LOOP

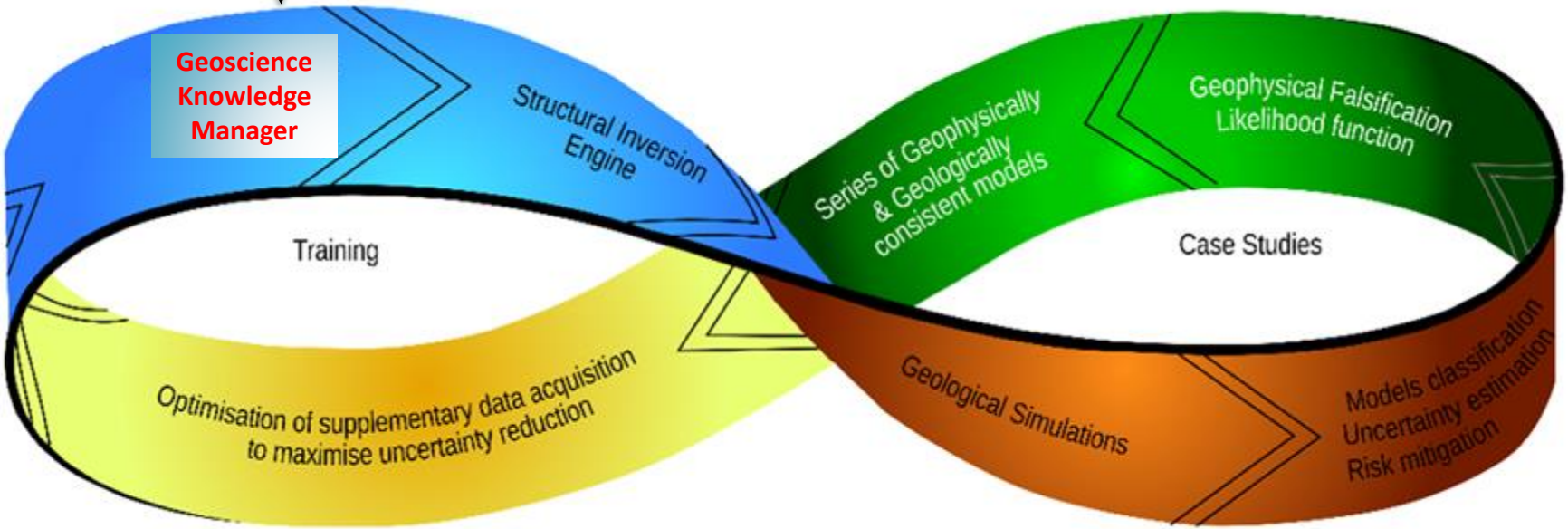
Geoscience Knowledge Manager

WP1

Boyan Brodaric

Steve Richard

WP1



WP1: OneGeology team (GSC, GA, BGS, BRGM)
Steve Richard

Knowledge Challenge

Knowledge-rich

Maps, x-sections, articles, diagrams, rules, …
- partially integrated into modelling

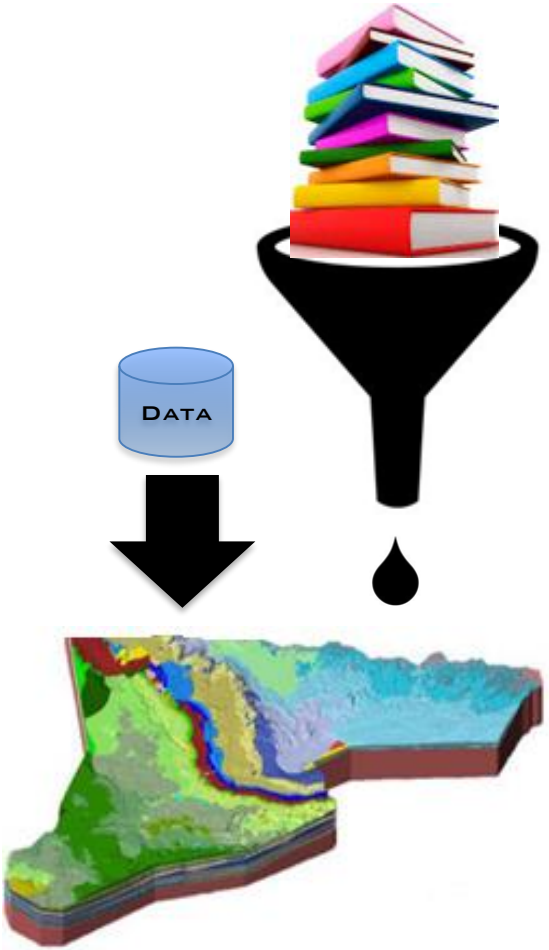
Data-poor

Few subsurface observations relative to area
- highly integrated into modelling

Challenge

Better leverage knowledge

- use more knowledge
- use knowledge more

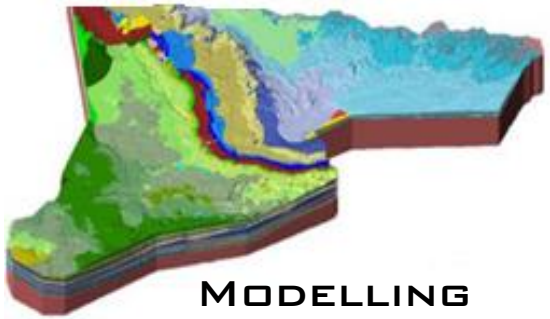


Geoscience Knowledge Manager (GKM)

Concept

Geoscience Knowledge Manager
interpretations

- stratigraphy
- topology
- …



Modelling
models

Data
observations

- boreholes, field data, geophys,…

Knowledge

Atemporal (Theoretical)
- laws
- theories
- rules
- universal types
- classification schemes
- atemporal relations

Temporal (Situational)
- processes and events
- temporal relations
- histories
- particular types

Spatial (Situational)
- map legends, conceptual models
- qualitative spatial relations
- interpretations: units, structures,…
- feature descriptions without quantitative geometry & location

WP1 – Knowledge Manager

Data

Observational
- field data
- boreholes
- geophysics
- remote sensing
- laboratory analyses

1D, 2D Models (locational)
- 1D: borehole interpretations
- 2D: maps, cross-sections (geometries)

WP2 – Data Manager

Models

3D Models
- input data
- input knowledge
- configurations
- functions, scalar fields
- output geometries
- output uncertainties
- method provenance

WP3-5 – Modelling

Knowledge constraints on topology
Includes stratigraphic relations / correlations

Geoscience Knowledge Manager

Components

extract

GeoSciML?

-unstructured info:
text documents
databases
-ontology-driven
-relation discovery

store

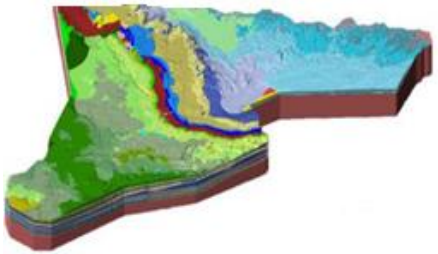
Geoscience Knowledge Manager

Structure Repository

use

GeoSciML?

-Select algorithm
-Assign weights
-Infer event history
-Infer data points
-Consistency check



- know structure: ontology
- know repository: RDF store
  load, store, find, reason, get

BGS GSC, GA GSC

GKM Mid-term Results

WORK IN PROGRESS

1. Knowledge Extraction

Mon 9:40 AI (ML)-driven strat info from text documents (R.E. Heaven)

2. Knowledge Management

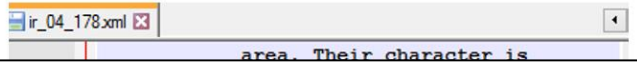
Mon 9:20	requirements	(E. de Kemp – GSC)
Mon 9:40	draft ontology I	(B. Brodaric – GSC, R.E. Heaven – BGS)
Mon 10:00	draft ontology II	(S. Richard – GSC/USGIN)
Mon 10:50	draft repository	(D. Lecinsky/W. Francis – GA)

3. Model Validation

Tues 11:00 consistency checker (M. Parquer - GSC)

Prepare training data

1. XML semantic markup experiments (early 2000s)



2. Converted to training data

NER model training

3. Train a custom Named Entity Recognition (NER) model for stratigraphic terms
https://github.com/BritishGeologicalSurvey/geo-ner-model

```
java -cp stanford-ner.jar
```

NER tagging

4. Run text mining using custom NER model for stratigraphy and core NER for locations

Stanford CoreNLP

Entity resolving

5. Stratigraphy entities matched to BGS Linked Data vocabs

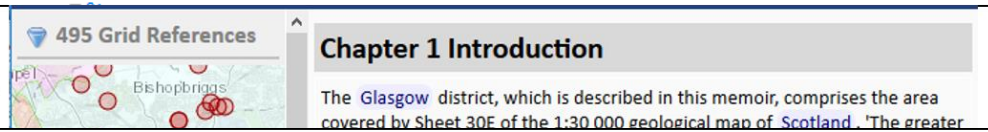
6. Locations resolved to geometry using gazetteer

Searching

7. Database of mined data with API

Visualising

8. Marked up text linked to map
http://194.66.252.183:81/TextMiningDemo



Analysing

9. API supports frequency and correlation analysis

Source	Code	Title	Ages	Rocks	Places	Points	Total
	B01879	Geology of the Glasgow district. Sheet 30E Sheet memoir (Scotland)	335	591	920	476	2322
	1		335	591	920	476	2322

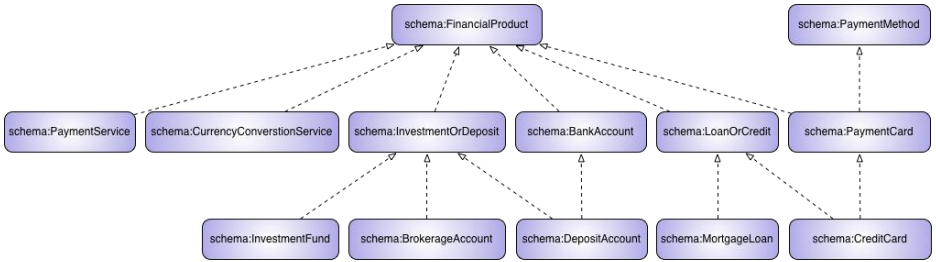
British Geological Survey
Expert | Impartial | Innovative

2020/03/03 Rachel Heaven reh@bgs.ac.uk

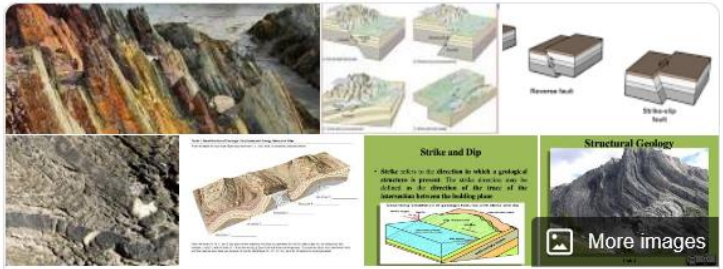
Knowledge Management

is Hot!... Google, Amazon, MicroSoft...

Google Knowledge Graph (2016)
570 million objects
70 billion facts
schema.org ontology
query API (e.g. SPARQL)
powers Google's **infobox**



schema.org financial product



Structural geology **infobox**

Field of study

Structural geology is the study of the three-dimensional distribution of rock units with respect to their deformational histories. Wikipedia

People also search for View 15+ more

Stratigra... Petrology Economic geology Mineralo... Paleonto...

GeoScience Ontology (GSO)

WORK IN PROGRESS

PREDECESSORS

NADM	(NADM 2004)	conceptual only
GeoSciML	(GeoSciML 2017)	data language
DOLCE Rocks	(Brodaric & Probst 2009)	DOLCE, broad not deep

RECENT WORK (selected for breadth)

GeoSciML ontology	(OGC SWG, in progress)	no upper ontology
OntoGeonous ontology	(Lombardo et al. 2018)	no upper ontology
GeoCore ontology	(Garcia, Abel, Perrin 2020)	BFO, broad not deep

GSO

GeoScience Ontology	(in progress)	DOLCE + BFO + GeoSciML

GSO Development

Requirements

Broad and Deep:	basic geologic entities (min GeoSciML)
Modular:	core ontology + plug-and-play modules
Stand-alone:	no imports – for compactness and consistency

Methods

✓	GKM GitHub	https://github.com/Loop3D/GKM
✓	Ontological analysis	(First-Order Logic, Dolce, BFO)
✓	UML diagrams	(Enterprise Architect)
✓	OWL encoding	(ttl: TopBraid, Protégé)
✓	Examples	(ttl: Units, Faults, Folds, Time Scale)
	Testing	(more synthetic examples)
	Implementation	(triplestore, real data)
	3D Integration	(with modelling code)

GSO Modular Organization

Layercake Ontology

Common (static)

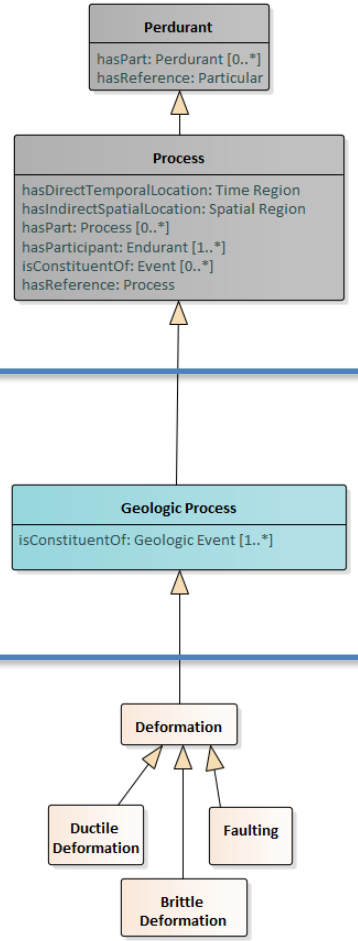
general framework: e.g. objects, processes, qualities,…

Geology (static)

geology framework: e.g. geologic units, structures, events, time,…

Modules (extensible, exchangeable)

geology modules: e.g. properties, settings, structures, lithologies, minerals, elements,…



GSO GitHub (https://github.com/Loop3D/GKM)

Layercake Ontology

Common
general framework

BFO: space and time
Dolce: physical objects, situations
BFO + Dolce: qualities
GSO: features
Other: processes, events

GSO-Common.ttl

Geology
geology framework

GSO

GSO-Geology.ttl

Modules
geology modules

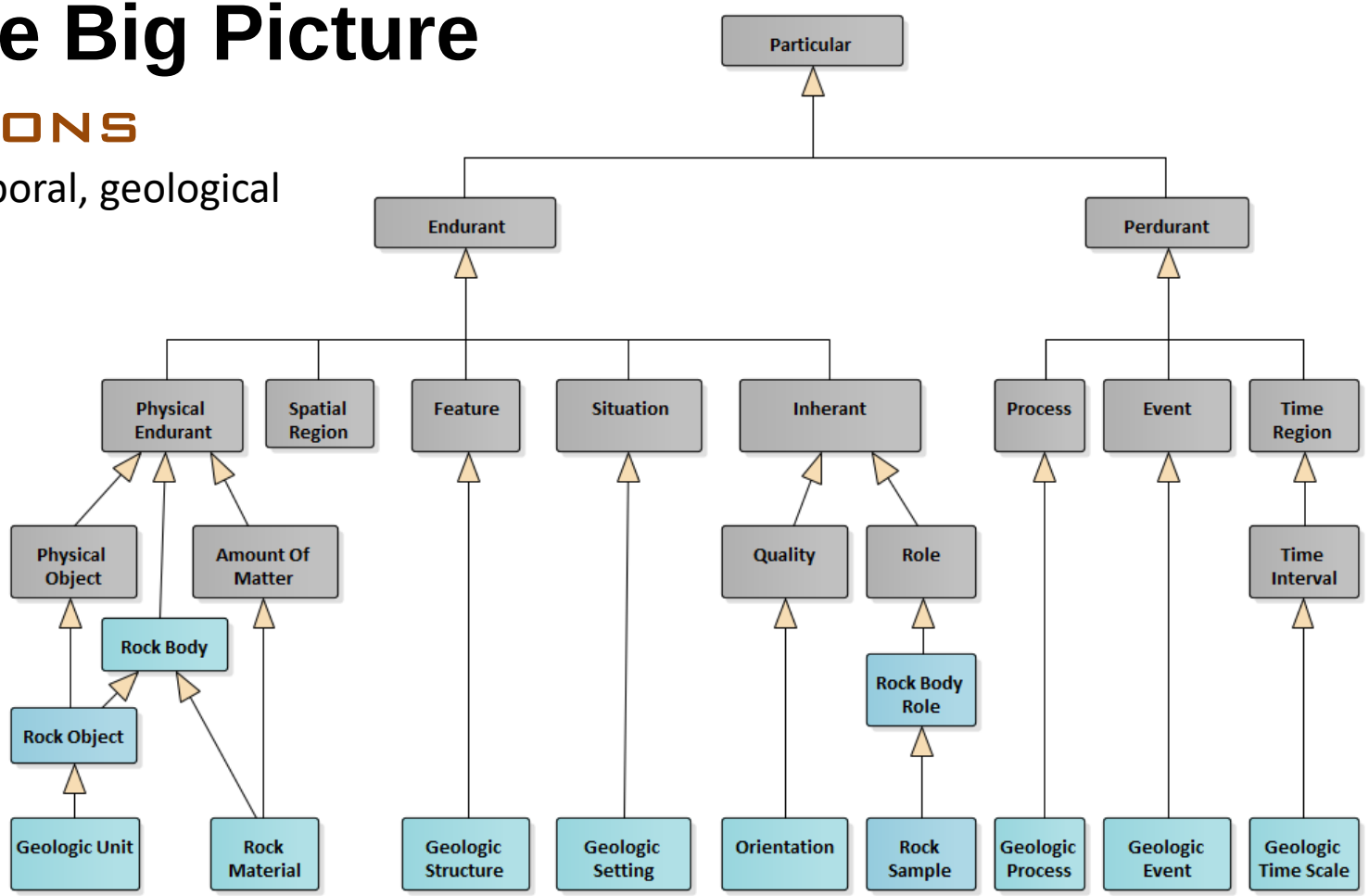
GeoSciML

- GSO-Elements.ttl
- GSO-Geologic_Process.ttl
- GSO-Geologic_Time_Interval.ttl
- GSO-Geologic_Unit.ttl
- GSO-Hydrology.ttl
- GSO-Minerals.ttl
- GSO-Physical_Property.ttl
- GSO-Physical_Setting.ttl
- GSO-Rock_Material.ttl
- GSO-Structure-Contact.ttl
- GSO-Structure-Fault.ttl
- GSO-Structure-Fold.ttl
- GSO-Structure-Foliation.ttl
- GSO-Structure-Lineation.ttl

GSO The Big Picture

+ RELATIONS

spatial, temporal, geological



GSO Qualities

after BFO + DOLCE

qualities are characteristics: e.g. *orientation*, *colour*, *density*,…
qualities are **not** relations: **not** *over*, *below*, *older*,…
qualities inhere-in things: in *rock bodies*, *materials*, *processes*, *events*,…
qualities are singly-dependent: they depend on the one thing they *inhere-in*

inheres-in is a special kind of dependence relation

qualities can have values	*strike* = 45
values are datatypes or qualities	*strike* = 45 *strike* = northeast
qualities can have qualities	*planar orientation* = {*strike* = 45, *dip* = 75}
unit of measure is a quality	*planar orientation* = {*strike* = 45°, *dip* = 75°}

qualities are a module, so easily extended

GSO Space and Time

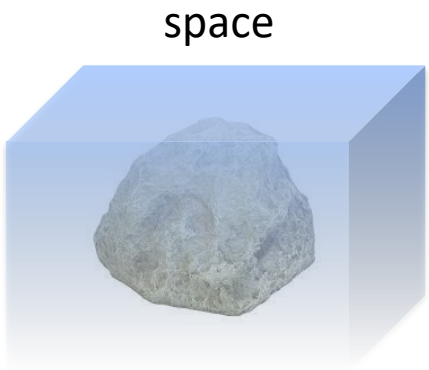
after BFO

space and time are not qualities
space and time are containers occupied by things
space is occupied by object-like things (Endurants)
time is occupied by processes and events (Perdurants)

physical endurants are directly located in space
- spatial location (WKT) is a quality of a chunk of space

physical endurants are indirectly located in time
- by virtue of the processes and events they participate in

processes and events are directly located in time
processes and events are indirectly located in space
- by virtue of the objects participating in them



GSO Time Scale

type of time region →

instance of time region →

= collection of geologic time regions

e.g. ICS 2014

geologic time region

time instant: time of a geochron boundary, GSSA

time interval: has a geologic referent

geologic referents for time regions

time instant: stratigraphic point (contact) referent, GSSP

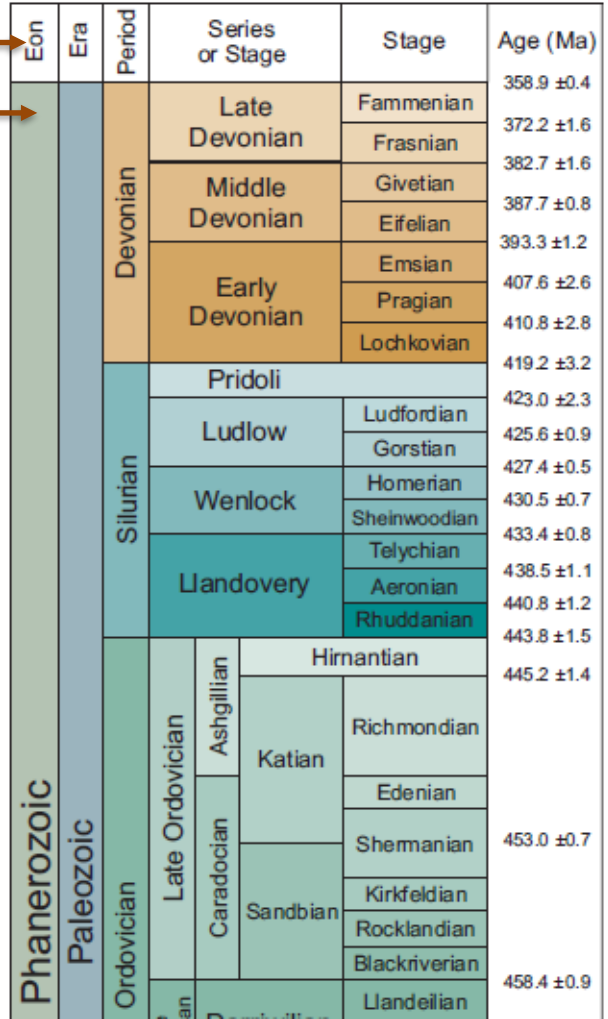
time interval: chronostratigraphic unit referent

types of geologic time intervals

Eon, Era, Period,…

instances of geologic time intervals

Precambrian, Archean, Devonian



GSO Rock Materials

after DOLCE and Hahmann & Brodaric 2014

amounts of matter (AoM) are not abstract
- contra GeoSciML: can be touched, moved, have mass,…

AoM can be unified or not
- contra Dolce, Abel *et al.* 2015
- sandstone layer (unified, bounded)
- sandstone in formation X (not unified, unbounded)

AoM are constituents, not parts of physical objects
- `Rock Material` constitutes `Rock Object`
- `Rock Object` has constituent `Rock Material`
- `Rock Material` has constituent other AoM

AoM can have parts and constituents that are AoM
- top part of sandstone layer
- sandstone layer constituted by sand particles

Kinds of AoM (GSO)

```
Amount of Matter
    Rock Material
    Granular Material
    Glass
    Mineral
    Element
```

GSO Rock Objects

are physical endurants made of rock material
- directly located in space (occupy a chunk of space)
- indirectly located in time (occupy time via processes and events)
- are wholes unified by internal characteristics **or** external topology
  - internal characteristics = composition, geometry, structures

are not spatial regions or situations
- they directly occupy chunks of space (so are not chunks of space)
- situations indirectly occupy space (by virtue of their parts)

are more general then geologic units
- rock objects are defined by internal traits **or** topology
  - e.g. facies (compositional) or stratigraphic section (topological)
- geologic units are defined by internal traits **and** topology
  - units have uniform composition and specific relations to other rock bodies

Kinds of Rock Objects (GSO)

```
Rock Object
    Crystal
    Fossil
    Facies
    Lithosome
    Stratigraphic Section
        Stratotype
```

GSO Geologic Units

units are physical objects
- participate in processes and events
- have direct spatial location
- defined by internal characteristics + external topology

units are wholes with fragment parts
- contra types (as per theory)… for practicality

distinguish unit types vs instances
- `Lithostratigraphic Unit` (LU) is a type
- `Dakota Formation` is an instance of LU

unit ranks are types of units
- `Formation` is a subtype of `Lithostratigraphic Unit`
- `Dakota Formation` is an instance of `Formation`

- `Eonothem` is a subtype of `Chronostratigraphic Unit`
- `Archean` chronstratunit is an instance of `Eonothem`

Kinds of Units (GSO)

```
Lithostratigraphic Unit
    Supergroup
    Group
    Formation
    Member
    Bed
    …
Chronostratigraphic Unit
    Eonothem
    Erathem
    System
    Series
    …
```

GSO Structures

after Brodaric 2019

geological structures are **`Features`**

- emergent: derived from at least 2 things, hosts and focals
- relational: derived from a **relation** between the things
- multi-dependent: cannot exist without the things
- spatial: directly located in space
- singular: a single entity

`Inherent Feature`

derived from a property such as shape: `Fold`, `Esker`

`Integral Feature`

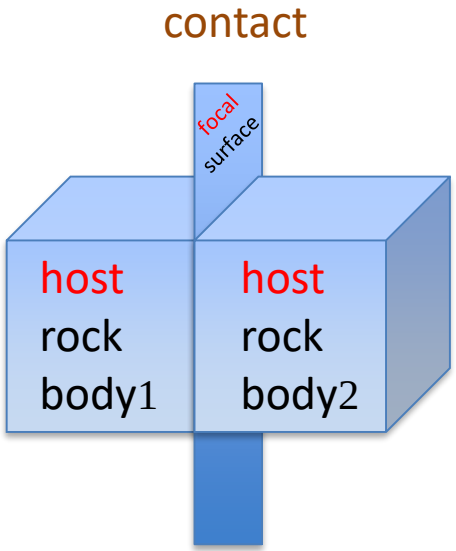
derived from an organization of parts: `Foliation`, `Lineation`

`Material Spatial Feature`

derived from a spatial relation: the rock between faults = `Fault Zone`

`Low Dimension Feature`

point or surface between host rock bodies: `Contact`, `Fault`, `Fracture`



GSO Relations

Relations connect 2 or more things

- dependent, emergent, plural
- `meets` (A, B)
- `younger` (A, B)
- `intrudes` (A, B)

Relations are aspatial and temporal

- Relations do not occupy space directly
- Relations occupy time directly

Relations are not Features!

- Features occupy space directly
- Features derive from relations
- Features are singular
- `meets` (A, B) – touching of rock bodies A and B
- `hosts` (A, B, C) – C is a contact hosted by A and B

Geological relations

- hypothesis: spatial + temporal + process aspects
- intrudes (A,B) = meets (A,B) ^ younger (A,B) ^ partic (A,P) ^ intrusion (P)

Kinds of Relations

`spatial`
- **(Egenhofer/RCC)**
- `contains`
- `meets`
- …

`temporal`
- **(Allen's interval)**
- `during`
- `before (older)`
- `after (younger)`
- …

`geological`*
- `intrudes`
- `cross-cuts`
- …
- *requires work

GSO Roles

after BFO + UFO

a role is played by something:

- *sample* is played by some *rock object*
- *pendant* is played by some overlying *rock object*
- *clast* is played by some *amount of matter*

playing is a special kind of inheres-in relation

a role inheres-in so depends-on the playing thing

- like qualities: e.g. a sample inheres-in and depends-on a rock object

a role has a related host:

- *sample* is sample of some *rock body*
- *pendant* is hosted by some underlying *rock body*
- *clast* is hosted by surrounding *amount of matter*

Kinds of Roles

```
Rock Body Role
    Rock Object Role
        Pendant
    Rock Sample
    …
Amount of Matter Role
    Clast
    Inclusion
    Xenolith
    Xenocryst
    …
```

GSO Processes and Events

after Galton & Mizoguchi 2009

processes and events unfold in time

- only a part is present at any one time
- have endurants as participants
- have parts to form event/process histories

processes are ways of happening

events are the happenings

Process	Event
running in Boston Marathon	Boston Marathon
shaking in San Fran earthquake 1989	San Fran earthquake 1989
rifting in Late Triassic	Pangea break-up
deposition process of Formation X	deposition of Formation X

processes constitute events

- analogous to matter constituting objects
- e.g. Boston Marathon constituted by the running in it

Kinds of Process

after Perrin *et al.* 2005

```
Geologic Process
    Additive
        Deposition…
    Subtractive
        Erosion…
    Transformational
        Metamorphism…
    Deformational
        Faulting…
```

GSO Settings

after DOLCE

situations are fragments of the real world
- situations are endurants: wholly present at a time

situations have endurants as parts
- situations have at least one endurant part

situations are indirectly located in space & time
- located in space by virtue of their endurant parts
- located in time by virtue of the processes/events they engage

settings are geologic fragments
- under-developed in GSO

Kinds of Setting (GSO)

```
Situation
    Geologic Setting
        Marine
        Alluvial Fan Setting
        …
```

GSO What's Missing?

rules	
relations	(many)
modules	(many)
testing ontology	(real examples)
testing repository	(load, query, reason)
3D code integration	

GSO What's Next?

work group engagement	(open GitHub)
testing	(ontology + repository)
refinement	(all the above)

THANKS!

WP1 Breakouts

Outcomes

1.	Complete GSO draft v1	May 2020
2.	Open GitHub to LOOP participants	May 2020
3.	Start regular monthly meetings (open)	Apr 2020
4.	Start wider testing GSO v1	Apr 2020 (GSC), June 2020 (GA), tbd (BGS)
5.	Start testing GKM repository	June 2020
	- populate with GSO examples	GA, GSC, WP2
	- develop retrieval protocols	GA
6.	Start integration with WP2, WP3	ongoing
	- **possible workshop in Ottawa**	**Aug, Sept, or Oct 2020**
	- WP2: add M2L vocabularies	
	- WP2: explore consistency check integration	
	- WP3: refine ontology for WP3 needs	Apr 2020 – Aug 2020
	- WP3: integrate GKM with WP3 code	Ottawa workshop