



HaIRST

Harvesting Institutional Resources in Scotland Testbed

hairst.cdlr.strath.ac.uk
2002-2005

cluster coordinates

scope and assumptions in the eFair context



NAPIER UNIVERSITY
EDINBURGH

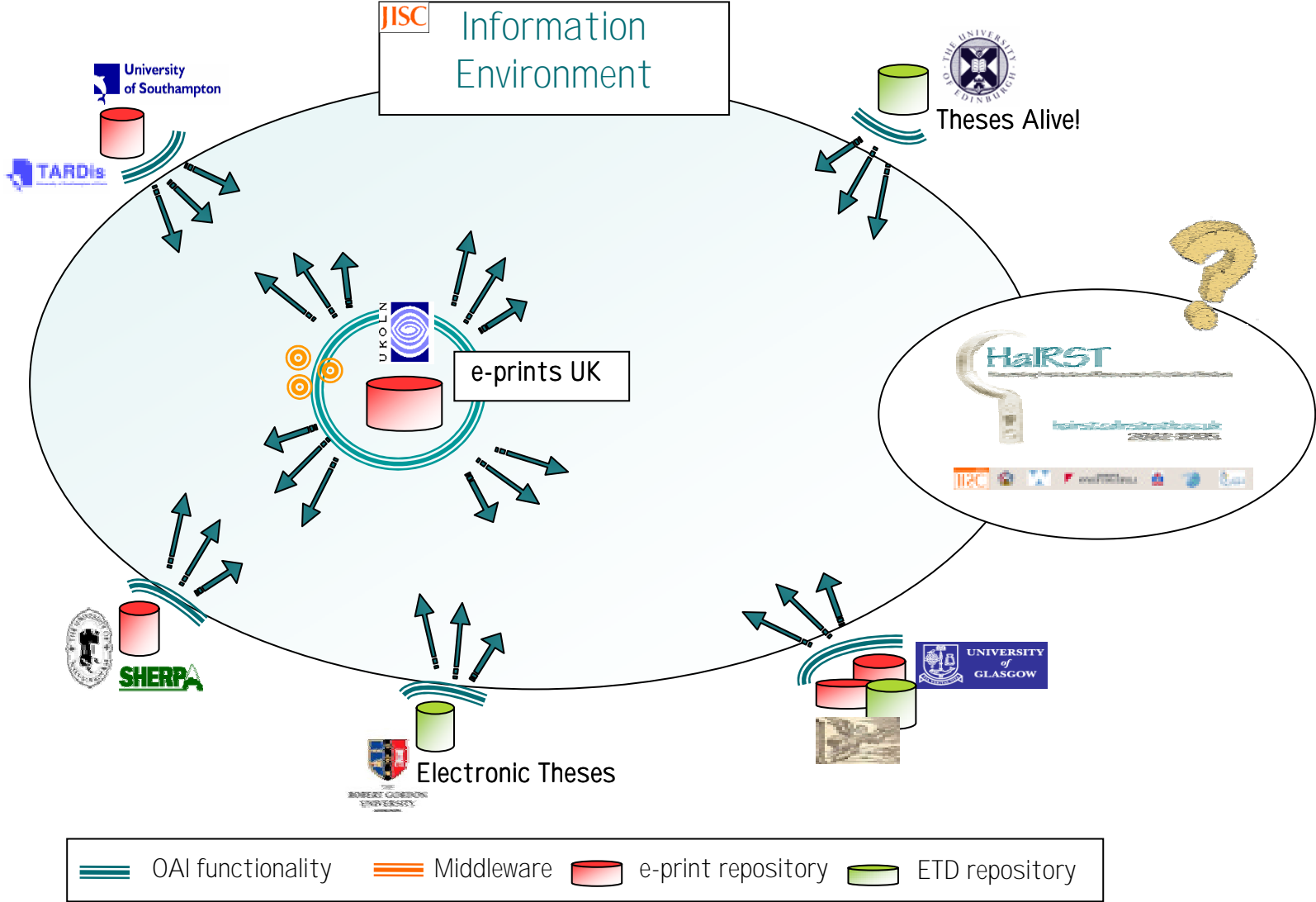


CDLR@Strathclyde Univ. · Napier Univ. · St. Andrews Univ. · Glasgow College Group · John Wheatley College

cluster traits

- FAIR mission: *develop* the service infrastructure of the IE;
 - deposit*: build local DNER collections at *institutional* IE nodes;
 - disclosure*: open local DNER collections to remote IE services;
 - discovery*: locate remote DNER components from local IE services;
 - middleware*: support interoperability & promote participation;
- eFair primary strand: deposit & disclosure services at IE 'end-points':
 - focus on *research* output: *e-prints* and *ETDs*;
 - OAI assumption: *pre-harvesting* services;
 - some technical enquiries: 'what software, what formats, what interface?'
 - many organisational enquiries: 'what policies as to content, preservation, IPR, submission, classification, ...?';
 - plenty of advocacy!
- eFair secondary strand: *post-harvesting* services:
 - discovery services;
 - middleware services (e.g. metadata enhancements);
 - DC assumption;

cluster map

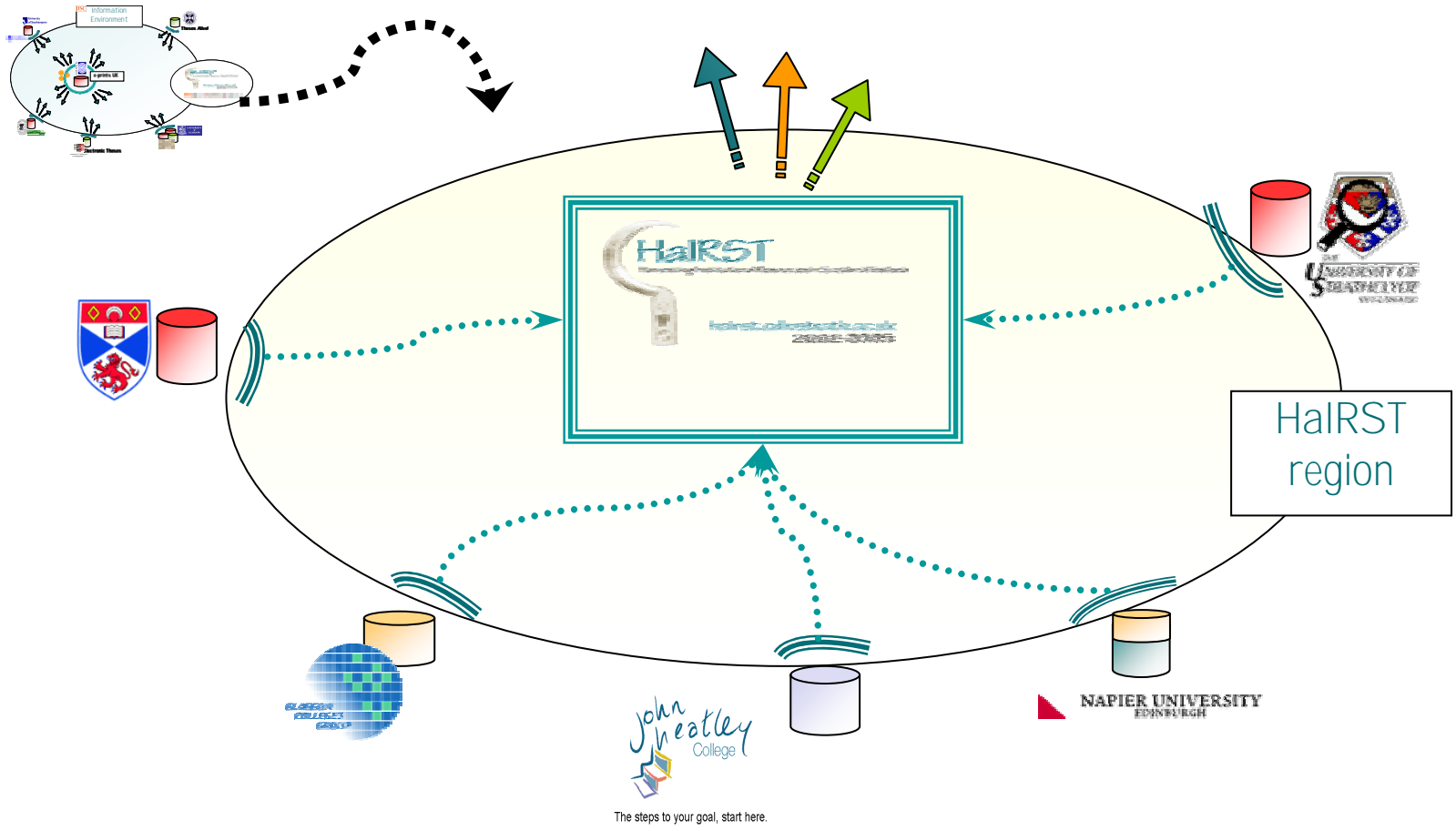


project traits

- 🖥️ A 'complete' exercise in OAI-based interoperability within the IE:
 - 🖥️ *multiple, cross-sector partners-as-data-providers:*
 - 🖥️ 3 Scottish Universities & 10 Glasgow Colleges;
 - 🖥️ *technical & cultural diversity:* data, metadata, hardware, software, people, requirements, policies, cultures, agendas, ...;
 - 🖥️ often *competitive* environment;
 - 🖥️ *pre-harvesting services:* local deposit and disclosure services
 - 🖥️ *post-harvesting services:* discovery, metadata mapping, and further disclosure services;

- 🖥️ A rich investigation:
 - 🖥️ *technically:* identify & explore different software pathways to OAI compliance:
 - 🖥️ from '*thick*', comprehensive solutions (e.g. *eprints.org*)...
 - 🖥️ ...to '*thin*', ad-hoc layers on top of pre-existing back-ends (e.g. *OAICat*), including bare file-systems;
 - 🖥️ *culturally:* understand different milieus to tailor and promote the JISC message of interoperability;

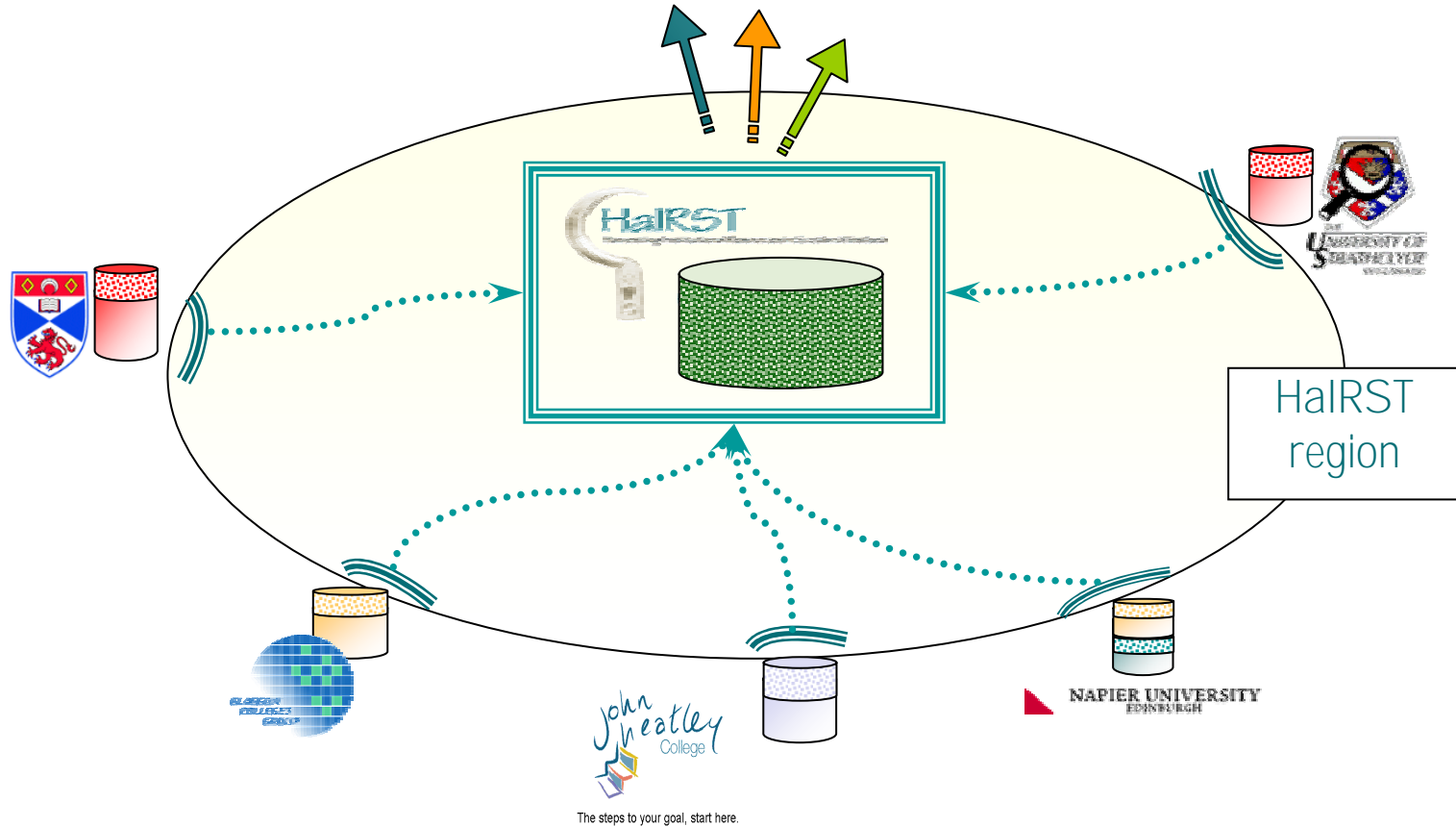
HaIRST region



metadata

- ❏ the challenge of **exporting local diversity** in the IE:
 - ❏ loosen the DC assumption: no single *metadata standard* or *application profile*;
 - ❏ would simplify post-harvesting service deployment, but...
 - ❏ ...if simple, impoverish overall IE functionality: *one-stop entry-points* with *GCD* functionality only...;
 - ❏ ...if complex, limit the IE scope only to '*rich*' nodes;
 - ❏ harvest **multiple standards or profiles**:
 - ❏ local, inter-provider, project-wide, wider...;
 - ❏ typically, *modularly* composed from multiple international standards (DC, MARC, DCMI-Edu, LOM, IMS, ...);
 - ❏ through local *extensions* and/or *refinements*;
 - ❏ from *simple* to *complex*, enough to satisfy local requirements;
- ❏ deploy a post-harvesting discovery service that exploits metadata diversity to the benefits of users...
- ❏ ...*how?*

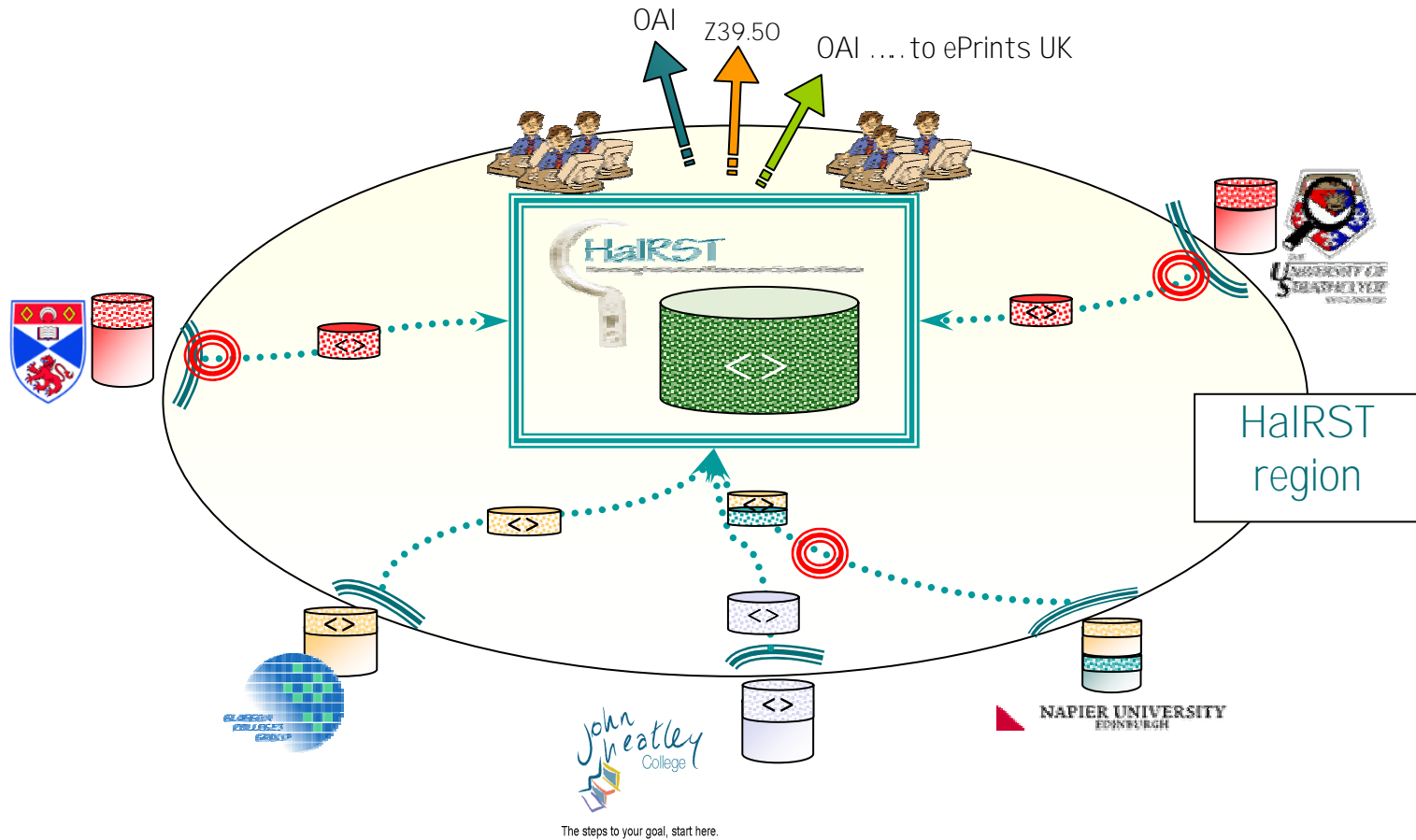
HaIRST region (rev.)



approach

- 🖥️ application profiles are:
 - 🖥️ arbitrarily-grained *extensions/refinements* of single, top-level, DC-like profile;
 - 🖥️ syntactically incarnated in *XML* (syntactic interoperability), possibly *RDF-based* (structural interoperability);
 - 🖥️ *native* or *derived* from pre-existing forms (e.g. MARC, IMS);
 - 🖥️ if derived, locally by providers or remotely by harvesters/brokers/aggregators, etc. through *mapping services*;
 - 🖥️ harvested into semi-structured repository (file-based or native XML DB);
 - 🖥️ **queried** with an *XQuery-like* language via a *graphical Web interface* that:
 - 🖥️ allows incremental structuring of queries;
 - 🖥️ from *keyword-* to *field-based* to a mixture of the two
 - 🖥️ *simpler-query=larger-input, harder-query = smaller-input...and yet as large as* it may be:
 - 🖥️ requires some *dumb-down* mechanism;
 - 🖥️ **further disclosed** into the IE via *OAI* (hierarchical harvesting) or *Z39.50* (distributed searching);

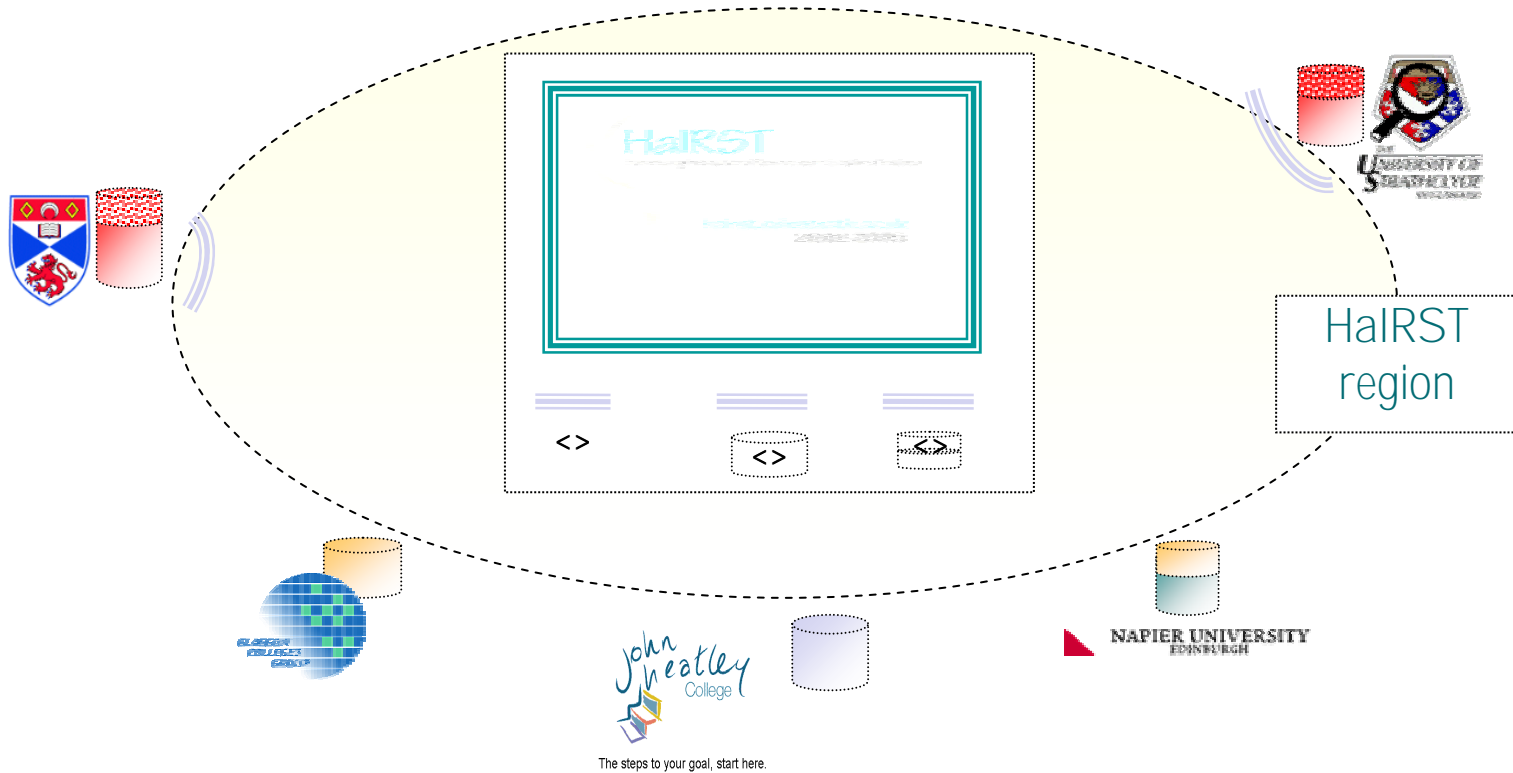
HaIRST region (rev.)



progress

- 🖥️ slow!....coordination is difficult, small steps on all different fronts:
- 🖥️ **Resource Assessment Exercise** across partners (Dec02-Feb03);
 - 🖥️ of staff, hardware, software, and management resources;
 - 🖥️ for logistic arrangements of: staffing & purchasing policies, resource identification, per-partner OAI software pathways design;
 - 🖥️ *resource identification is difficult!* Different agenda and motivations inter- and mostly cross-sector;
 - 🖥️ a *delegation* model of participation is often preferred;
- 🖥️ **next steps:**
 - 🖥️ *Metadata Orientation Event* (Apr03): Assessing standards and practices for local metadata profiles;
 - 🖥️ start deploying and developing partner-specific deposit and disclosure services;
- 🖥️ **summer goal:** complete harvesting of DC-only sample metadata into XML-based repository;

HaIRST region (rev.)



≡ OAI functionality  repository  metadata