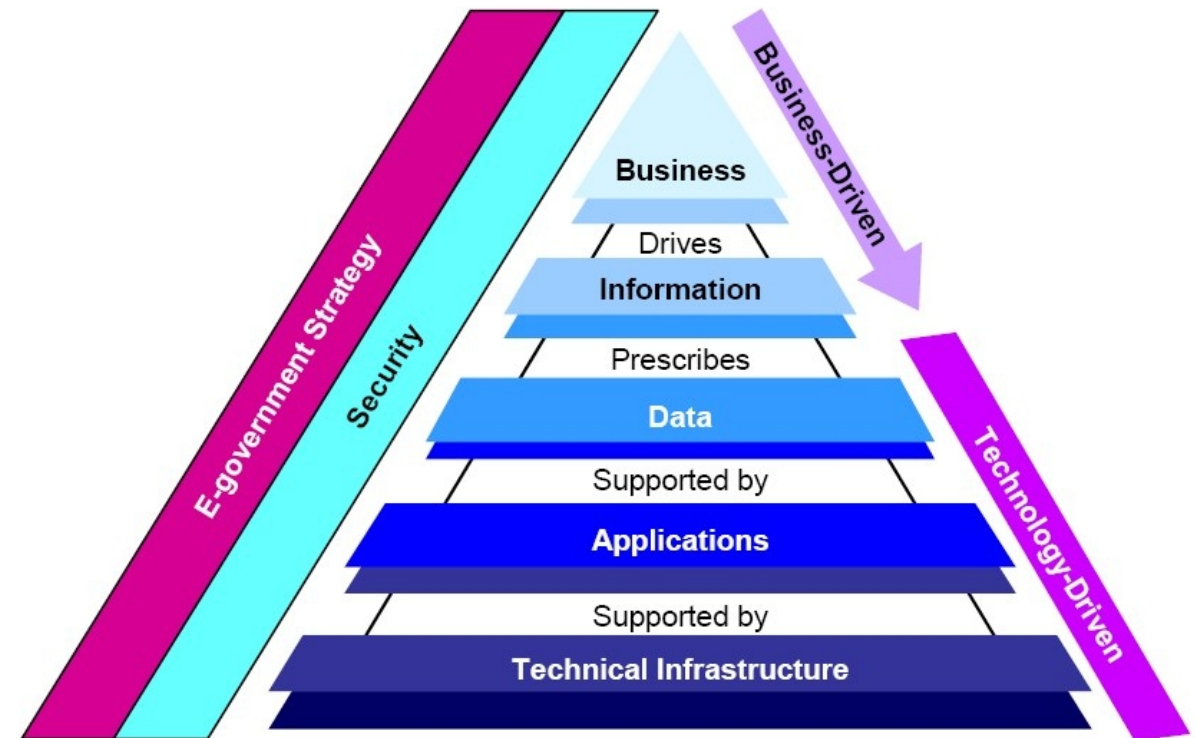
Largest not presently well catered for
Align the system with larger initiatives

How to organise support?

Business defines needs

Technology defines possibilities

Data provides linkage for Enterprise



Enterprise Architecture

Shells of linked Interests

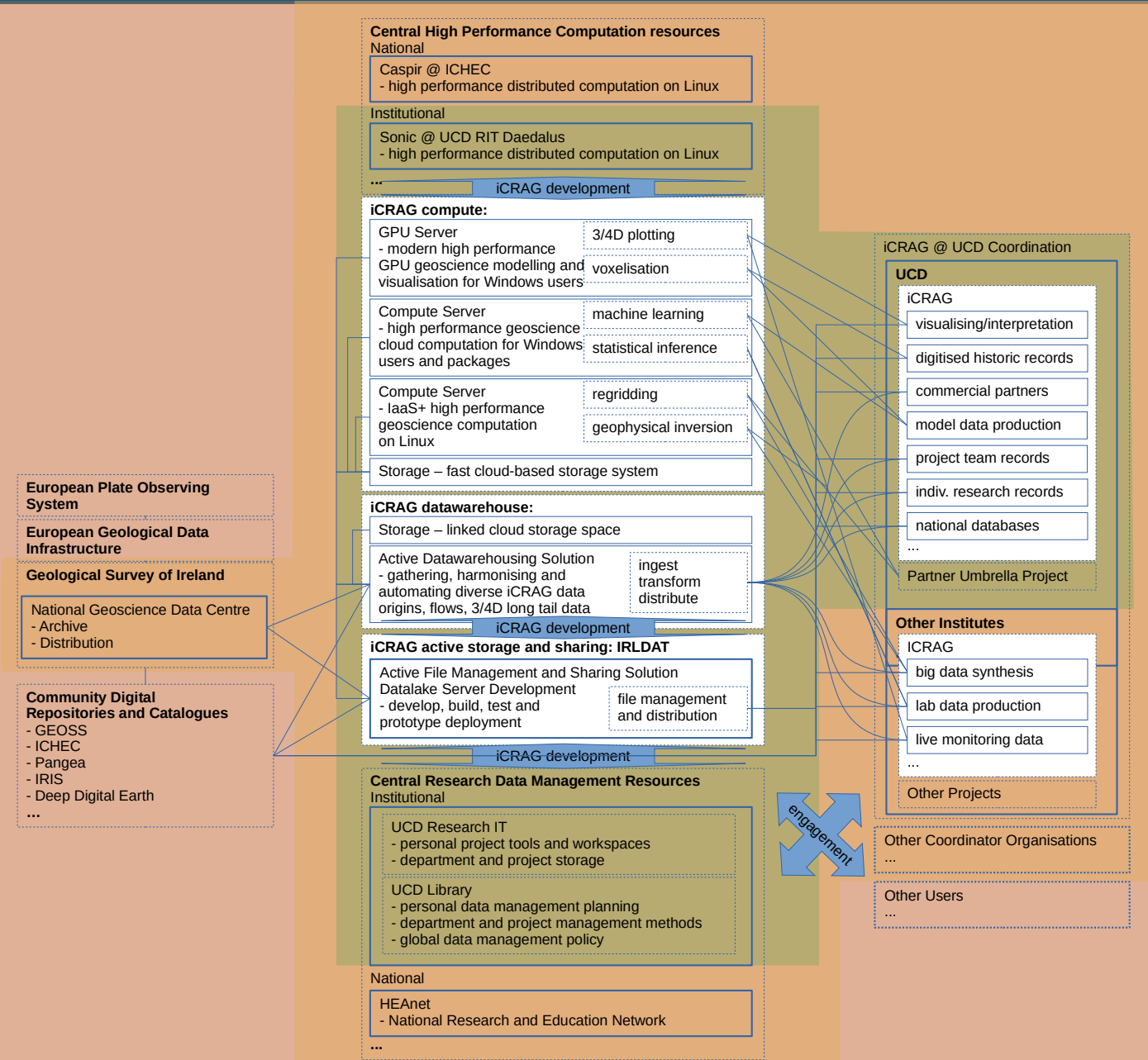
- individuals
- centre
- institutional
- national
- international

Making datasets citable

Heterogeneous Information Network

Three infrastructure foci

- accessible compute and GPU
- datawarehousing
- **object storage & file sharing**



Object Storage and File Sharing

IRLDAT - Piloting Irish Active Data Storage

B2Drop (+ publication via Share in EUDAT...)

Active file management production need was est ~300TB
+ Making cold storage more accessible would mean Pbs



An all-iCRAG federated cloud based approach?



<https://www.cultureslate.com/lists/u7a57oxsczugt8bgwxhkrk6uhi1up2>

iCRAG IRLDAT Overview

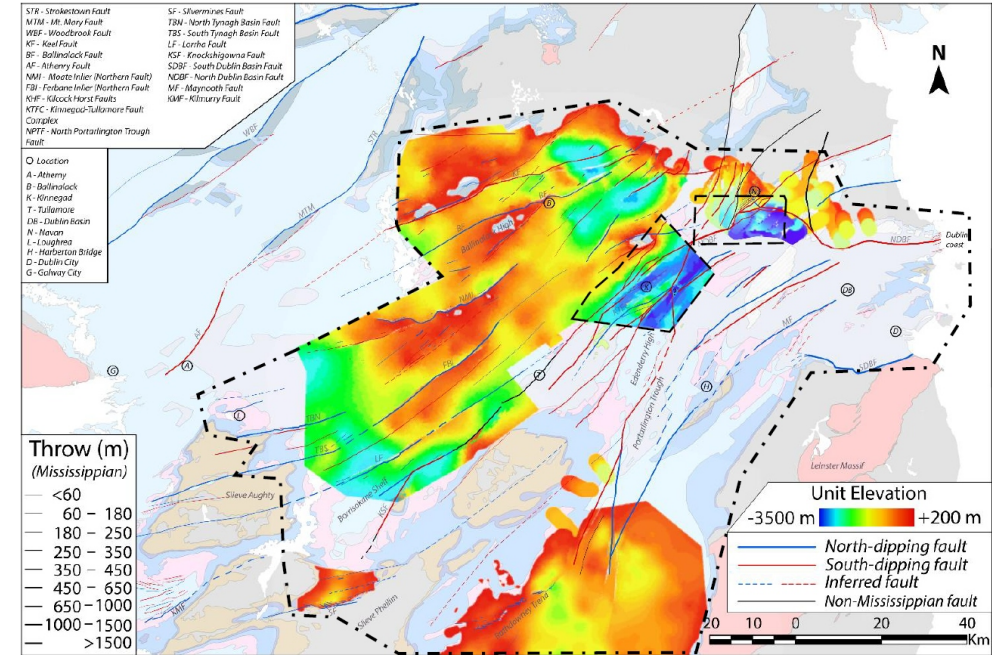
IRLDAT Phase 1 - iCRAG on EUDAT

1. Central folder structure: administration
2. Federated accounts: registration, setup
3. Larger projects: EUDAT quotes + DMPs

IRLDAT Phase 2 - iCRAG case hosted by ICHEC on AWS S3

1. Folder structure recreated + B2Access groups.
2. Smaller/simpler project data synced down from EUDAT and other sources to iCRAG@UCD desktop, and up to IRLDAT using Nextcloud client.
3. Larger & new project groups being onboarded for direct sync to IRLDAT using Nextcloud client (and down to iCRAG@UCD).

- *dedicated vm in iCRAG@UCD, for UCD data exchanges*



Torremans, K. 2023. iCRAG Challenge Project. Area to be combined in 3D and volume models.

iCRAG IRLDAT Use Case Examples

- Marine Blue Carbon modelling	M Chatting	97MB synced
- iCRAG pXRF Database	E Burton	0.5GB synced
- Level 0 Geophysical Modelling	I Lokmer	to be onboarded and synced
- National Geothermal DB	K Torremans	to be onboarded and synced
- EOSC Synthetic Datasets Contract	Burbidge, Crowley	32GB synced growing to 1TB external company onboarded
- National Petrophysics DB	A Melo	120GB growing to 1TB to be onboarded and synced
- Earth:Chec ICHEC EuroCC2 Digital Flagship	English, Kanaan, Burbidge	Project shared working environment. Team being onboarded.

EOSC Future Synthetic Datasets

European Open Science Cloud

- experience in geoscience area and market

iCRAG community and institution

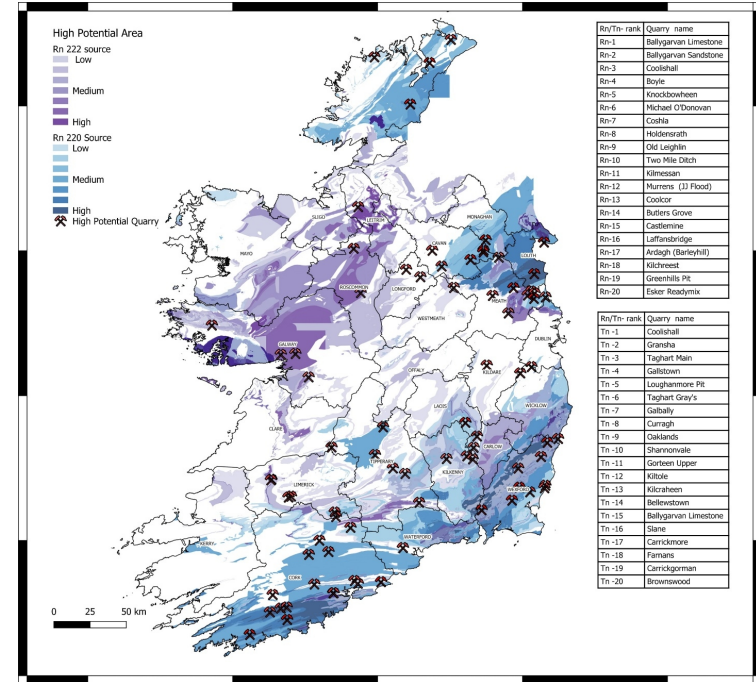
- experience in EOSC mechanisms inc. voucher
- new radiometric and lidar datasets support BSS

Geochron Ltd – contract + EOSC marketplace

IRLDAT – substantial inter-organisational use case

To date:

- 32 GB geodata file mix sync, including >GB images and zips
- Windows – B2Drop – Ubuntu via Nextcloud client with Single Sign On
- 1st dataset provided and reported on, 5 more currently being set up



Radon and Thoron potential and Irish quarries for EOSC data supply prioritisation. Geochron Ltd., with permission



Earth:Chec - Metadata and Architecture

- Datawarehousing needed to efficiently access **The Long Tail**
- Challenging to resource & in-house work needed anyway
 - Build, harmonise, verify the metadata foundation

FAIR ≈ Machine Readable

EARTH:CHEC

- ICHEC Euro CC2 Digital Innovation Flagship project
- iCRAG and ICHEC staff with GSI and UCD
 - Mapping and Cataloguing Data Resources
 - Formalised planning and cross checks
 - Harmonising use of terminology
 - “Enterprise Architecture”
 - Position papers
 - Prepare for wider discussion and development

	A	B	F	G	H	I	J	K	L	M	N	O
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												
32												
33												
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												



iCRAG IRLDAT Nextcloud Phase 2 Experience

Folder administration works once understood: connect a user to a folder with many permissions then revoke rather than just adding

Better not move a big sync once created!

... but Nextcloud sorts itself out impressively given time / bandwidth

Integration with UCD and so nationwide single sign on is really great in many ways, but login via separate username and password could be more stable.

‘Tricks’ around login could disrupt access for inexperienced users

- Ubuntu: repeated requests to login via browser – resolved, but
- Windows: login (via SSO) can also give repeated pop-up, also on EUDAT.

Where nextcloud-account link has not broken editing the address in the browser down to <https://b2drop.irdat.ichec.ie> gets a user into their folders.

Multiple accounts for one (super)user complicate this...

Conclusions and Prospects

Storage and Sharing being actively developed

- HEAnet and ICHEC engagement
- EOSC, EUDAT, EuroCC2, NORF projects driving each other
- moderate scale IRLDAT deployment
- all data types of interest
- largely as backup
- seeking funding to sustain and contribute to national infrastructure

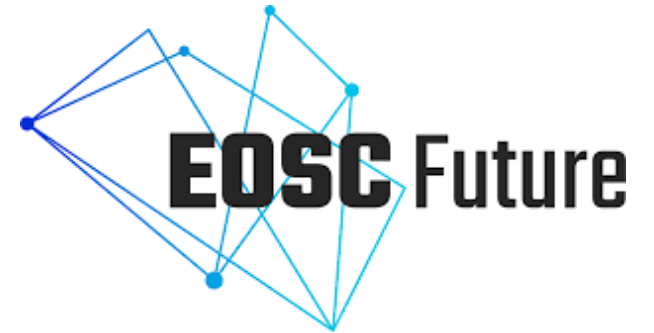
Metadata - linking stakeholder interests and architecture

Datawarehousing - the main challenge for FAIR: ideas laid out and evolving

*key is that we know enough to know what we need
requires activity in all the EA shells of linked interests*



SFI RESEARCH CENTRE
IN APPLIED GEOSCIENCES



IRLDAT GeoData Use Case

Chris Burbidge

iCRAG, University College Dublin

IRLDAT SAF 2023-12-08