SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Linked Open Heritage Data in Sweden:
SOCH & SweDigArch

Marcus Smith, Swedish National Heritage Board

Linked Open Cultural Heritage Data

- FAIR, LOD, RDF…
- Swedish Open Cultural Heritage
- Urdar and SweDigArch

Inmatning av rundata
CC BY Runverket, Riksantikvarieämbetet

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

FAIR Data
Linked Open Data
RDF

FAIR data principles

Recommendations for how data should be made *machine*-readable so that *humans* can use it.

Findable

I want to be able to **find** information/images/data/publications, etc
Requires **structure** and **consistency**

Accessible

I want to be able to **download** and **process** the information
Requires **persistent resolvable identifiers** (e.g. URIs, DOIs)

Interoperable

I want to be able to **combine** data from **multiple sources**
Requires standards for **terminology** and shared **referents**
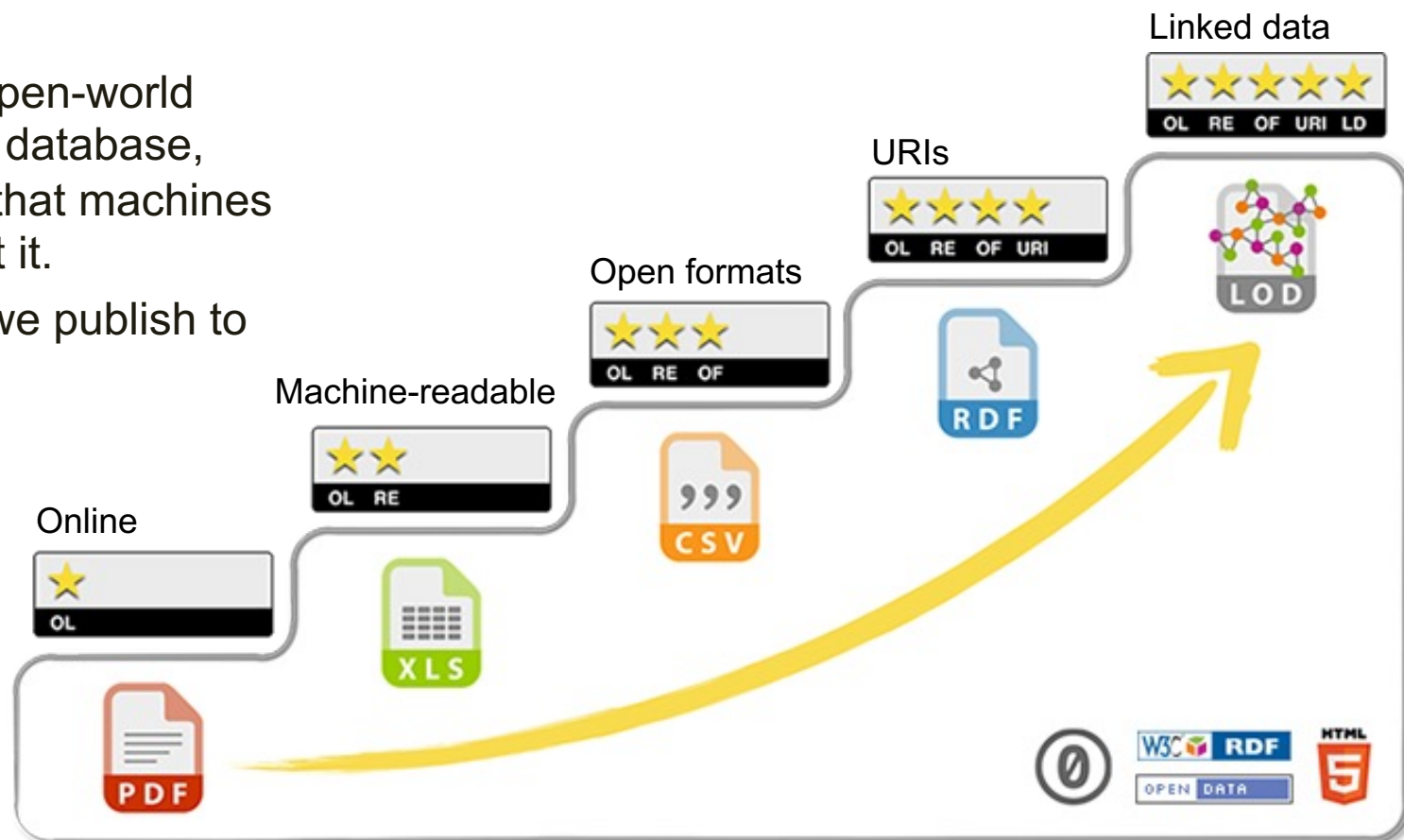
Reusable

I want to know how I may **use** and **publish** what I find
Requires clear **rights statements** for both humans and machines

https://www.go-fair.org/fair-principles/

SWEDISH NATIONAL HERITAGE BOARD
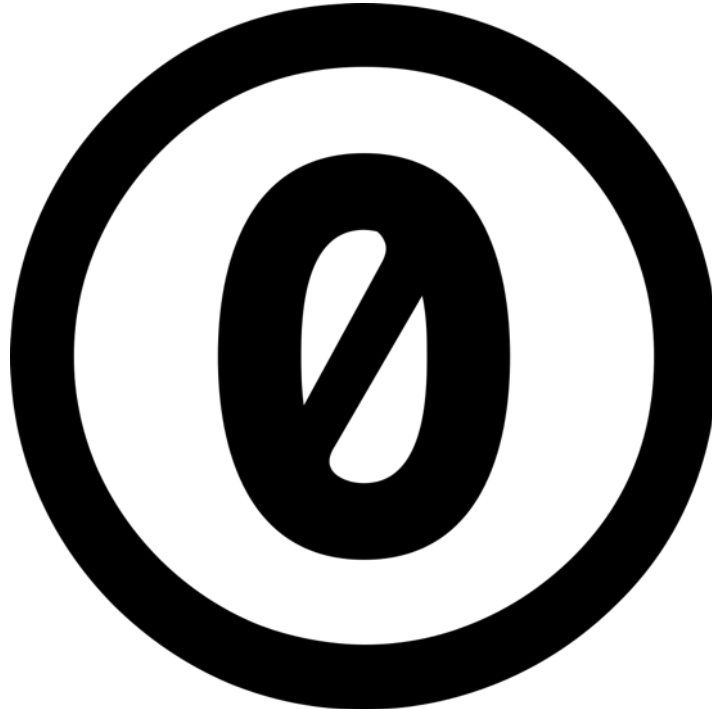RIKSANTIKVARIEÄMBETET

Linked (Open) Data and the Semantic Web

- The semantic web is an open-world distributed directed graph database, expressed in such a way that machines can use and reason about it.

- Linked open data is how we publish to enable this.

- Beyond FAIR publication, it's allows data to be described using shared vocabularies, and relating datasets to one another across the web, lifting data out of silos.

https://5stardata.info/

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Open Data

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET
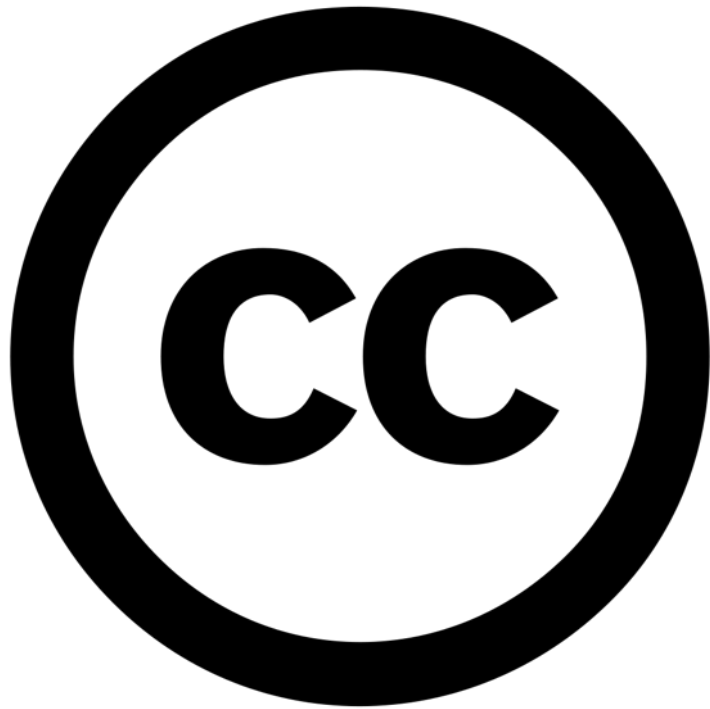
Linked Data
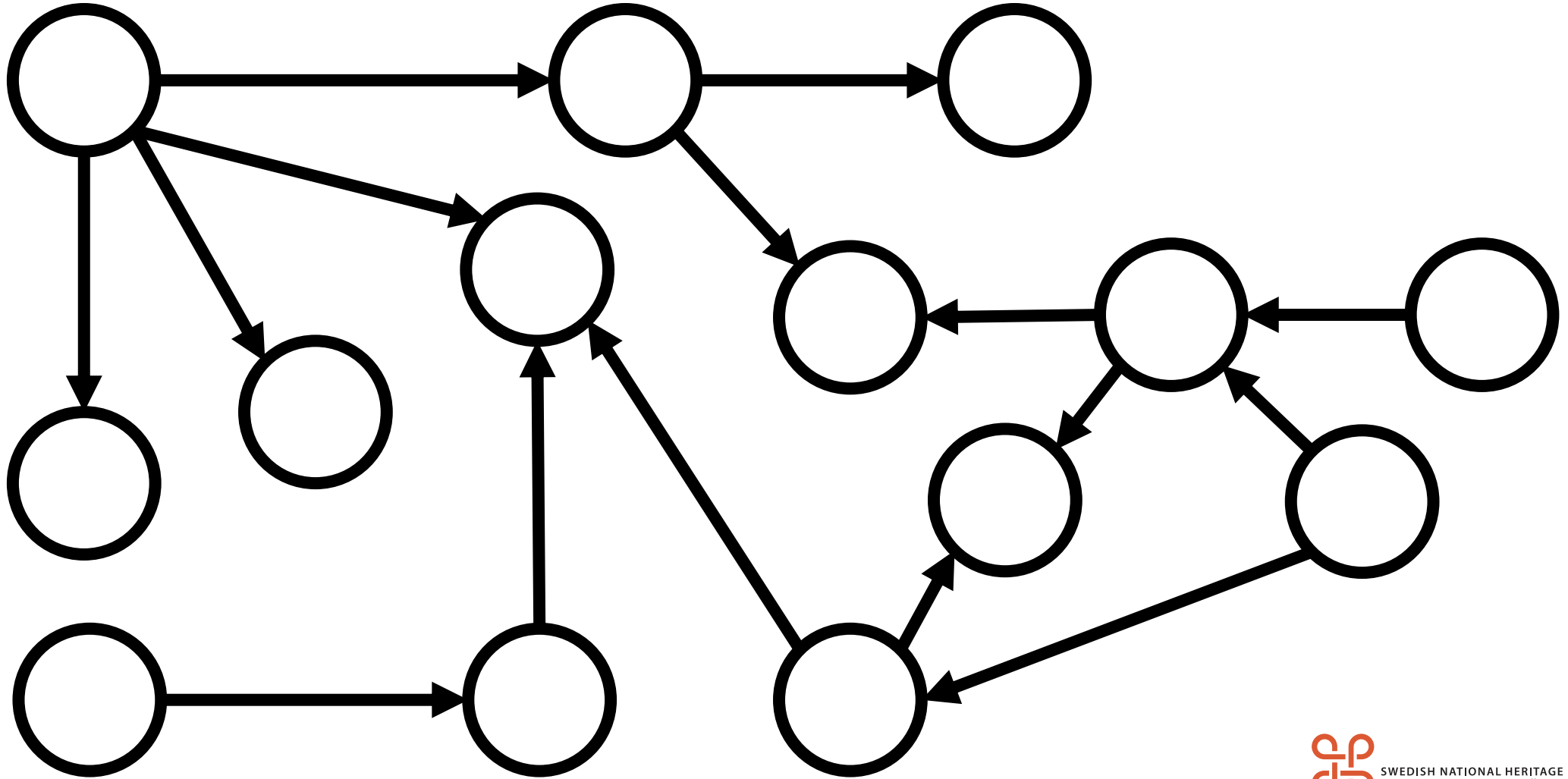
RDF

- URIs as identifiers
- Statements (triples) composed of a subject, a predicate, and an object
- Combined together to make graphs
- Serialisation formats:
  - RDF/XML
  - Turtle
  - JSON-LD
  - (and some others e.g. canonical N-Triples)
- Querying using SPARQL:

```
SELECT ?subject ?predicate ?object
WHERE {
  ?subject ?predicate ?object .
}
```

RDF – the Resource Description Framework

Subject   **Predicate**   **Object**
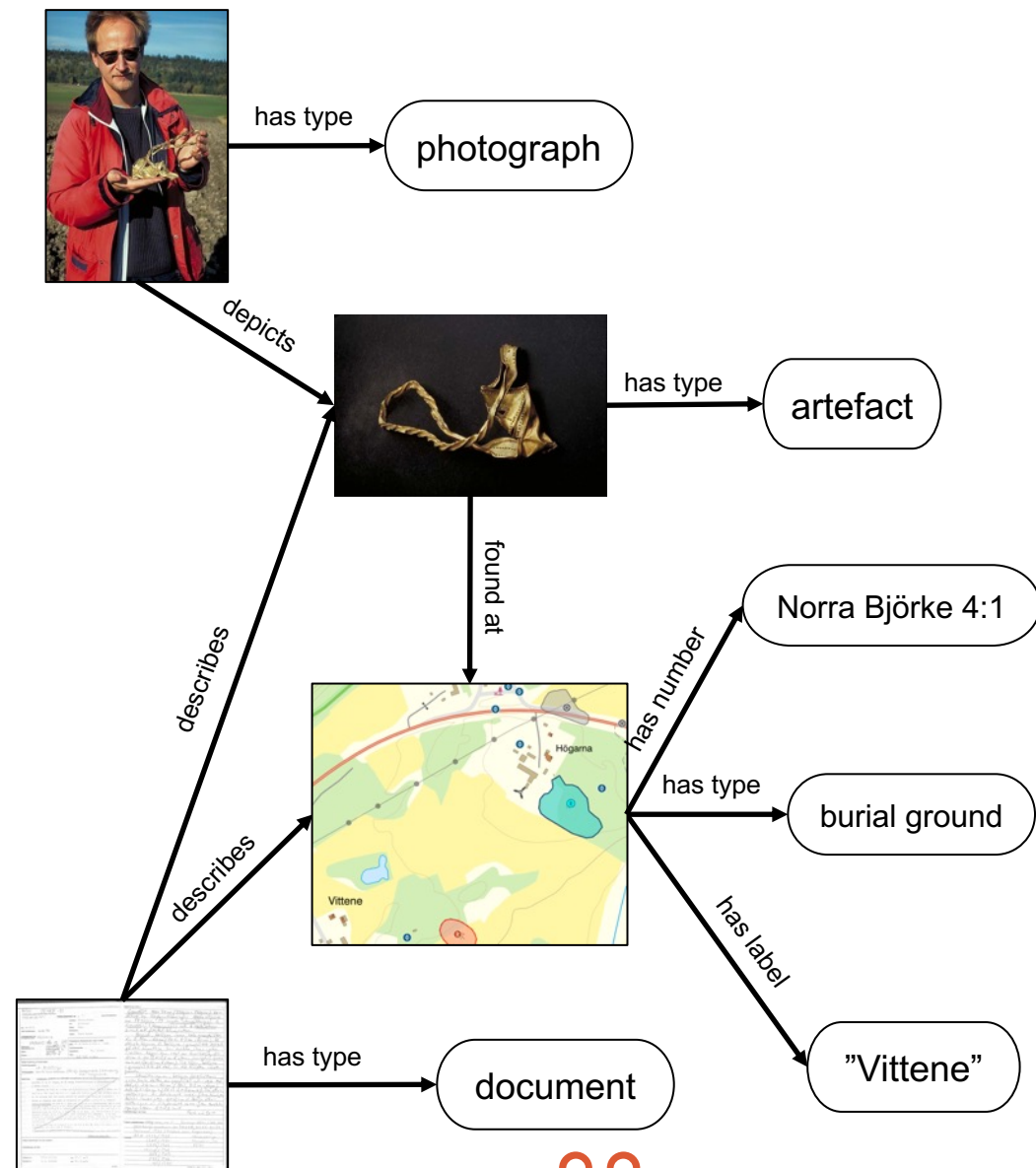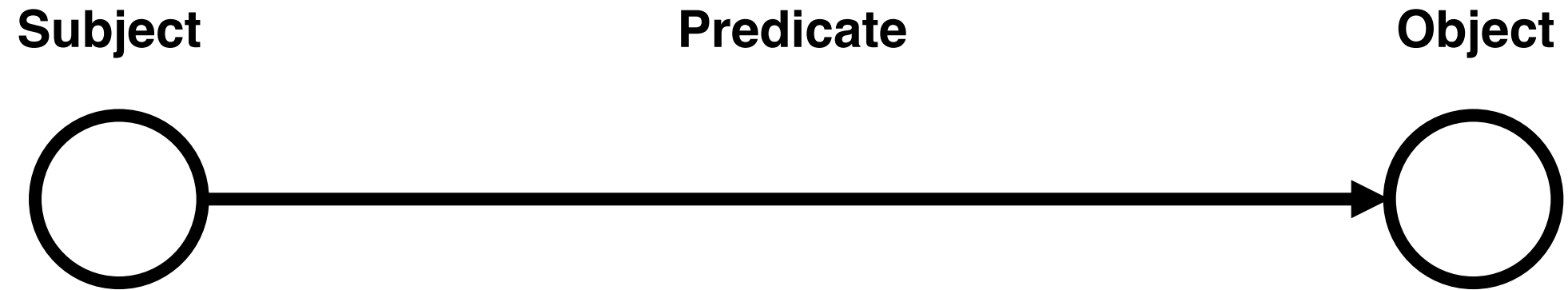
RDF – the Resource Description Framework

<http://www.example.com/subject>   <http://www.example.com/predicate>   <http://www.example.com/object>
.

RDF – the Resource Description Framework

@prefix ex: <http://www.example.com/> .

ex:subject   ex:predicate   "Object" .

Some LOD applications you might be familiar with

- **PAS** <https://finds.org.uk/> – Portable Antiquities Scheme

- **Nomisma** <http://nomisma.org/> – Coins data operated by the American Numismatic Society

- **Finto** <https://finto.fi/> – Finnish national vocabulary and ontology service

- **Hipla** <http://dev.hipla.fi/> – Finnish Ontology Service of Historical Places and Maps

- **Finna** <https://finna.fi/> – Finnish cultural heritage aggregator

- **SuALT/FindSampo** <https://blogs.helsinki.fi/sualt-project/> – Finnish Archaeological Finds Recording Linked Open Database

- **WarSampo** <https://seco.cs.aalto.fi/projects/sotasampo/> – Finnish World War II on the Semantic Web
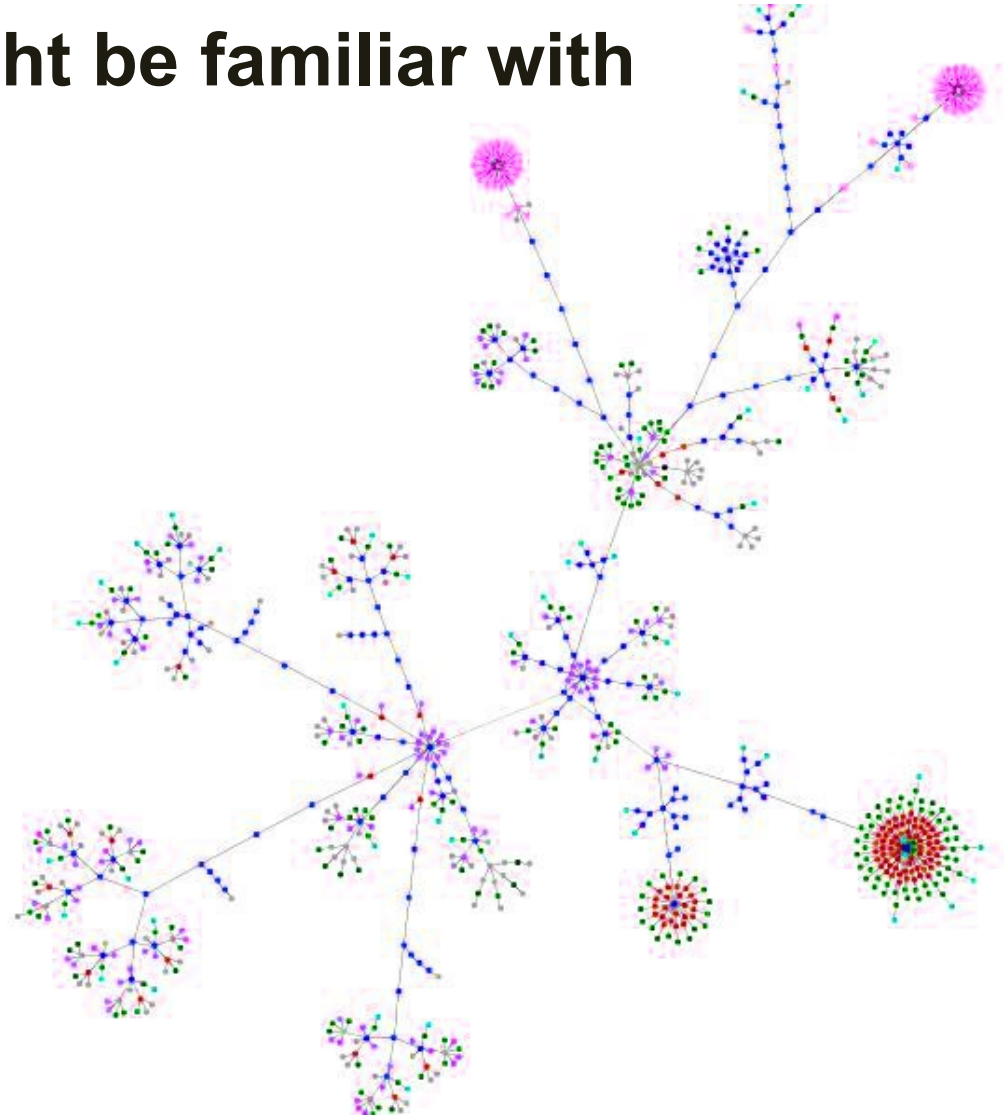
Swedish Open Cultural Heritage (SOCH)

SOCH (Swedish Open Cultural Heritage)

- **SOCH** (K-samsök in Swedish) is a linked data aggregator and web service for cultural heritage data, operated by the Swedish National Heritage Board in partnership with 80 other CHIs since 2010 <https://www.raa.se/ksamsok>

Aeroseum

Ájtte

ArkDes

Armémuseum

ARSENALEN
SVERIGES FÖRSVARSFORDONSMUSEUM

BLEKINGE MUSEUM

BOHUSLÄNS MUSEUM

Dramawebben

ENKÖPINGS KOMMUN

ETNOGRAFISKAMUSEET

ARLANDA FLYGSAMLINGAR

FALBYGDENS MUSEUM
FALKÖPING

GÖTEBORGS NATURHISTORISKA MUSEUM

Hemslöjden

Hälsinglands Museum

JAMTLI

FÖRSVARS museum
BODEN ★ MMVI

GOTLANDS FÖRSVARSMUSEUM

GRENNA MUSEUM
ANDRÉEXPEDITIONEN POLARCENTER

Hallwylska museet

Historiska

flygvapenmuseum

GÖTEBORGS STADSMUSEUM

Livrustkammaren

Linköping
Där idéer blir verklighet

JÄRNVÄGSMUSEET

JÖNKÖPINGS LÄNS MUSEUM

KARLSBORGS FÄSTNINGSMUSEUM

KULTUREN

Kulturmagasinet

KULTUR PARKEN SMÅLAND

KÖPINGS KOMMUN

Litografiska Museet
LITOGRAFKLUBBENS ARKIV

LÄNSMUSEET GÄVLEBORG

Malmö Museer

MARINMUSEUM KARLSKRONA

MEDELHAVSMUSEET

MILISEUM SKILLINGARYD

POSTMUSEUM

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

RKC MALMÖ

SJÖHISTORISKA MUSEET

SKANSEN

Skoklosters slott

SNUS & TÄNDSTICKS MUSEUM

SOUNDS OF CHANGES

NORDISKA MUSEET

STOCKHOLMS LÄNS MUSEUM

Swedish Rock Art Research Archives
Svenskt HällristningsForskningsArkiv
ETT NATIONELLT ARKIV FÖR DOKUMENTATION OCH FORSKNING

SVERIGES MILITÄRHISTORISKA ARV
UPPLEVELSEMUSEER

sörmlands museum

TEKNIKLAND

TELESEUM
SVERIGES FÖRSVARSTELEMUSEUM

TEKNISKA MUSEET

THE UNSTRAIGHT MUSEUM beta

TRELLEBORGS MUSEUM

UPPLANDSMUSEET

VASA MUSEET

vänermuseet

VÄNERSBORGS MUSEUM

VÄRLDSKULTURMUSEET

VÄRMDÖ KOMMUN

VÄRMLANDS MUSEUM

VÄSTRA GÖTALANDSREGIONEN
VÄSTARVET

VÄSTERGÖTLANDS MUSEUM

västmanlands läns MUSEUM
– en del av Region Västmanland

ÖREBRO LÄNS MUSEUM

ÖSTASIATISKAMUSEET

ÖSTERLENS MUSEUM

ÖSTERGÖTLANDS MUSEUM

SOCH (Swedish Open Cultural Heritage)

- **SOCH** (K-samsök in Swedish) is a linked data aggregator and web service for cultural heritage data, operated by the Swedish National Heritage Board in partnership with 80 other CHIs since 2010 <https://www.raa.se/ksamsok>

- Mission: *To make it easier for people to find and use digital cultural heritage for work, for education, and for pleasure.*

- SOCH consists of three parts:
  - Technical infrastructure
  - Content and data
  - Network of partners

- Partnership in SOCH is free and voluntary but requires:
  - Stable long-term management
  - That records are already published online
  - Rights statements for all media
  - CC0 for all metadata
  - Data mapping
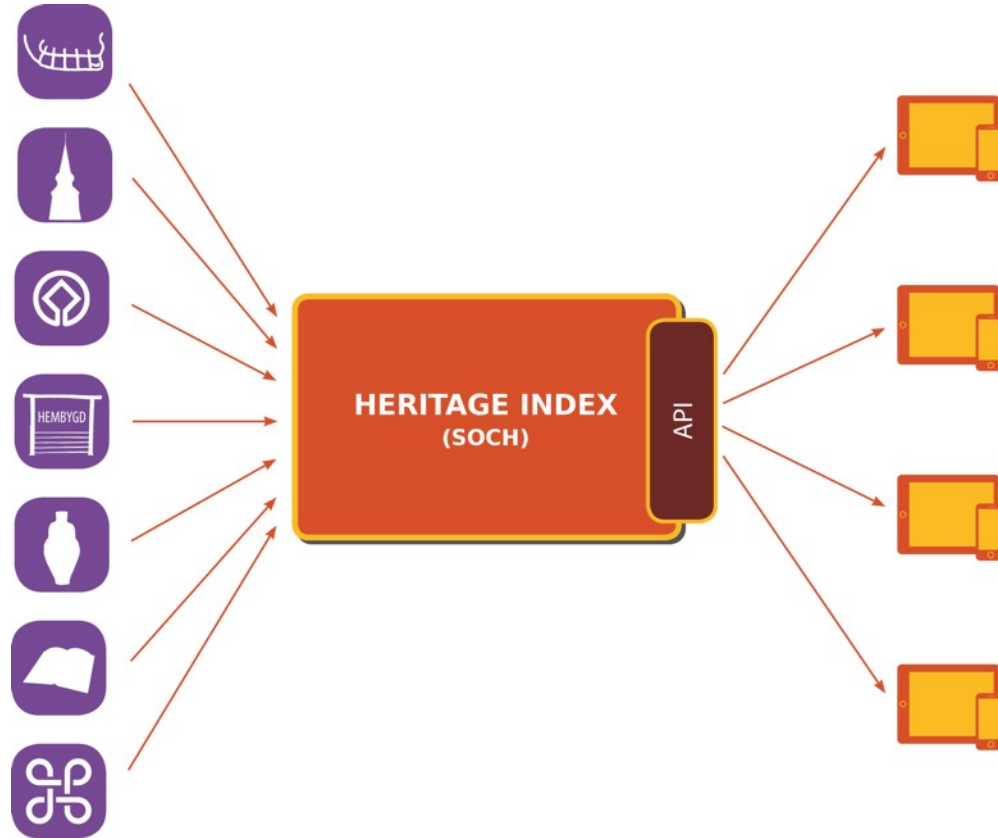

Stativ. Tekniska museet, CC BY (cc)

SOCH technical infrastructure
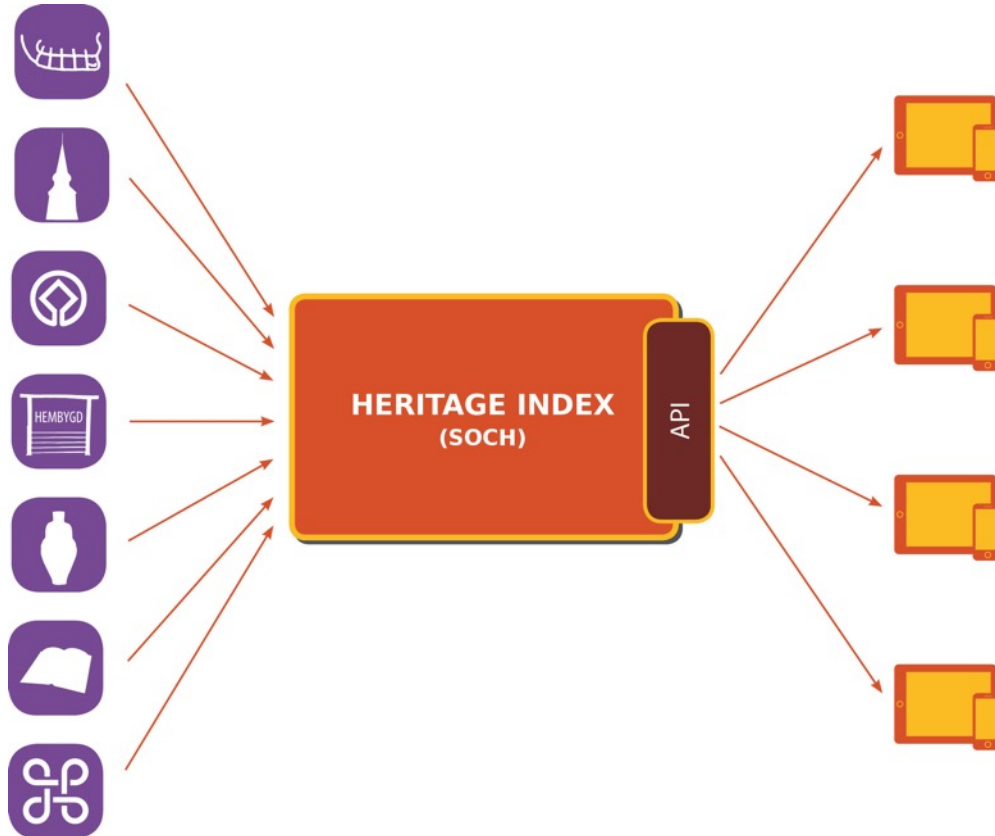
- A unified interface to heritage data from multiple systems and institutions

- Harmonised to a common data model

- Assigns persistent identifiers

- Published as linked data

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

SOCH technical infrastructure

- Metadata is harvested via OAI-PMH and is stored as RDF/XML

- The metadata is indexed

- Calls to the web API search against the index

- Resolving a record's URI returns its RDF document from the database, as RDF/XML or JSON-LD

OAI-PMH – http/XML-based protocol for efficient metadata harvesting

RDF – "Resource Description Framework", metamodel and W3C-standard for linked data

RDF/XML & JSON-LD – serialisation formats for RDF data

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

SOCH contents

- The index currently contains over 9.5 million records, mostly artefacts and photographs from museums but also monuments from the Cultural Environment Register (KMR) and historic buildings from BeBR.
- Of those 9.5 million,
  - c 5.1 million with images
  - c 2 million with geodata
- Record types include:
  - photographs
  - artefacts
  - ancient monuments
  - built heritage
  - documents
  - publications

  - maps
  - audio
  - video
  - stories
  - people
  - events
  - …

En kvinna demonstrerar Frösöstenen, J RS1928:66
Bengt Wellert, Östersund, Stiftelsen Jamtli

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

SOCH contents

- Only metadata is harvested and stored

- All metadata is open (CC0)

- All media must be marked up with rights statements; there is no requirement for an open license, but it is strongly encouraged (58% are PD, CC0, or CC BY; a further 36% are other CC BY-*)

- SOCH uses its own bespoke RDF data model <https://www.raa.se/hitta-information/k-samsok/att-anvanda-k-samsok/protokoll-och-parametrar/>

- The SOCH data model is primarily attribute- and relation-based, with some influence from CIDOC-CRM: SOCH uses "contexts" to describe events in the life of artefacts and monuments: creation, ownership, presence, etc. Contexts have associated dates and locations.

- URIs are on the kulturarvsdata.se domain

```
    ],
    ":itemClassName": "Uppstoppade djur (Naturalier, Djur)",
    "itemDescription": "_:b92",
    ":itemKeyWord": [
      "Häst",
      "Stormaktstid",
      "Slaget vid Lützen",
      "Gustav II Adolf",
      "Gustav II Adolfs häst"
    ],
    ":itemLabel": "Gustav II Adolfs häst från slaget vid Lützen 1632-11-06, kallad Streiff.",
    "itemLicense": "http://kulturarvsdata.se/resurser/License#cc0",
    "itemLicenseUrl": "http://creativecommons.org/publicdomain/zero/1.0/",
    "itemMeasurement": [
      "_:b60",
      "_:b63",
      "_:b66",
      "_:b98"
    ],
    "itemName": "_:b33",
    "itemNumber": "_:b11",
    "itemSuperType": "http://kulturarvsdata.se/resurser/EntitySuperType#object",
    "itemType": "http://kulturarvsdata.se/resurser/EntityType#object",
    ":ksamsokVersion": "1.1",
    ":lastChangedDate": "2015-08-18",
    ":mediaType": "text/html",
```

SOCH and Linked Open Data

- Links describe *how* records and resource relate to one another, both within a dataset/collection, across datasets, and to resources outside SOCH, e.g.:

  – an artefact *is depicted by* an image;

  – a document *describes* an ancient monument;

  – a person *owned* or *used* an object;

  – a record *consists of*, or *is a part of*, another record;

  – Identities: a record is the *same as*, or *supersedes*, another record

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET
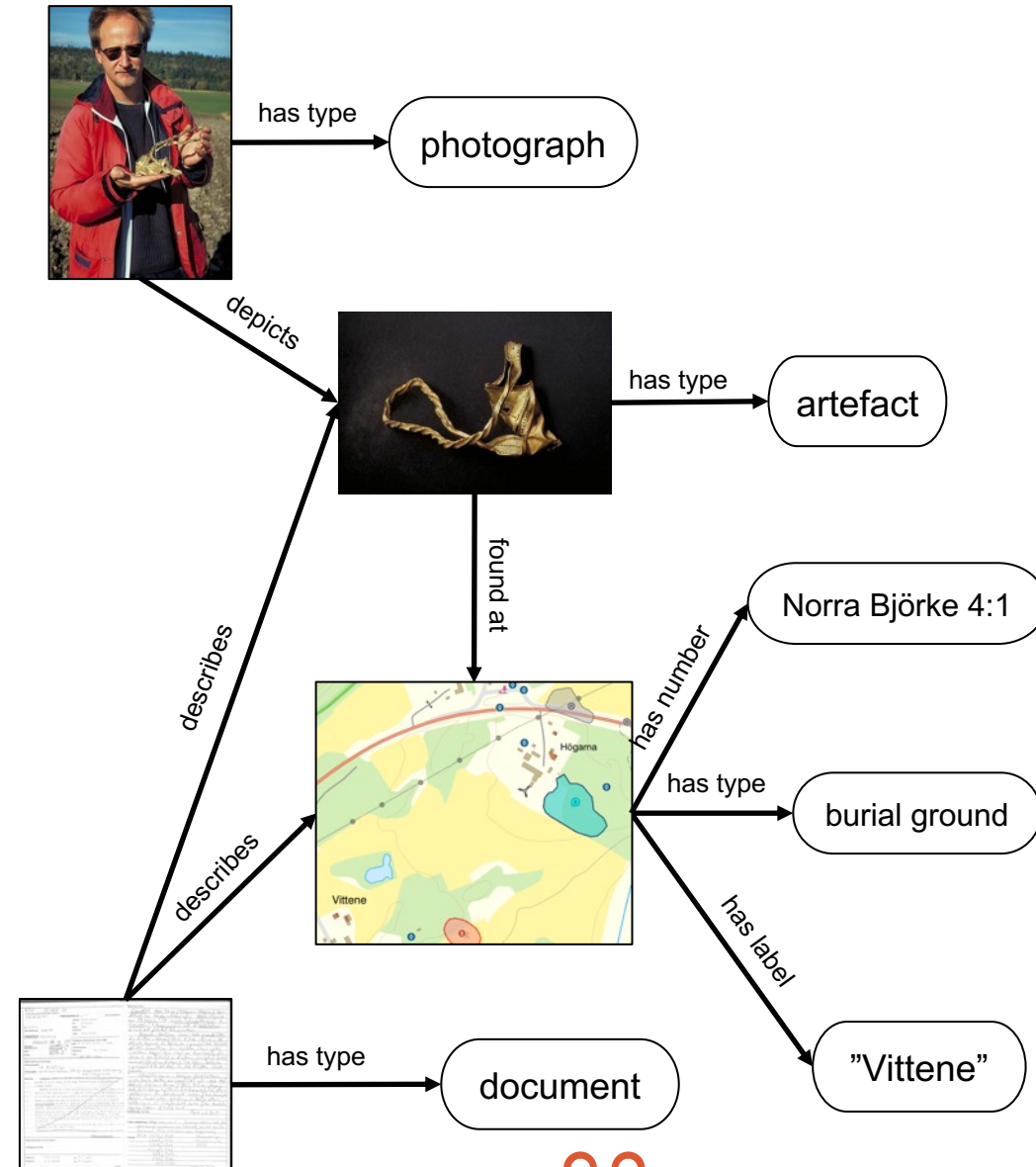
SOCH and Linked Open Data

- These relations can also point out to resources outside of SOCH e.g.:
  - Wikipedia articles
  - Wikimedia images
  - Wikidata resources
  - VIAF (Virtual International Authority File, for e.g. authors and publications)
  - Shared structured vocabularies: Getty AAT, PeriodO, KulturNav, FISH/HeritageData
  - Books and articles at the Swedish National Library
  - Still more at Europeana

Standard Radio factory in Ulvsunda, Stockholm. Arrangement of pylons for telephony equipment. Public Domain, Tekniska museet

SOCH API

- SOCH also has its own search API (Application Programming Interface):

```
https://kulturarvsdata.se/ksamsok/api
?method=search&query=text=runsten
%20AND%20itemType=foto
```

```
https://kulturarvsdata.se/ksamsok/api
?method=getRelations&relation=all
&objectId=raa/lamning/257068c5-8b5f-
4c3e-926a-dbd60b223c49
```


That belongs in a museum..

available online through a Semantic Web Endpoint

Applications

- **Kringla** web interface at <https://www.kringla.nu/>
- Numerous mobile apps and websites using SOCH data and API https://www.raa.se/hitta-information/k-samsok/att-anvanda-k-samsok/exempel-pa-anvandning-av-k-samsok/ e.g. :
  - Runor <https://app.raa.se/open/runor/>
  - GeoStory <https://www.geostory.se/>
  - Fornfynd
  - etc
- SOCH partners also use it to access their own data and create digital exhibitions
- Integration with Wikimedia Commons and Wikidata, Libris, etc
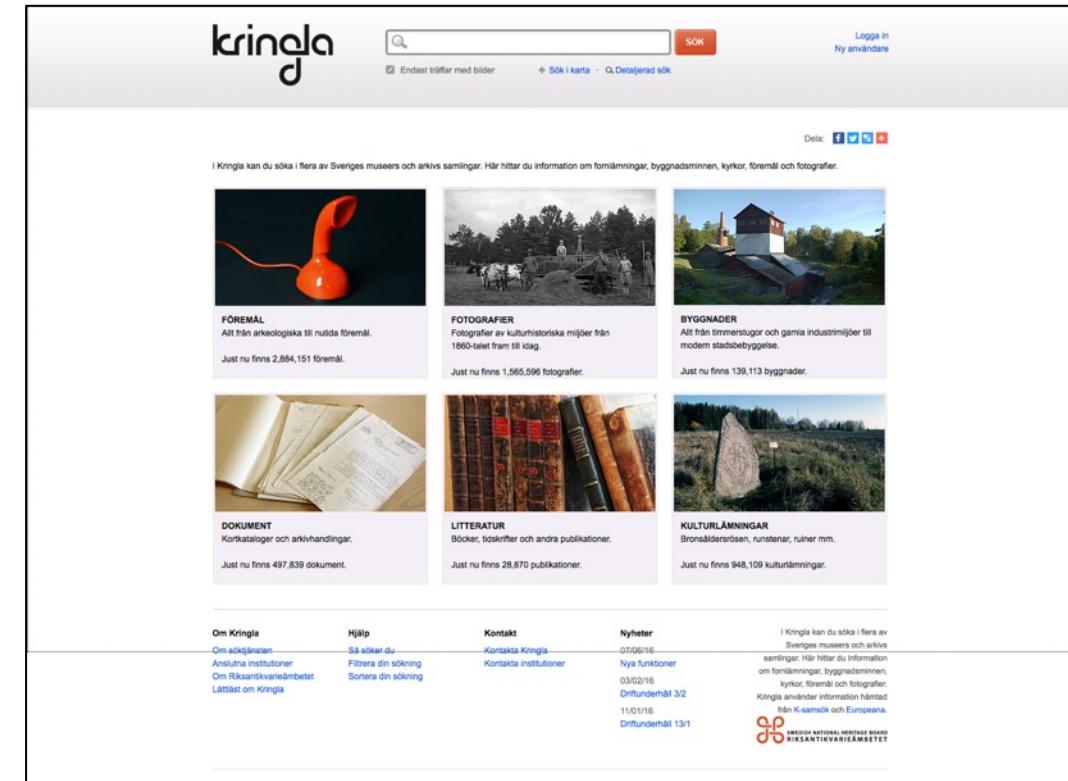
User-Generated Content (UGC)

- SOCH has a separate, parallel service for third parties (users) to add links between records or to external resources

- User-contributed links are accessible via a separate API: <https://ugc.kulturarvsdata.se/>

- Links can be contributed via this API by trusted services allotted an API-key
(and, formerly, via user log in on Kringla)

- 2.7 million links at present

F6 Karlsborg 1957.Vinterövning i Norrland.
Tankningsanslutningar sammankopplas.
Public Domain, Karlsborgs fästningsmuseum

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Europeana

- **Europeana** <https://europeana.eu/> is the European cultural heritage aggregator
- It works in much the same way as SOCH, but on a larger scale, harvesting metadata from national, domain, and topic aggregators across Europe (including SOCH)
- More of an archives and libraries focus, in addition to museums and galleries, with an emphasis on visual media
- Europeana uses **EDM**, the Europeana Data Model
- Europeana also has its own API.
- In contrast to SOCH, Europeana works hard to enrich the incoming data it harvests, using machine translation, keyword matching, links to Wikidata etc.
- Europeana also maintains hand-curated digital collections, to aid in browsing the vast amount of material it aggregates
- Europeana also works closely with aggregators and CHIs to improve metadata quality
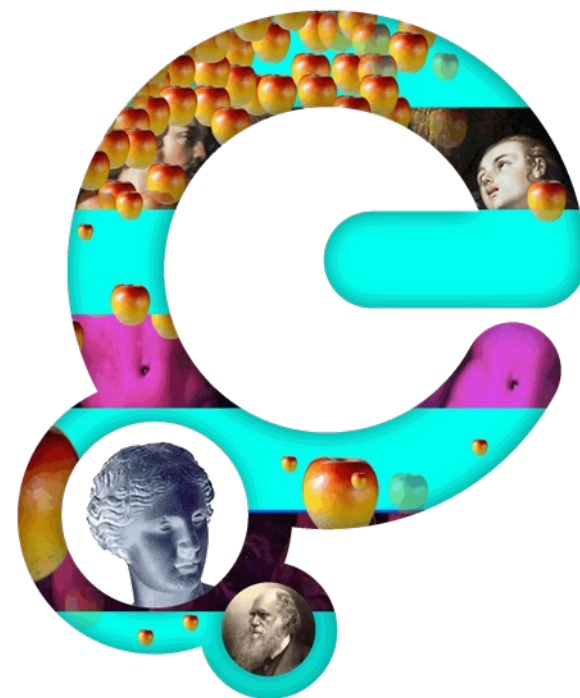
europeana
think culture

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Citizen Science

- SOCH and our open data services are focused on ensuring that everyone – including researchers and the general public – can find, access, and use the data that exists in order to aid in their research.

- Finds information comes in to the Cultural Environment Register (and on to SOCH) from county administrative boards – either in connection to excavations, or as newly reported find-spots – or to SOCH from museums

- Our focus is on enabling access to useful information

- Case study: FornPunkt <https://fornpunkt.se/>

  - A new third-party service built on SOCH and our open data services – primarily sites and monuments, and images

  - Allows users to report finds and sites

  - Published as linked open data

  - Integrated with Wikidata, Wikimedia

  - Already in use by some county administrative boards as a planning tool in conjunction with the CER and source of potential sites to investigate (when confirmed, they get officially registered)
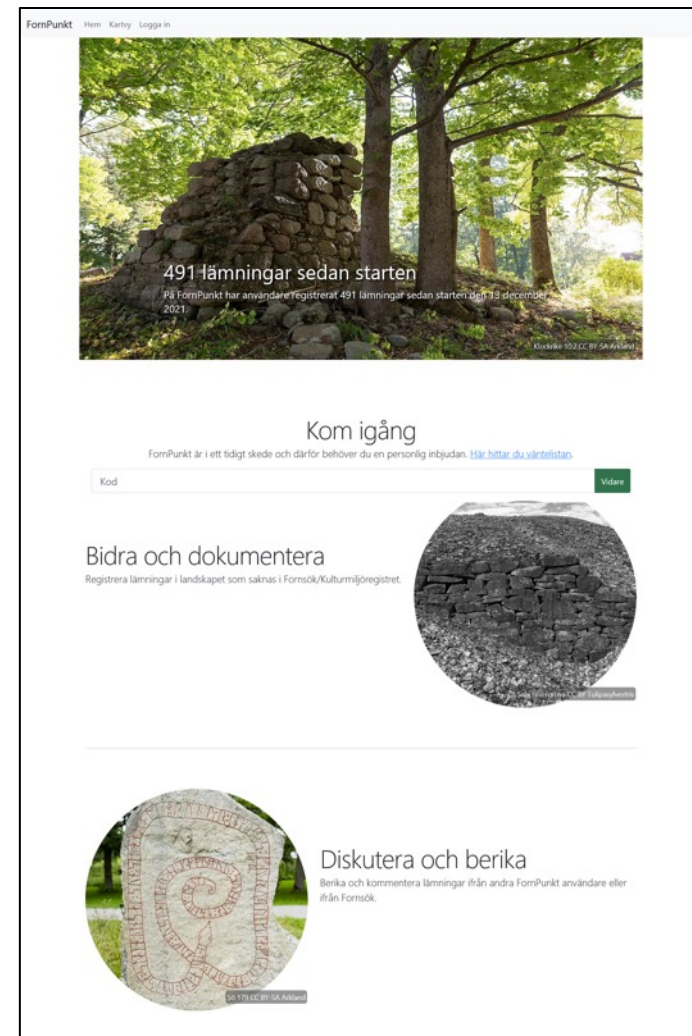
Current and future developments: SOCH 2.0, Urdar, and SweDigArch

SOCH is Showing its Age

- The landscape has changed since SOCH was created – there are new requirements, new expectations… and new possibilities
- There is an increasing technical debt:
  We need a modern, stable, future-proof platform
- SOCH uses a bespoke data model and API:
  We want to align closer to established standards –
  EDM & CIDOC-CRM, REST & SPARQL
- There is also a need to simplify harvesting for smaller CHIs, to lower the barrier to entry

- Time for a new SOCH!

Stickad strumpa
CC BY Lilli Zickerman, Hemslöjdens samlingar

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Part of Something Greater

- We see an increasing need within the Swedish heritage sector for a shared national infrastructure
- A shared digital infrastructure for CHIs
- Increased focus on linked open data, both in terms of data quality and in links between datasets/to shared authorities
- For CHIs, for developers, for researchers, and for the general public

Adam Bromée (3 år 11 mån) fascineras av CD-ROM.
CC BY-NC-SA 1998 Göran Hofstedt, Sörmlands museum

New Data Model

- **Layer 0:** Base-level; bare minimum metadata requirements. Simple Dublin Core & Schema.org statements, + rights statements.

- **Layer 1:** Normal level; requires Layer 0. EDM!

- **Layer 2:** Advanced level; requires Layer 1. CIDOC-CRM and its extensions (especially CRMarchaeo).

- **IIIF-layer:** Independent of L1 & L2. IIIF

Cat and Meccano
CC BY-NC Hilding Mickelsson, Hälsinglands museum

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Next Steps

- New data model, new technical infrastructure
- Better support for modern standards: SPARQL, etc
- Authorities/vocabularies for monument types, fieldwork intervention types
- Better support for external resources – a gradual move from aggregator to index
- Planning and preparation 2022
- Development to begin 2023
- Watch this space!

Dammode, kvinnlig modell poserar iklädd draperad aftonklänning och vidbrättad hatt.
CC BY-NC-ND Erik Holmén 1938, Nordiska museet

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET

Urdar

- 2020–2023 funded by the Swedish Foundation for Humanities and Social Sciences
- Collaboration between the National Heritage Board and Uppsala University, with Umeå, Lund and others.
- Much in common with Norwegian ADED.
- Aims:
  - **Preserve** archaeological documentation from 3 600 Intrasis-projects produced by Archaeological Unit of the National Heritage Board 2000–2014 in archivable formats
  - **Develop an information model** for digitally born and digitised analogue archaeological documentation stored by the NHB Archive
  - Make the digital documentation delivered by the project **FAIR** via Fornsök, Arkivsök, SOCH
  - **Develop methods** to enable data-driven research on digitalised archaeological documentation
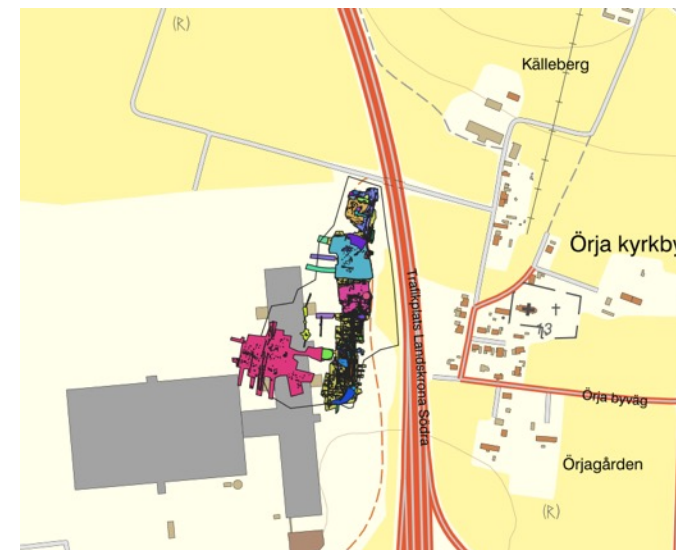
SweDigArk

- 2022–2027 funded by the Swedish Research Council
- Consortium of 6 universities headed by Uppsala, plus NHB and Swedish Historical Museums
- Enable data-driven analysis of combinations of complex archaeological information
  - Ancient monuments, GIS-data, museum collections and archival materials, 3D models of sites and artefacts, laboratory analysis of geology, isotopes, lipids, paleoecology, paleoentomology, C14 and dendrochronology… etc
- Increase cross-disciplinary and international access to Swedish archaeological and cultural heritage data
- Support research on environmental history and sustainability
- Provide a resource for planning, environmental- and sustainability analyses

http://swedigarch.se/

SWEDISH NATIONAL HERITAGE BOARD
RIKSANTIKVARIEÄMBETET
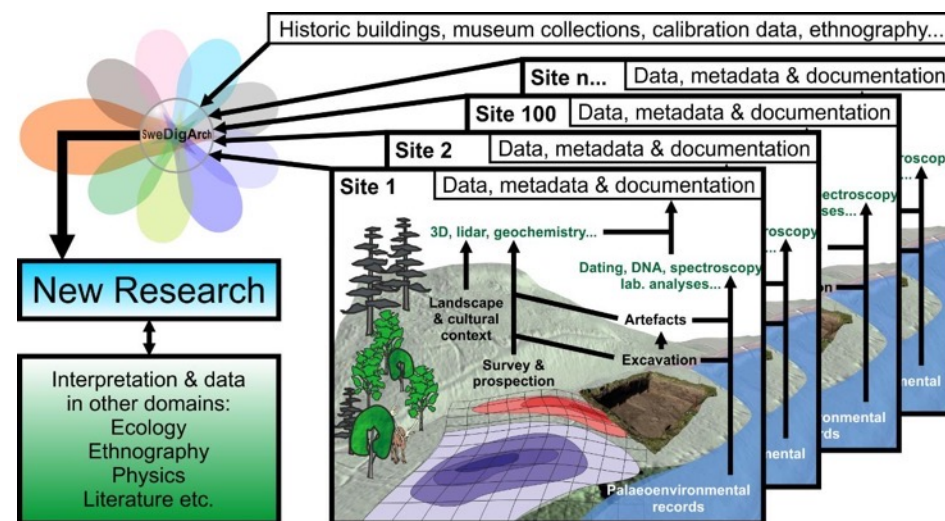
SweDigArk

- Implement FAIR data principles for SweDigArch resources: collections, national records and research databases

- Aggregate the content in the SweDigArch resources and publish the metadata, links and persistent identifiers through REST API and SPARQL endpoint

- Help increase digital competence among researchers, archaeologists, museums, government agencies etc
  - Workshops, hackathons, networking, courses etc

- Built on top of a new SOCH 2.0
  - A new data model with modern standards
  - New output services
  - Open up for delivery from research data repositories

http://swedigarch.se/
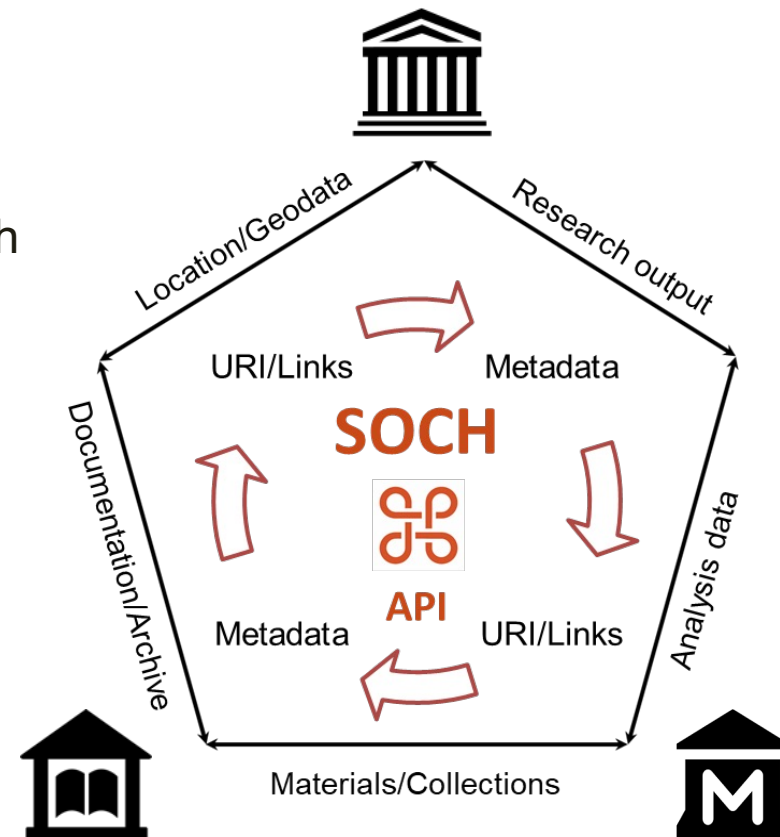
Thank you!

Marcus Smith
marcus.smith@raa.se
@carwash

https://www.raa.se/ksamsok
https://www.kringla.nu/

Tjänsteman vid dator
CC BY-SA, Järnvägsmuseet